

**Modelos propuestos
para la continuidad
en 2015 y 2016 del MCS-ENIGH**

Ajuste demográfico por posestratificación

Víctor Alfredo Bustos y de la Tijera

Ajuste demográfico por imputación

Benito Durán Romo

Ajuste del MCS 2015 por alineación de deciles

Rodrigo Negrete Prieto y Benito Durán Romo

**Propuesta para recuperar la continuidad del MCS 2015
usando máquinas de soporte vectorial**

José Alejandro Ruiz Sánchez

Ajuste de una base de datos vía modelos lineales

Víctor Alfredo Bustos y de la Tijera

**Ajuste estadístico a la distribución del ingreso en
el Módulo de Condiciones Socioeconómicas 2015
mediante imputaciones múltiples**

Delfino Vargas Chanes y Servando Valdés Cruz

**Ajuste por factor de crecimiento con base en
el indicador de gasto de la ENIGH**

Edgar Vielma Orozco, Octavio Heredia Hernández y José Vences Rivera

**Propuesta de ajuste de la distribución del ingreso
corriente total per cápita mediante el gasto corriente
total per cápita**

Alida M. Gutiérrez-Landeros, Marco A. Romero-Navarro, Mariana Galindo-Orozco, Carolina Chávez-Ruelas, David Rojas-Rosey, David López-Lira Bayod y Ricardo Aparicio-Jiménez

**Modelo estadístico 2016 para la continuidad
del MCS-ENIGH**

José Alejandro Ruiz Sánchez y Ana Miriam Romo Anaya

**Incentivos, error de medición y estimación
de la pobreza en México**

Alfonso Miranda y Jaime Sainz Santamaría



Contenido

Introducción	3
Ajuste demográfico por posestratificación <i>Post-stratification demographic adjustment</i> Víctor Alfredo Bustos y de la Tijera	12
Ajuste demográfico por imputación <i>Demographic adjustment by imputation</i> Benito Durán Romo	28
Ajuste del MCS 2015 por alineación de deciles <i>Adjustment of MCS 2015 by means of decile alignment</i> Rodrigo Negrete Prieto y Benito Durán Romo	58
Propuesta para recuperar la continuidad del MCS 2015 usando máquinas de soporte vectorial <i>A proposal for recovering MCS 2015's continuity using Support Vector Machines</i> José Alejandro Ruiz Sánchez	88
Ajuste de una base de datos vía modelos lineales <i>Setting up a database via linear models</i> Víctor Alfredo Bustos y de la Tijera	112
Ajuste estadístico a la distribución del ingreso en el Módulo de Condiciones Socioeconómicas 2015 mediante imputaciones múltiples <i>Statistical Adjustment to Income Distribution in the MCS 2015 through Multiple Imputations</i> Delfino Vargas Chanes y Servando Valdés Cruz	144
Ajuste por factor de crecimiento con base en el indicador de gasto de la ENIGH <i>Adjustment by means of a factor growth based on an ENIGH's expenditure indicator</i> Edgar Vielma Orozco, Octavio Heredia Hernández y José Vences Rivera	162
Propuesta de ajuste de la distribución del ingreso corriente total per cápita mediante el gasto corriente total per cápita <i>Adjustment of per capita income distribution by per capita expenditure: a proposal</i> Alida M. Gutiérrez-Landeros, Marco A. Romero-Navarro, Mariana Galindo-Orozco, Carolina Chávez-Ruelas, David Rojas-Rosey, David López-Lira Bayod y Ricardo Aparicio-Jiménez	200
Modelo estadístico 2016 para la continuidad del MCS-ENIGH <i>2016 Statistical Model for MCS-ENIGH's Continuity</i> José Alejandro Ruiz Sánchez y Ana Miriam Romo Anaya	214
Incentivos, error de medición y estimación de la pobreza en México <i>Incentives, measurement error, and poverty estimation in Mexico</i> Alfonso Miranda y Jaime Sainz Santamaría	240
Colaboran en este número	258

INSTITUTO NACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Presidente del Instituto

Julio Alfonso Santaella Castell

Vicepresidentes

Enrique de Alba Guerra

Mario Palma Rojo

Paloma Merodio Gómez

Dirección General de Estadísticas Sociodemográficas

Edgar Vielma Orozco

Dirección General de Estadísticas de Gobierno, Seguridad Pública y Justicia

Adrián Franco Barrios

Dirección General de Estadísticas Económicas

José Arturo Blancas Espejo

Dirección General de Geografía y Medio Ambiente

Claudio Martínez Topete, encargado del despacho

Dirección General de Integración, Análisis e Investigación

Enrique Jesús Ordaz López

Dirección General de Coordinación del Sistema Nacional de Información Estadística y Geográfica

María Isabel Monterrubio Gómez

Dirección General de Vinculación y Servicio Público de Información

Eduardo Javier Gracida Campos

Dirección General de Administración

Marcos Benerice González Tejeda

Contraloría Interna

Francisco Hugo Gutiérrez Dávila, encargado del despacho

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA

Editor responsable

Enrique Jesús Ordaz López

Editor técnico

Gerardo Leyva Parra

Coordinación editorial

Virginia Abrín Batule y Mercedes Pedrosa Islas

Corrección de estilo

José Pablo Covarrubias Ordiales y Laura Elena López Ortiz

Corrección de textos en inglés

Gerardo Piña

Diseño y edición

Juan Carlos Martínez Méndez y Eduardo Javier Ramírez Espino

Indizada en: Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal *Latindex Catálogo*; Citas Latinoamericanas en Ciencias Sociales y Humanidades (*CLASE*) y en la Red Iberoamericana de Innovación y Conocimiento (REDIB).

Número especial de este volumen 9 del 2018 de REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, editada por el Instituto Nacional de Estadística y Geografía, Avenida Héroe de Nacozari Sur 2301, Fraccionamiento Jardines del Parque, 20276, Aguascalientes, Aguascalientes, Aguascalientes, entre la calle INEGI, Avenida del Lago y Avenida Paseo de las Garzas, México. Teléfono 55 52781069. Toda correspondencia deberá dirigirse al correo: rde@inegi.org.mx

Editor responsable: Enrique Jesús Ordaz López. Reserva de Derechos al Uso Exclusivo del Título Núm. 04-2012-121909394300-102, ISSN Núm. 2007-2961, ambos otorgados por el Instituto Nacional del Derecho de Autor. Certificado de Licitud de Título y Contenido Núm. 15099, otorgado por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación. Domicilio de la publicación, imprenta y distribución: Avenida Héroe de Nacozari Sur 2301, Fraccionamiento Jardines del Parque, 20276, Aguascalientes, Aguascalientes, Aguascalientes, entre la calle INEGI, Avenida del Lago y Avenida Paseo de las Garzas, México.

El contenido de los artículos, así como sus títulos y, en su caso, fotografías y gráficos utilizados son responsabilidad del autor, lo cual no refleja necesariamente el criterio editorial institucional. Asimismo, la Revista se reserva el derecho de modificar los títulos de los artículos, previo acuerdo con los autores. La mención de empresas o productos específicos en las páginas de la Revista no implica el respaldo por el Instituto Nacional de Estadística y Geografía.

Se permite la reproducción total o parcial del material incluido en la Revista, sujeto a citar la fuente. Esta publicación consta de 400 ejemplares y se terminó de imprimir en abril del 2018.

Versión electrónica: <http://rde.inegi.org.mx>

ISSN 2395-8537

CONSEJO EDITORIAL

Enrique de Alba Guerra

Presidente del Consejo

Fernando Cortés Cáceres

Profesor Emérito de FLACSO

PUED de la UNAM

México

Gerardo Bocco Verdinelli

Universidad Nacional Autónoma de México

México

Juan Carlos Chávez Martín del Campo

Banco de México

México

Lidia Bratanova

UNECE Statistical Division

Switzerland

Tonatiuh Guillén López

El Colegio de la Frontera Norte, AC

México

Víctor Manuel Guerrero Guzmán

Instituto Tecnológico Autónomo de México

México

Introducción

En diciembre de 2017, a solicitud del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), el Instituto Nacional de Estadística y Geografía (INEGI) dio a conocer en su página de internet los resultados del Modelo Estadístico 2015 para la continuidad del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH). El Modelo se presentó con el objetivo de "...proveer a los usuarios de la información con los insumos necesarios para la medición quinquenal de la pobreza multidimensional municipal, correspondiente a 2015...". La necesidad de este modelo, a su vez, se derivó de que los resultados del MCS 2015, que el INEGI publicó en julio del 2016, presentaban una discontinuidad en el ingreso de los hogares en relación con los datos del MCS-ENIGH 2014, ya que mostraba variaciones inconsistentes, respecto de lo que se esperaba como razonable dada la dinámica de la actividad económica y del empleo ocurrida entre el 2014 y 2015. Era, entonces, evidente que las cifras de los ingresos de los hogares del MCS 2015 no encajaban con la serie bienal del MCS-ENIGH que comenzó en el 2008. Esta discontinuidad resultó fundamentalmente de impactos no esperados de mejoras en la capacitación y en la estrategia operativa, orientados a lograr un mayor apego al diseño original de la Encuesta. Las mejoras en la capacitación hicieron énfasis en la importancia de evitar que los informantes evadieran reportar su ingreso, que ofrecieran información sobre sus ingresos claramente inconsistente con lo necesario para subsistir, con su evidente nivel de vida o con el resto de la información entregada en la ENIGH. Por su parte, la estrategia operativa implementó sistemas más acuciosos del seguimiento del avance en campo y estableció mecanismos de reconsulta que se detonaban desde oficinas centrales para verificar los reportes de ingresos nulos y otras inconsistencias que se detectaban en el análisis de gabinete.¹

Dado que el objetivo fundamental del MCS 2015 era la medición de la pobreza multidimensional por parte del CONEVAL y que dicha institución requería que las cifras de ingresos mantuvieran la continuidad con la serie previa, junto con el INEGI decidieron integrar un grupo técnico de trabajo (al que luego se sumó un conjunto de es-

¹ Para una descripción más detallada de la investigación de las causas de la discontinuidad del ingreso de los hogares que apareció en el MCS 2015, ver: <http://www.beta.inegi.org.mx/proyectos/investigacion/invenc/>

pecialistas académicos y de la sociedad civil) al cual se denominó Grupo Técnico Ampliado (GTA), cuyo propósito era realizar y discutir una serie de trabajos de diagnóstico sobre lo ocurrido a partir de la cual se consideró que era posible generar un modelo estadístico que permitiera recuperar la continuidad de la serie histórica mediante la combinación de los resultados del MCS 2015 con los provenientes de levantamientos previos, así como del posible uso de fuentes complementarias que ayudaran a establecer el contexto de los cambios en el tiempo. La elección del modelo, sin embargo, no era obvia, por lo que se decidió avanzar de manera simultánea en varias avenidas metodológicas, cada una con sus propios supuestos, alcances y limitaciones. De entre las distintas alternativas consideradas se eligió la que finalmente se consideró como la más adecuada para atender el problema en cuestión, la cual se publicó en la sección de investigación de la página del INEGI en internet,² y cuya metodología se usó también en el Modelo Estadístico 2016 para la continuidad del MCS-ENIGH.³ Los usuarios especializados podrán encontrar ahí no solo los resultados del Modelo sino, también, una nota técnica que lo describe. Sin embargo, en aras de una mayor transparencia y por considerarlo de utilidad e interés para los especialistas, en este número especial de REALIDAD, DATOS Y ESPACIO. REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA, el INEGI presenta un artículo que amplía y detalla la nota técnica arriba referida, así como los trabajos correspondientes a las alternativas que se ensayaron al interior del GTA. Con ello se espera dejar testimonio del esfuerzo realizado, a la vez de ofrecer, a través de estos trabajos, elementos adicionales que abordan, con distintos énfasis y perspectivas, aristas complementarias de una misma realidad, las cuales, al final, pueden resultar en aspectos informativos útiles para tener una apreciación más cabal de lo que está detrás de la discontinuidad de los ingresos de la serie del MCS-ENIGH en el 2015 y 2016, así como de varios de los principales retos que se enfrentan al momento de generar información vía encuestas de ingresos aplicadas a los hogares en México.

2 <http://www.beta.inegi.org.mx/proyectos/investigacion/eash/2015/>

3 <http://www.beta.inegi.org.mx/proyectos/investigacion/eash/2016/>

Cuando comparamos entre dos encuestas que hacen las mismas preguntas sobre ingreso corriente total de los hogares a la misma población, pero en momentos distintos, sabemos que los ingresos reportados pueden ser diferentes porque cambió: 1) el número total de hogares, 2) el ingreso promedio por perceptor en el hogar y 3) el número de perceptores en el hogar. Los dos primeros artículos de este número especial abordan la pregunta de en qué medida el cambio de los ingresos entre el 2014 y 2015 responde a un incremento en el número de hogares reportados en la muestra que pudiera ir más allá de lo que las tendencias de la dinámica demográfica nos harían esperar. Cabe subrayar que estos dos trabajos no pretenden alinear las otras dos fuentes de cambio del ingreso para el año en cuestión, sino única y exclusivamente la que proviene del incremento en el número de hogares. De esta manera, el artículo de Víctor Alfredo Bustos y de la Tijera, titulado *Ajuste demográfico por posestratificación*, modifica el número de hogares mediante una técnica de posestratificación que busca corregir los factores de expansión de los hogares según su estrato sociodemográfico, para acercarlos a la estructura de la Encuesta Intercensal 2015. Por su parte, el de Benito Durán Romo, que se presenta con el título de *Ajuste demográfico por imputación*, utiliza técnicas de imputación tipo *Cold Deck* con las cuales asigna integrantes adicionales a los hogares levantados en el MCS 2015, de manera que, con unos más grandes, sea necesaria una menor cantidad de ellos para expandir a la misma población total de referencia, atemperando, así, la tendencia del tamaño promedio de los hogares hasta hacerla consistente con las tendencias implícitas de las proyecciones de población del Consejo Nacional de Población (CONAPO). Es importante notar que cada uno de estos dos enfoques responde a una hipótesis distinta sobre el origen de la diferencia en el número de hogares observado en el 2015. Así, la posestratificación supone que, en tanto la información captada es precisa, el problema a corregir está en las discrepancias entre tamaños relativos de subpoblaciones tanto según la Encuesta Intercensal como los resultantes de la selección de la muestra y del levantamiento. En contraste, el enfoque de imputación *Cold Deck* supone que el problema proviene de campo, en el sentido de que al momento de aplicar el cuestionario se hubiera tenido alguna subenumeración de pobladores de los hogares, dando como resultado unos de tamaño promedio menor al real

y, en consecuencia, un mayor número de hogares (necesario para cuadrar con el total de población de la proyección del CONAPO utilizada como referente o *benchmark*).

Ahora, considerando los ejercicios propiamente diseñados para tratar de recuperar la continuidad de los ingresos del MCS 2015 con respecto a la serie previa, comenzamos por el propuesto por Rodrigo Negrete Prieto y Benito Durán Romo que se presenta con el nombre de *Ajuste del MCS 2015 por alineación de deciles*. En éste se presenta una técnica que ellos denominan *Delta*, la cual parte de la premisa de que las mejoras en la captación del ingreso implementadas en el 2015 incidieron en una menor cantidad de hogares con ingresos cero o con unos bajos de manera desproporcionada (*pseudoceros*), lo cual implicó un desplazamiento hacia arriba de los valores de los ingresos que delimitan cada uno de los deciles de hogares cuando éstos se ordenan según su nivel de ingreso corriente total. A partir de lo anterior, los autores notan que cada decil de hogares reportado en el MCS 2015 combina hogares con ingresos que cabría esperar en ese decil, con otros con ingresos que en los levantamientos anteriores se habrían presentado en un decil más alto. Esto significa que al comparar los deciles del MCS 2014 con los del MCS 2015 en realidad no se están comparando *peras con peras*, sino que serían *peras con peranzanas*. En este sentido, la recuperación de la continuidad histórica de la serie supondría quitar en lo posible la fracción *manzana* de los hogares de cada decil (puesto que en los levantamientos previos no correspondería a ese estrato de ingreso, sino a uno mayor) y modificar los factores de expansión de manera que se deje el estrato integrado casi totalmente por hogares con ingresos tipo *pera*. Al hacer esto se estaría reconociendo que el mejor esfuerzo de campo del MCS 2015 habría hecho que los hogares con menores ingresos de este levantamiento tuvieran ingresos promedio mayores a los de los levantamientos inmediatos anteriores, afectando —en una especie de efecto dominó— los ingresos de cada uno de los cuantiles superiores en que es posible ordenar a los hogares según su ingreso.

Otra opción que se exploró consistió en explotar la relación entre el ingreso y sus covariables según fueron captadas en los levantamientos anteriores utilizando téc-

nicas de *Machine Learning*, que es lo que se hace en el trabajo de José Alejandro Ruiz Sánchez denominado *Propuesta para recuperar la continuidad del MCS 2015 usando máquinas de soporte vectorial*. Esta propuesta es similar a la de las *peranzanas* (método *Delta*) descrita arriba en el sentido de que identifica a los hogares de cada decil en el 2015 que, conforme a las experiencias pasadas, corresponderían a ese decil y ajusta los factores de expansión para que ese subconjunto de hogares represente 10% del total de hogares. Sin embargo, la manera de hacer el *pronóstico* del decil al que correspondería cada hogar es diferente, en virtud del uso de técnicas de clasificación. Asimismo, el uso de este método facilita emplear un conjunto de referencia más amplio (en este caso los MCS-ENIGH del 2010, 2012 y 2014) que sirvió como entrenamiento a partir del cual se caracterizaron relaciones entre las covariables y el decil de pertenencia de cada hogar; esta relación es trasladada al MCS 2015 para pronosticar el decil al que pertenecería un hogar en ese año.

Un enfoque más tradicional para intentar recuperar la continuidad de la serie de ingresos de los hogares es el que sigue Alfredo Bustos en su documento *Ajuste de una base de datos vía modelos lineales*, en el que se explota la relación observada en el pasado entre un conjunto amplio de covariables recabadas por la ENIGH y el ingreso, de manera que se pueda aprovechar esa asociación en un modelo de regresión para generar una versión estimada del ingreso a partir de sus covariables en el 2015. A efecto de tomar en consideración las diferencias de los pesos relativos que esas covariables pudieran tener en diferentes grupos de la población, el trabajo modela por separado el comportamiento del ingreso al interior de cada uno de los cuatro estratos socioeconómicos en los que se agrupan las unidades primarias de muestreo en el Marco Nacional de Viviendas, de manera que se tiene un modelo lineal por cada estrato. En adición a lo anterior, este enfoque busca tomar en consideración especificidades coyunturales del año para el que se hizo la estimación (2015) por medio de la explotación de la información implícita en la distribución de los errores. Además de lo anterior, se corrige el número de hogares usando el modelo de posestratificación mencionado arriba.

Partiendo del supuesto de que solo hay que corregir los ingresos y de que se conoce el conjunto de hogares en la muestra del 2015 cuyos ingresos deben ser corregidos por no corresponder con lo que se esperaría a partir de la serie previa, es posible, también, construir un modelo para llevar a cabo imputaciones múltiples, como el que proponen Delfino Vargas Chanes y Servando Valdés Cruz en su *Ajuste estadístico a la distribución del ingreso en el Módulo de Condiciones Socioeconómicas 2015 mediante imputaciones múltiples*. En este modelo se descartan los valores que se van a corregir y se les trata como si fueran en realidad datos faltantes *Missing at Random*, por lo que se supone que la propensión de que algún dato (de ingreso, en este caso) sea faltante se relaciona con algunos de los datos observados, de manera que es posible utilizarlos (p. ej. las covariables del ingreso) para estimar parámetros que permitan modelar bajo ciertos supuestos los valores de los datos faltantes. Dado que lo que se busca reconstruir es la parte baja de la distribución del ingreso, los autores seleccionaron un conjunto amplio de covariables especialmente relacionadas con los ingresos de la población pobre, las cuales se articulan por medio de un modelo estadístico a partir del cual se estiman los valores esperados de las observaciones *faltantes* que se utilizarán para hacer las imputaciones correspondientes. En este trabajo no se estima de forma directa un valor imputado, sino que se recurre a un proceso Monte Carlo de simulación que resulta en 10 distintas bases de datos, que se resumen al final en un solo modelo de regresión con el fin de hacer imputaciones que tomen en consideración la variabilidad de los datos a modificar y, así, tener imputaciones finales lo mejor informadas y más realistas posible.

Una covariable del ingreso que tiene una asociación teórica y empírica relevante con el ingreso es el gasto. En el trabajo de Edgar Vielma Orozco, Octavio Heredia Hernández y José Vences Rivera, *Ajuste por factor de crecimiento con base en el indicador de gasto de la ENIGH*, se busca aprovechar esta asociación y el hecho de que en la ENIGH 2016 el gasto no parece mostrar el desfase con la serie previa que se observa en la información de ingreso. Así las cosas, el modelo de regresión lineal simple que se utiliza busca establecer la relación histórica entre gasto e ingreso y usarla para estimar el ingreso corriente total de los hogares en el 2016 a partir del gasto mone-

tario reportado en ese año y, finalmente, dado que en el MCS 2015 no se levantó información de gastos, recurrir a una interpolación para asignar las estimaciones de ingreso que corresponderían al MCS 2015. Cabe señalar que en este ejercicio no se corrige todo el vector de ingresos sino solo aquellos casos justificados correspondientes a conjuntos de hogares (definidos por entidad federativa, tamaño de localidad y estrato socioeconómico) en los que se observa un desfase entre la dinámica del ingreso y la de gasto del 2014 al 2016.

Entrando ahora en el ámbito de la generación de modelos para reestablecer la continuidad de la serie para el 2016 y siguiendo en la línea de explorar el gasto como elemento articulador en la serie de tiempo del ingreso corriente de los hogares, *Propuesta de ajuste de la distribución del ingreso corriente total per cápita mediante el gasto corriente total per cápita*, artículo desarrollado por Alida M. Gutiérrez-Landeros, Marco A. Romero-Navarro, Mariana Galindo-Orozco, Carolina Chávez-Ruelas, David Rojas-Rosey, David López-Lira Bayod y Ricardo Aparicio-Jiménez, parte del supuesto de que una buena estimación del ingreso de los hogares en el 2016 resulta de suponer que la tasa de crecimiento del ingreso corriente total promedio per cápita del 2014 al 2016 es la misma que la del gasto corriente total per cápita, distinguiendo, para ello, entre la dinámica observada para el ámbito urbano y la del rural. Con ello se logra generar una nueva versión del vector de ingresos de los hogares para el 2016, que es consistente con la trayectoria previamente observada (hasta el 2014).

Luego de considerar todas las alternativas a la mano, el enfoque que finalmente se transformó en la base del Modelo Estadístico 2016 para la continuidad del MCS-ENIGH, así como del Modelo Estadístico 2015 para la continuidad del MCS-ENIGH, corresponde al trabajo desarrollado por Alejandro Ruiz y Ana Miriam Romo Anaya, el cual propone "...una metodología de edición de microdatos basada en el ajuste de una función de distribución bajo restricciones...". El proceso comienza por tomar en consideración el crecimiento de la mediana del ingreso laboral reportado por la ENOE del 2014 al 2016 (para el Modelo 2016) y del 2014 al 2015

(para el Modelo 2015) para cada una de las 32 entidades federativas del país. Este crecimiento es luego aplicado a la mediana del ingreso corriente total del MCS-ENIGH del 2014 de cada uno de los estados. Ello arroja como resultado 32 medianas objetivo del ingreso corriente total de los hogares en el 2015 o en el 2016, según sea el caso, las cuales, junto con el valor máximo observado del ingreso corriente total en cada entidad federativa, se erigen luego en restricciones que debe cumplir la estimación por máxima verosimilitud de los cuatro parámetros de una función de distribución *Beta Generalizada* del tipo II (*GB2*), ajustada a los datos de la distribución del ingreso recolectado en campo correspondiente al MCS 2015 o a la ENIGH 2016. La versión de la distribución del ingreso resultante es luego comparada con el ajuste sin restricciones de la misma forma funcional *GB2* a los datos de ingreso corriente total del MCS 2015 y la ENIGH 2016. Al comparar ambas estimaciones de la misma forma funcional es posible obtener un cociente entre el ingreso restringido y el ingreso sin restringir para cada hogar en la muestra. Este cociente se usa para corregir hogar por hogar el ingreso laboral (en estricto sentido, es el ingreso por trabajo principal) a partir de los datos de ingreso publicados originalmente para el MCS 2015 y la ENIGH 2016. Enseguida, partiendo de una constante empírica de la relación entre el ingreso laboral con su complemento se establece, también para cada hogar, el valor que corresponde al ingreso corriente total del 2015 o del 2016, según sea el caso, cerrándose de esa manera el proceso. Los resultados de estos modelos, con sus correspondientes microdatos, están publicados en la página del INEGI en el apartado de *Investigación* y se puede acceder a ellos desde la liga <http://www.beta.inegi.org.mx/proyectos/investigacion/eash/2016/>.

Finalmente, como artículo de cierre de este número especial se presenta el trabajo *Incentivos, error de medición y estimación de la pobreza en México*, desarrollado por Alfonso Miranda y Jaime Sainz Santamaría. En él se hace una revisión de lo ocurrido en torno a la publicación de los resultados del MCS 2015 y algunas consideraciones respecto a las posibles causas de la discontinuidad de los niveles de los ingresos corrientes de los hogares, en especial en la parte izquierda de la distribución, respecto a la serie histórica previa de los MCS-ENIGH. Indican que tales cam-

bios probablemente son resultado de modificaciones en los incentivos enfrentados por los entrevistadores (por vía de un fortalecimiento en apariencia marginal de su capacitación, así como de un más estricto control de operaciones en campo y de supervisión, orientados a que el personal de campo cumpliera de la mejor manera posible con el desempeño de sus funciones) y que si bien parece ser que la discontinuidad obedece a mejoras operativas y de capacitación en la ENIGH, la ruptura con la serie previa es inobjetable, con la desventaja de que no se cuenta con algún estudio previo que documentara de manera oportuna las ventajas para la medición del ingreso de los cambios implementados. Por otra parte, los autores hacen una revisión de las ventajas y desventajas de los distintos modelos alternativos que se trabajaron desde el INEGI y ofrecen sugerencias del tipo de experimentos que sería necesario realizar para tener una mayor claridad de lo sucedido, además de que proponen líneas de trabajo para profundizar en la mejoría de las estadísticas micro-económicas sobre distribución del ingreso, subrayando en especial las ventajas de los estudios longitudinales de tipo panel.

Con la presentación de esta colección de trabajos, el INEGI busca hacer del conocimiento de las y los especialistas interesadas(os) en la materia los detalles técnicos de los diferentes esfuerzos emprendidos para recuperar la continuidad de la serie de ingresos de los hogares del MCS-ENIGH que se requieren por parte del CONEVAL para dar continuidad a las mediciones de pobreza multidimensional que dicha institución está obligada a realizar a partir de la información que le entrega el INEGI. Con este ejercicio de transparencia, el Instituto quiere, por una parte, reconocer y agradecer los esfuerzos que en el marco del GTA desarrollaron especialistas de la sociedad civil, la Academia, el CONEVAL y el mismo INEGI. Por otra, la publicación de estos trabajos tiene también el objetivo de aportar al acervo de conocimientos sobre esta materia tan relevante, conscientes de que el progreso científico se fortalece con el intercambio libre y abierto de ideas, de manera que esperamos que estos artículos sirvan no solo como reportes de métodos y hallazgos, sino que también se transformen en referentes para apuntalar futuras investigaciones.

Ajuste demográfico por **posestratificación**

Post-stratification demographic adjustment

Víctor Alfredo Bustos y de la Tijera*

* Instituto Nacional de Estadística y Geografía, alfredo.bustos@inegi.org.mx

Easter Weekend On the Beach in Mazatlan Sinaloa Mexico/Matt Mawson/Getty Images



El tamaño promedio de los hogares en el Módulo de Condiciones Socioeconómicas 2015 decreció más de lo que sería razonable esperar, según el crecimiento poblacional establecido en las proyecciones del Consejo Nacional de Población (CONAPO); como consecuencia, se ajustó al alza el número de hogares para que el tamaño de población coincidiera con el proyectado por el CONAPO. En otras palabras, suponiendo que las respuestas recogidas (por ejemplo: número de miembros del hogar) son precisas, los hogares pequeños están sobrerrepresentados en la muestra; por su parte, los grandes están subrepresentados. Acción: corregir los tamaños relativos de estas subpoblaciones con base en ejercicios estadísticos recientes, como la Encuesta Intercensal 2015. Es decir, corregir los tamaños relativos de distintas subpoblaciones a través de ajustes menores a los factores de expansión y no considerar solo el tamaño poblacional proyectado.

Palabra clave: posestratificación.

The average household size in the Socioeconomic Conditions Module 2015 decreased, according to the population growth established in the CONAPO projections, more than it would be reasonable to expect. As a consequence, the number of households was adjusted upwards so that the population size coincides with that projected by CONAPO. In other words, assuming the collected responses (eg, household members) are accurate, small households are over-represented in the sample. For their part, the large are underrepresented. Action: Correct the relative sizes of these sub-populations based on recent statistical exercises, such as the Intercensus Survey 2015. That is, to correct the relative sizes of different sub-populations through minor adjustments to the expansion factors and not only consider the projected population size.

Key word: post-stratification.

Principales resultados y conclusión

El tamaño promedio de los hogares, después de la posestratificación, coincide con el del levantamiento del 2014. Algo semejante ocurre con los tamaños promedio de los hogares por estrato socioeconómico; por otro lado, los tamaños de las entidades federativas experimentan pequeños cambios cuya significancia está por ser establecida. De este modo, el número total de hogares requerido para cubrir la proyección poblacional del CONAPO para el 2015 se reduce de manera que, en lugar del crecimiento anual incluido en la publicación original del MCS 2015 de 3.3%, se tiene un incremento de solo 1.3%, que está en coincidencia con el crecimiento poblacional para el mismo periodo, que es de 1.29 por ciento.

1. Introducción

Una de las características más notorias del Módulo de Condiciones Socioeconómicas (MCS) 2015 fue el hecho de que el tamaño promedio de los hogares resultara menor que lo que era demográficamente razonable esperar. Lo anterior dio lugar a una nueva ponderación de las personas, de modo que el total poblacional producido por el MCS coincidiera con las proyecciones demográficas del Consejo Nacional de Población (CONAPO) para ese año. Una consecuencia no deseada, resultado de

la calibración por población del MCS 2015 de la misma manera en que se hacía en los proyectos Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)-MCS anteriores, fue que el número de hogares entre los levantamientos del 2014 y 2015 creciera 3.3 por ciento.¹ Para aquellas encuestas, cuya unidad de análisis es el individuo, el ajuste al alza de los factores de expansión no había tenido consecuencias. No obstante, el MCS

¹ El crecimiento en el número de hogares entre el 2014 y 2015 es significativo: los intervalos de confianza estimados son (31 918 332, 32 382 468) y (32 997 771, 33 438 303) para el 2014 y 2015, respectivamente.

incluyó, además, al hogar como unidad de análisis, por lo que dicho crecimiento no podía ser aceptado. Cabe hacer notar, sin embargo, que la práctica de ajustar del mismo modo los factores de expansión en anteriores levantamientos del MCS había estado ya presente; por ejemplo, en el del 2014, el Módulo reportó un total de poco más de 32.1 millones de hogares; por su parte, la Encuesta Intercensal (EIC) 2015, levantada seis meses después, solamente expandió a poco más de 31.9 millones de hogares.

Como quiera que sea, la expansión a más de 33 millones de hogares (según se publicó en julio del 2016) dejó en claro que en ese levantamiento pudieron estar presentes una o más circunstancias alternativas. Entre las múltiples hipótesis planteadas para explicar el comportamiento en apariencia anómalo del ingreso corriente total (ICT) en el levantamiento del 2015 estuvo, precisamente, el de la alteración de la estructura demográfica de los hogares en relación con levantamientos anteriores. Es esta consideración la que da lugar al ejercicio que se reporta en este trabajo.

2. Metodología

2.1 Su aplicación

El procedimiento elegido para corregir las estructuras demográficas del MCS es el conocido como posestratificación. La motivación detrás de dicho procedimiento está dada por el hecho de que, al referirse a muestras aleatorias de poblaciones finitas, la aleatoriedad presente da lugar a desviaciones, más o menos grandes, de las estructuras demográficas en la muestra de, por ejemplo, las obtenidas mediante ejercicios censales recientes. Cuando los resultados de la encuesta están condicionados por las estructuras demográficas captadas en la muestra, se busca acercarlas a las poblacionales mediante la modificación de los factores de expansión, bajo el supuesto de que las respuestas recogidas (por ejemplo: número de miembros del hogar) son precisas. De este modo,

los factores asociados a poblaciones subrepresentadas en la muestra ven incrementado su valor; en tanto que, los de las poblaciones sobrerrepresentadas lo verán reducido.

Para el caso que nos ocupa, en vista de que, como ya se mencionó, el hogar es también unidad de estudio del MCS, no podríamos simplemente modificar las ponderaciones de los individuos, pues ello daría lugar a más de una nueva ponderación para el hogar. Del mismo modo, el uso del promedio de los factores asociados a cada uno de los integrantes podría llevar a la cancelación del efecto deseado. Por lo anterior, se decidió llevar a cabo una posestratificación de hogares en lugar de individuos. Más aún, la publicación reciente de los resultados de la EIC 2015 dio lugar a tabulados al nivel de los hogares. Ya que el tamaño promedio de los hogares debía ser corregido, se hizo uso de tabulados que hicieran referencia a esta variable. Sin embargo, se añadieron dos más (sexo y edad del jefe del hogar) con el fin de aumentar la consistencia de los resultados.

2.2 Nota técnica

- Descripción del software y paquetes usados para la obtención de los resultados. Se hizo uso del paquete Stata para hacer los ajustes de los factores de expansión del Módulo. La instrucción utilizada se ejemplifica enseguida: **svyset** upm [pweight = factor], strata(est_socio) fpc(nh_1) poststrata(pstpst) postweight(pstwt) vce(linearized) singleunit(certainty) || folioviv

En esta instrucción, la variable *factor* se refiere a los factores de expansión del MCS 2015; *est_socio*, a los estratos sociodemográficos en los que se clasifican las unidades primarias de muestreo; la variable *nh_1*, a los tamaños de dichos estratos; *pstpst*, al identificador de cada uno de los 156 grupos formados a partir de las combinaciones dadas por el sexo y el grupo de edad del jefe del hogar, así como por el tamaño del mismo, en el que se ubica cada hogar de acuerdo con

las variables utilizadas para posestratificar; la variable *pstwt* representa al tamaño de cada uno de esos grupos.

- Fuentes utilizadas. La Encuesta Intercensal 2015 aporta resultados sobre estructuras relevantes a nuestro problema. La tabla 1 muestra la información que da lugar a las variables *pstspst* y *pstwt* incluidas en la definición de la posestratificación.

3. Resultados

3.1 Resultados a nivel nacional, por entidad federativa y por decil con la nueva base del MCS 2015

La tabla 2 presenta el resultado de realizar la misma posestratificación en las bases de datos correspondientes al 2014 y 2015. Queda de manifiesto,

Tabla 1

Estimadores de los hogares y su distribución porcentual según número de integrantes por tamaño de localidad, sexo y grupos quinquenales de edad del jefe o la jefa

Sexo del jefe o la jefa	Grupos quinquenales de edad del jefe o la jefa	Hogares	Número de integrantes					
			1	2	3	4	5	6 y más
Hombres	Total	22 683 498	7.79	15.56	19.44	24.81	16.95	15.46
Hombres	12-19 años	121 883	18.20	30.78	35.41	9.35	3.52	2.75
Hombres	20-24 años	926 257	9.99	19.45	39.05	21.47	6.50	3.54
Hombres	25-29 años	1 810 935	7.35	14.22	29.30	30.11	13.02	6.00
...								
Hombres	60-64 años	1 537 682	8.86	25.28	20.52	16.47	11.35	17.52
Hombres	65-69 años	1 185 559	10.60	30.96	19.46	13.62	9.59	15.77
Hombres	70-74 años	871 214	12.23	34.60	18.23	11.82	8.68	14.43
Hombres	75 años y más	1 214 213	15.74	35.24	17.53	10.93	7.75	12.82
Mujeres	Total	9 266 211	15.69	22.57	21.58	17.28	10.67	12.21
Mujeres	12-19 años	58 955	18.34	33.50	30.63	10.05	4.49	2.99
Mujeres	20-24 años	295 162	13.34	29.01	30.02	17.28	6.15	4.20
Mujeres	25-29 años	497 149	10.91	23.24	26.93	22.13	10.59	6.20
...								
Mujeres	60-64 años	783 667	21.73	25.85	17.43	12.80	8.83	13.36
Mujeres	65-69 años	671 402	26.40	25.93	15.90	11.21	8.24	12.32
Mujeres	70-74 años	552 064	29.87	25.72	15.22	10.53	7.73	10.93
Mujeres	75 años y más	929 762	33.29	26.24	15.08	9.42	6.68	9.29

Fuente: INEGI. Encuesta Intercensal 2015. Tabla Hogares 2, consultada en <http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/intercensal/>

de esta manera, que el levantamiento del 2014 dio lugar a una muestra mucho más consistente con la EIC 2015. Las primeras tres columnas de datos muestran los resultados obtenidos directamente de las bases de datos de cada año; las últimas presentan el efecto de la posestratificación. A nivel nacional, para el 2014, la posestratificación modifica en solamente dos el total nacional de hogares y en poco menos de 30 mil el de la población. Para ese mismo año, los tamaños promedio de los hogares tanto a nivel nacional como por estrato socioeconómico muestran modificaciones menores. En contraste, los cambios que sufre la base de datos en relación con las cantidades anteriores para el 2015 son de destacar: el número total de los hogares pasa de 33 218 037 a solo 32 567 268; sin embargo, el de las personas crece en casi 30 mil. De este modo, el tamaño promedio de los hogares, tanto a nivel nacional como por estrato socioeconómico, mostraría una mayor consistencia con los resultados obtenidos para el 2014. Bajo las anteriores condiciones, tanto el crecimiento poblacional como el del número de hogares entre el 2014 y 2015 se encontrarían cercanos a 1.3%, lo que contrasta con las publicaciones originales de resultados para las que los mencionados incrementos son, respectivamente, de 1.2 y 3.3 por ciento.

La figura 1 exhibe las discrepancias entre las estructuras relativas por grupo de edad y sexo del jefe del hogar que resultan de la base de datos original del MCS 2015 y de la EIC 2015, en las columnas primera y segunda, respectivamente, de cada grupo de edad. Por su parte, las columnas segunda y tercera exhiben la coincidencia entre la EIC y el propio MCS 2015 posestratificado. Este resultado no debe sorprendernos ya que, como se indicó con anterioridad, entre los elementos usados en la posestratificación están precisamente la edad y el sexo del jefe del hogar.

Algo semejante ocurre con la información presentada en la figura 2, ya que se refiere a otra de las variables utilizadas en el ejercicio de posestratificación; ésta ratifica lo que ya se había comentado con anterioridad en relación con la priorización en el MCS 2015 de los hogares de menor tamaño por encima de aquellos con un gran número de integrantes. De aquí se desprende el menor tamaño promedio de los hogares y la consecuente problemática que hemos tratado de atender.

Queda claro que, al menos en términos de las variables demográficas comentadas hasta ahora,

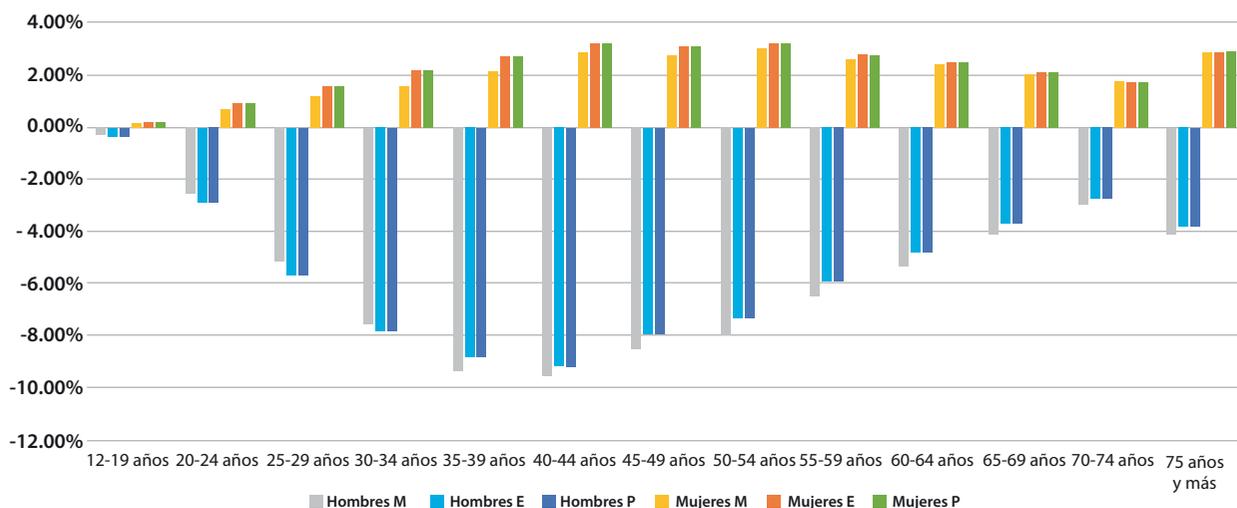
Tabla 2

Tamaño promedio del hogar por estrato socioeconómico

ENIGH-MCS 2014						
Estrato socioeconómico	Hogares	Población	Tamaño	hogs_post	pob_post	tam_hog
Bajo	5 837 414	23 889 861	4.09	5 785 519	23 694 703	4.10
Medio bajo	15 924 128	60 786 821	3.82	16 003 094	61 105 142	3.82
Medio alto	7 133 853	24 819 266	3.48	7 128 748	24 805 904	3.48
Alto	3 255 005	10 401 704	3.20	3 233 037	10 262 665	3.17
Total	32 150 400	119 897 652	3.73	32 150 398	119 868 414	3.73
MCS 2015						
Estrato socioeconómico	Hogares	Población	Tamaño	hogs_post	pob_post	tam_hog
Bajo	6 038 081	24 017 970	3.98	5 936 579	24 182 833	4.07
Medio bajo	16 588 330	62 122 995	3.74	16 374 533	62 580 419	3.82
Medio alto	7 288 686	24 673 826	3.39	7 074 247	24 400 733	3.45
Alto	3 302 940	10 566 939	3.20	3 181 909	10 254 259	3.22
Total	33 218 037	121 381 730	3.65	32 567 268	121 418 244	3.73

Figura 1

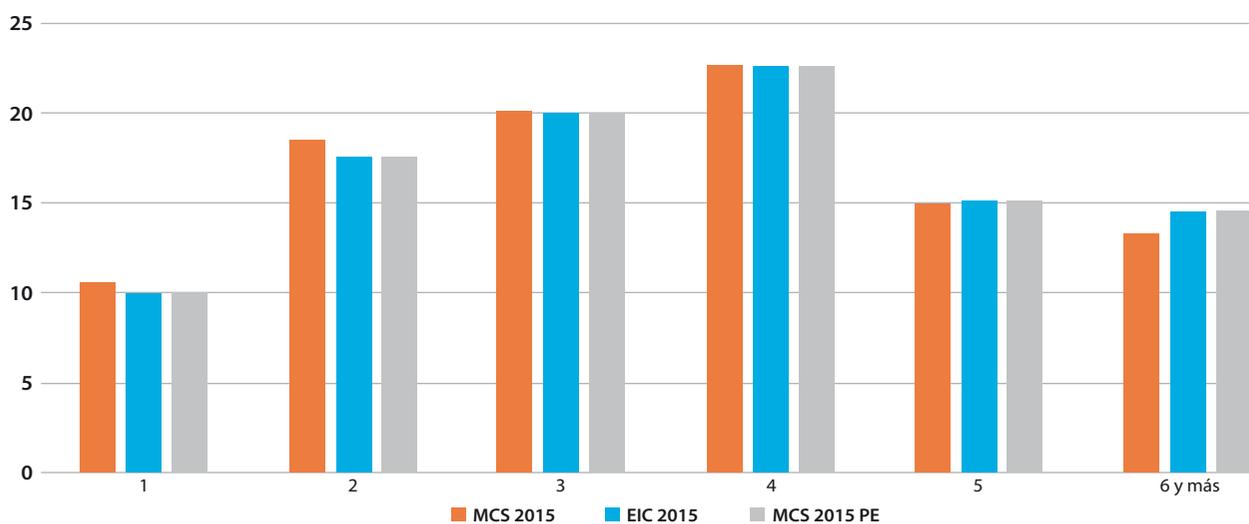
Efecto de la posestratificación en las estructuras relativas por edad y sexo del jefe del hogar, 2015



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

Figura 2

Efecto de la posestratificación en los tamaños del hogar, 2015



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

el ejercicio de posestratificación tiene un efecto favorable en los mayores niveles de agregación. Con el fin de ratificar o rectificar la anterior apreciación, enseguida se muestran resultados a nivel estatal por lo que toca a la distribución espacial de la población, así como las estructuras por edad y sexo de los jefes de los hogares en cada una de las entidades federativas, así como su comparación con los resultados publicados para esos mismos niveles por la EIC 2015.

La figura 3 exhibe la manera en que la posestratificación da lugar a un mayor promedio del número de integrantes del hogar que el que dio como resultado la base de datos original del MCS 2015. Asimismo, muestra las correcciones a las que da lugar la posestratificación, así como los efectos de la misma; además, se les compara con los levantamientos del 2014 y 2015 del MCS y con otro ejercicio de corrección demográfica realizado por el maestro Benito Romo, también del área de investigación del INEGI.

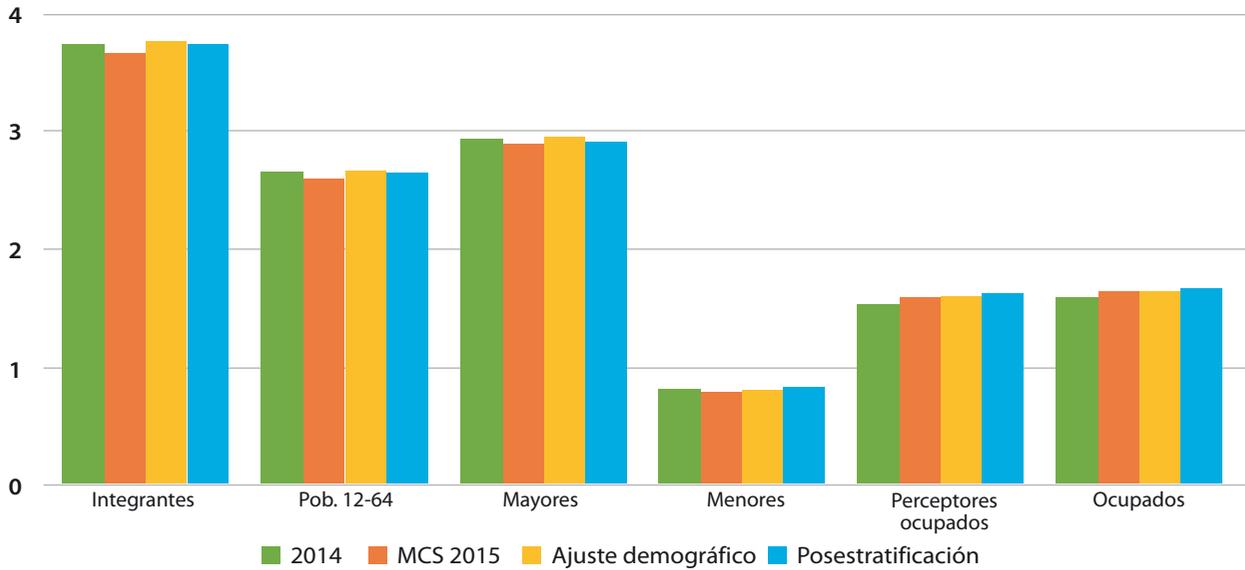
Por su parte, la figura 4 exhibe, en términos relativos debido a los diferentes tamaños poblacionales de cada momento, los tamaños de los totales de hogares por entidad federativa para el 2015. La consistencia es evidente, aunque se presentan variaciones de distintas magnitudes. Cabe recordar que la comparación tiene lugar entre diferentes versiones de un ejercicio de muestreo aleatorio en poblaciones finitas por lo que, de

cualquier manera, existe un elemento de incertidumbre alrededor de cada valor.

Por último, la figura 5 presenta la comparación de estructuras por edad y sexo para el total de la población. Los proyectos que se comparan son la EIC y el MCS posestratificado, ambos del 2015. De nuevo, aunque estas variables no fueron incluidas como condiciones en la posestratificación, se

Figura 3

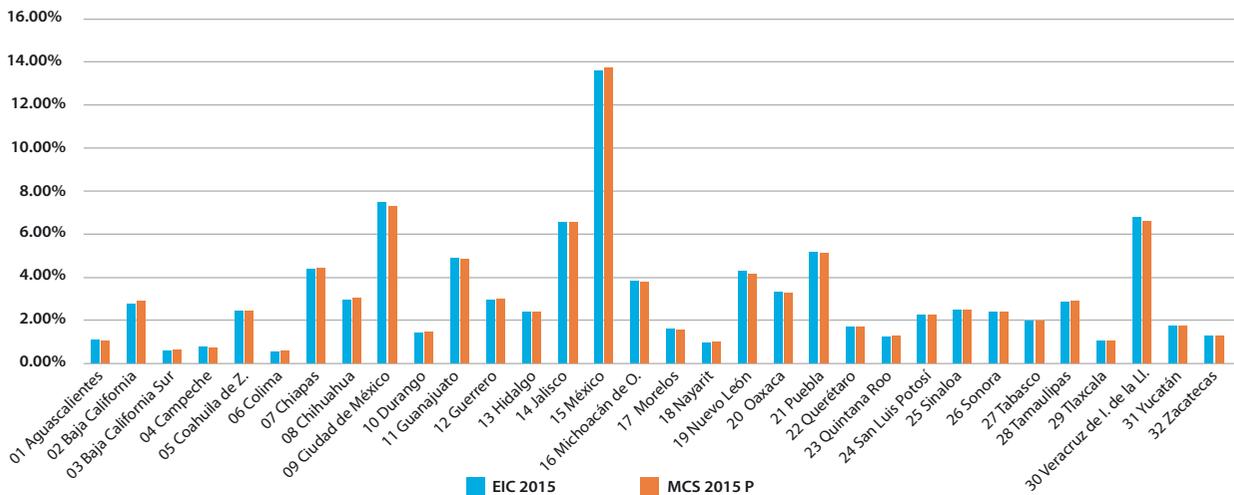
Efecto de la posestratificación en otras variables demográficas



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

Figura 4

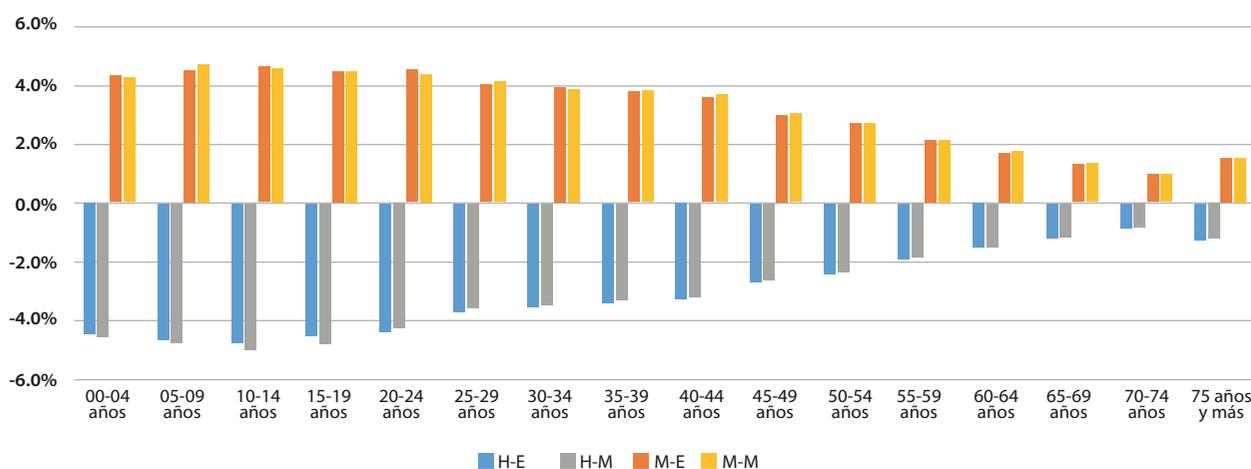
Tamaños relativos de la población, 2015



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

Figura 5

Estructuras relativas por edad y sexo, 2015



Fuentes: INEGI. Encuesta Intercensal 2015. // INEGI. Módulo de Condiciones Socioeconómicas 2015.

aprecia una coincidencia dentro de márgenes razonables. Las mayores diferencias se aprecian para ambos sexos en los grupos de edades jóvenes.

La tabla 3 muestra el efecto de la posestratificación en los incrementos identificados como relevantes a lo largo de los trabajos de análisis de los resultados del MCS 2015. A saber, el ingreso promedio por perceptor (IPP), el promedio del número de perceptores por

hogar (PH), el número de hogares (H) y, finalmente, el total del ICT. Esta información se presenta con desgloses por estrato socioeconómico, por tamaño de localidad y por decil de ingresos. En todos los casos, la posestratificación da lugar a un menor incremento 2014-2015 para el ICT. Lo mismo ocurre para el IPP y para H. La única excepción a lo anterior se da para el número promedio de perceptores por hogar, para el cual el incremento es mayor en todos los casos.

Tabla 3

Continúa

Efecto de posestratificar en el desglose de incrementos del ingreso corriente total, 2014-2015

	MCS 2014-2015 originales				MCS 2014 original/MCS-2015 posestratificado			
	IPP	PH	H	ICT	IPP	PH	H	ICT
Nacional	107.7%	103.9%	103.3%	115.6%	106.9%	105.4%	101.3%	114.1%
Por estrato socioeconómico								
Bajo	121.9%	100.5%	103.3%	126.5%	121.4%	102.0%	101.3%	125.4%
Medio bajo	111.5%	104.4%	103.3%	120.3%	110.5%	105.9%	101.3%	118.5%
Medio alto	102.9%	105.4%	103.3%	112.1%	102.0%	106.9%	101.3%	110.4%
Alto	100.2%	105.8%	103.3%	109.5%	97.9%	106.5%	101.3%	105.7%
Por tamaño de localidad								
100 mil o +	104.3%	103.6%	103.3%	111.7%	102.3%	105.0%	101.3%	108.9%
15 mil a 99 999	105.0%	108.3%	103.3%	117.6%	103.6%	110.1%	101.3%	115.5%
2 500 a 14 999	113.9%	103.0%	103.3%	121.2%	112.6%	104.5%	101.3%	119.2%
2 499 o -	121.9%	102.5%	103.3%	129.1%	120.9%	104.0%	101.3%	127.4%

Tabla 3

Concluye

Efecto de posestratificar en el desglose de incrementos del ingreso corriente total, 2014-2015

	MCS 2014-2015 originales				MCS 2014 original/MCS-2015 posestratificado			
	IPP	PH	H	ICT	IPP	PH	H	ICT
Por decil de ingreso								
I	130.7%	101.7%	103.3%	137.4%	130.1%	103.5%	101.3%	136.4%
II	123.7%	97.6%	103.3%	124.8%	123.5%	99.1%	101.3%	123.9%
III	115.5%	102.0%	103.3%	121.6%	115.2%	103.2%	101.3%	120.4%
IV	114.7%	102.3%	103.3%	121.2%	114.4%	103.4%	101.3%	119.8%
V	111.3%	104.5%	103.3%	120.2%	110.7%	105.8%	101.3%	118.6%
VI	112.3%	103.4%	103.3%	119.9%	111.5%	104.6%	101.3%	118.1%
VII	109.8%	105.3%	103.3%	119.4%	108.6%	106.7%	101.3%	117.4%
VIII	107.8%	105.0%	103.3%	117.0%	106.5%	106.6%	101.3%	114.9%
IX	103.8%	106.9%	103.3%	114.7%	102.1%	108.5%	101.3%	112.3%
X	100.3%	107.9%	103.3%	111.8%	97.0%	110.1%	101.3%	108.2%

4. Validación y evaluación de la metodología

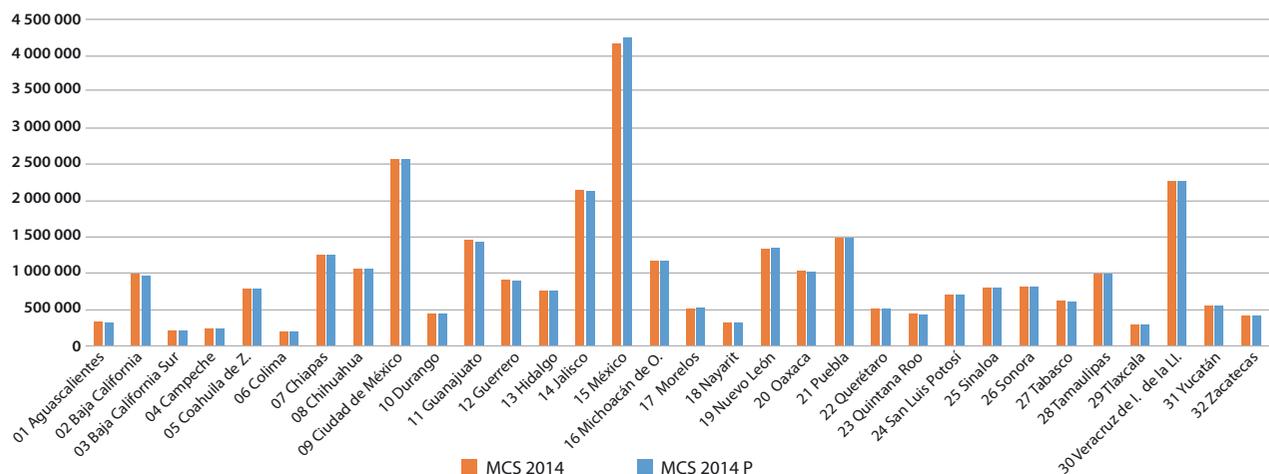
Como se indicó en el texto, la corrección aportada por la posestratificación para el MCS 2014 es irrelevante, como lo muestra la figura 6. El número de hogares por entidad presenta diferencias pequeñas en todos los casos.

En la figura 7 se muestran los errores relativos, definidos enseguida, correspondientes a las discrepancias mostradas en la figura 6:

$$e_j = \frac{2(Hog_{14p,j} - Hog_{14,j})}{(Hog_{14p,j} + Hog_{14,j})}$$

Figura 6

Número total de hogares por entidad, antes y después de posestratificar, 2014



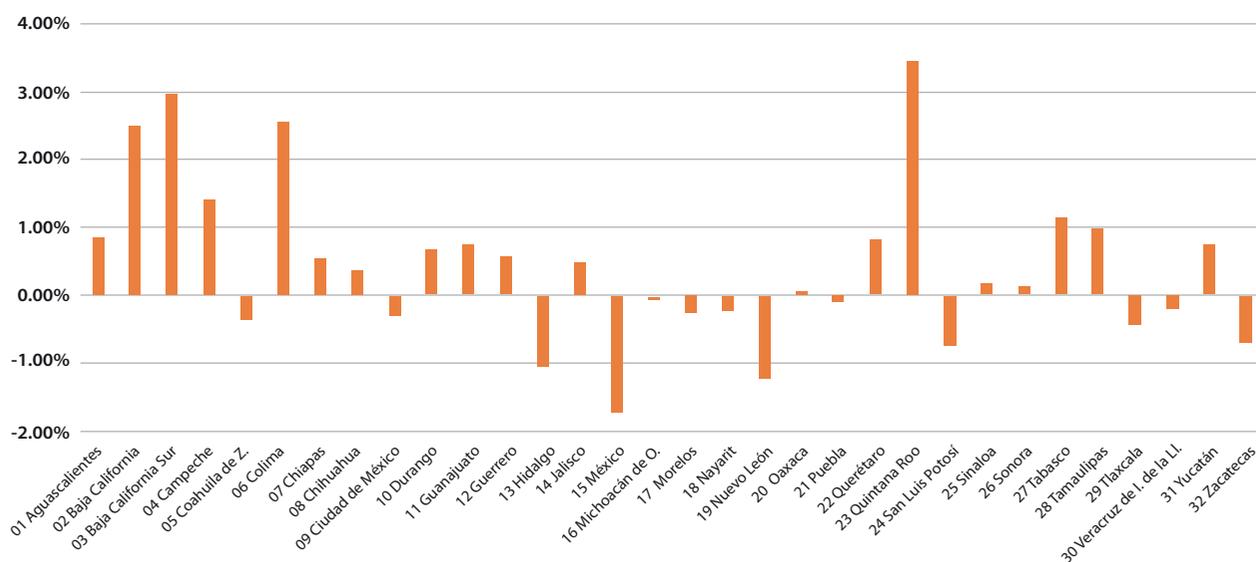
Fuente: MCS 2015 y cálculos propios.

donde $14p$ se refiere a la base de datos posestratificada del 2014 y j , a alguna entidad federativa. Por supuesto, los errores relativos de mayor magnitud se presentan principalmente para los estados más pequeños (Baja California, Baja California Sur, Colima y Quintana Roo; solo para esta entidad dicho error resulta ligeramente mayor a 3%). Aunque menor a 2%, el error de mayor consideración es

el que se presenta para el estado de México. Si se llegara a considerar que el tamaño de los errores cometidos al posestratificar es excesivamente grande, aún es posible repetir el ejercicio, pero entidad por entidad. La información requerida para este fin está contenida en la tabla *Hogares 3* de la *Encuesta Intercensal 2015*, una parte de la cual se reproduce en la tabla 4.

Figura 7

Diferencias relativas entre totales de hogares por entidad, 2014



Fuente: cálculos propios.

Tabla 4

Continúa

Estimadores de los hogares y su distribución porcentual según número de integrantes por entidad federativa, sexo y grupos quinquenales de edad del jefe o la jefa

Entidad federativa	Sexo del jefe o la jefa	Grupos quinquenales de edad del jefe o la jefa	Hogares	Número de integrantes					
				1	2	3	4	5	6 y más
01 Aguascalientes	Hombres	Total	247 011	6.35	13.03	17.78	24.89	20.02	17.93
01 Aguascalientes	Hombres	12-19 años	1 187	21.40	39.85	26.62	7.16	3.45	1.52
01 Aguascalientes	Hombres	20-24 años	10 337	6.34	18.79	41.87	25.30	4.29	3.42
...									

Tabla 4

Concluye

Estimadores de los hogares y su distribución porcentual según número de integrantes por entidad federativa, sexo y grupos quinquenales de edad del jefe o la jefa

Entidad federativa	Sexo del jefe o la jefa	Grupos quinquenales de edad del jefe o la jefa	Hogares	Número de integrantes					
				1	2	3	4	5	6 y más
01 Aguascalientes	Hombres	75 años y más	9 969	13.27	39.34	17.08	12.73	5.44	12.14
01 Aguascalientes	Hombres	No especificado	127	2.36	0.00	36.22	11.02	40.94	9.45
01 Aguascalientes	Mujeres	Total	87 578	15.51	21.38	22.58	17.02	11.07	12.43
01 Aguascalientes	Mujeres	12-19 años	479	13.36	38.83	43.01	2.92	0.84	1.04
01 Aguascalientes	Mujeres	20-24 años	2 503	11.71	26.89	30.76	20.62	4.04	5.99
...									
01 Aguascalientes	Mujeres	65-69 años	5 684	26.76	26.81	14.39	10.61	8.71	12.72
01 Aguascalientes	Mujeres	70-74 años	4 728	30.52	25.25	20.60	8.38	5.84	9.41
01 Aguascalientes	Mujeres	75 años y más	8 551	35.59	20.77	17.38	8.12	8.71	9.44
01 Aguascalientes	Mujeres	No especificado	87	17.24	9.20	64.37	9.20	0.00	0.00
02 Baja California	Hombres	Total	666 287	13.48	16.64	20.43	23.72	15.66	10.06
02 Baja California	Hombres	12-19 años	3 502	23.67	30.61	32.92	7.45	1.40	3.94
...									

Fuente: INEGI. Encuesta Intercensal 2015. Extracto de la tabla Hogares 3.

5. Conclusión y comentarios finales

La posestratificación corrige deficiencias en la demografía de los hogares. Sin embargo, a pesar de que su efecto en algunas variables fue menor para el levantamiento del 2014, el cual fue realizado para *emparejar el piso* por así decirlo, pensando en comparaciones que no favorezcan a ninguno de los ejercicios (ver tabla 5). Los efectos

son reducidos cuando ambas bases de datos son posestratificadas.

Más aún, tiene un efecto en la dirección correcta sobre variables asociadas al ingreso, aunque de magnitudes relativamente pequeñas.

Por otro lado, variables como el número promedio de perceptores por hogar toman valores mayores con respecto al MCS 2015 original.

Tabla 5

Continúa

Efecto de posestratificar en el desglose de incrementos del ingreso corriente total, 2014-2015

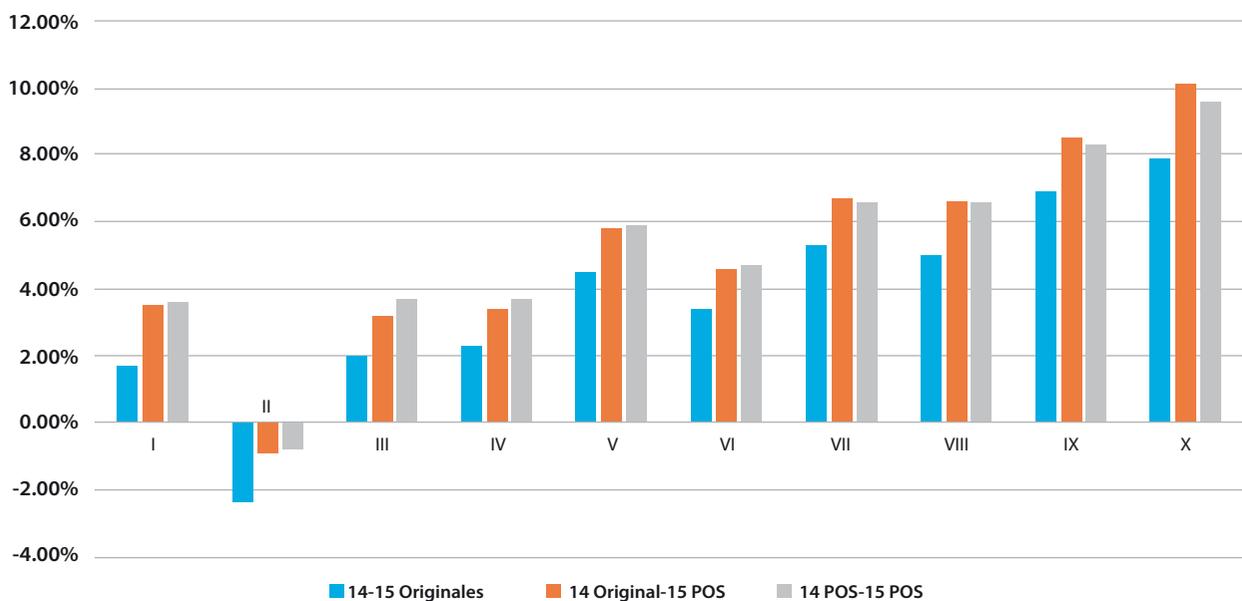
	MCS 2014 original/MCS-2015 posestratificado				MCS 2014-2015, ambos posestratificados			
	IPP	PH	H	ICT	IPP	PH	H	ICT
Nacional	106.0%	105.4%	101.3%	113.1%	106.9%	105.4%	101.3%	114.1%
Por estrato socioeconómico								
Bajo	121.4%	102.0%	101.3%	125.4%	120.9%	102.3%	101.3%	125.3%
Medio bajo	110.5%	105.9%	101.3%	118.5%	111.0%	106.0%	101.3%	119.2%

Efecto de posestratificar en el desglose de incrementos del ingreso corriente total, 2014-2015

	MCS 2014 original/MCS-2015 posestratificado				MCS 2014-2015, ambos posestratificados			
	IPP	PH	H	ICT	IPP	PH	H	ICT
Medio alto	102.0%	106.9%	101.3%	110.4%	103.1%	106.5%	101.3%	111.3%
Alto	97.9%	106.5%	101.3%	105.7%	99.7%	106.4%	101.3%	107.5%
Por tamaño de localidad								
100 mil o +	102.3%	105.0%	101.3%	108.9%	103.9%	104.9%	101.3%	110.4%
14 999 a 99 999	103.6%	110.1%	101.3%	115.5%	104.5%	109.9%	101.3%	116.3%
2 500 a 49 999	112.6%	104.5%	101.3%	119.2%	113.7%	104.4%	101.3%	120.2%
2 499 o -	120.9%	104.0%	101.3%	127.4%	120.2%	104.6%	101.3%	127.4%
Por decil de ingreso								
I	130.1%	103.5%	101.3%	136.4%	130.1%	103.6%	101.3%	136.5%
II	123.5%	99.1%	101.3%	123.9%	123.3%	99.2%	101.3%	123.9%
III	115.2%	103.2%	101.3%	120.4%	114.6%	103.7%	101.3%	120.4%
IV	114.4%	103.4%	101.3%	119.8%	114.0%	103.7%	101.3%	119.8%
V	110.7%	105.8%	101.3%	118.6%	110.5%	105.9%	101.3%	118.6%
VI	111.5%	104.6%	101.3%	118.1%	111.4%	104.7%	101.3%	118.1%
VII	108.6%	106.7%	101.3%	117.4%	108.7%	106.6%	101.3%	117.5%
VIII	106.5%	106.6%	101.3%	114.9%	106.5%	106.6%	101.3%	114.9%
IX	102.1%	108.5%	101.3%	112.3%	102.4%	108.3%	101.3%	112.4%
X	97.0%	110.1%	101.3%	108.2%	98.4%	109.6%	101.3%	109.2%

Figura 8

Incrementos en el promedio del número de perceptores por hogar, 2014-2015



Fuente: cálculos propios.

6. Anexo: pobreza multidimensional y carencias

Por lo que toca a temas de pobreza multidimensional, el efecto de posestratificar la base de datos del MCS 2015 es relativamente pequeño. Las tablas 6 a 9 se refieren a la pobreza multidimensional y a las carencias, de interés del CONEVAL, respectivamente. En ellas se presentan los valores de algunas estadísticas acerca de las diferencias entre los por-

centajes que se obtienen entre las bases de datos del MCS 2015 original y la posestratificada. Es de observarse que, salvo por un par de casos, todos estos valores se encuentran por debajo de un punto porcentual.

Por lo anterior, es necesario enfatizar que se ha de realizar un análisis a profundidad en relación con el efecto de la posestratificación sobre todas las variables relevantes.

Tabla 6

Pobreza multidimensional. Estadísticas de las diferencias MCS posestratificado vs. original, 2015, medidas entre las entidades federativas

Medida	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Mínimo	-0.01	-0.18	-0.14	-0.59	-0.12	-0.94
Máximo	0.83	0.57	0.59	0.35	0.20	-0.23
Media	0.46	0.35	0.10	0.00	0.02	-0.48
Desviación estándar	0.20	0.18	0.15	0.21	0.08	0.19
Mediana	0.45	0.37	0.08	0.02	0.01	-0.45

Fuente: cálculos propios a partir de la base de datos posestratificada MCS 2015

Tabla 7

Población con carencias. Estadísticas de las diferencias MCS posestratificado vs. original, 2015, medidas entre las entidades federativas

Medida	Población con al menos una carencia	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Mínimo	0.13	-0.07	-0.82	-0.11	0.14	0.17	-0.19	-0.03
Máximo	0.95	0.74	0.13	0.36	1.34	1.23	0.61	0.66
Media	0.45	0.35	-0.32	0.14	0.72	0.54	0.12	0.32
Desviación estándar	0.19	0.19	0.24	0.11	0.26	0.27	0.19	0.18
Mediana	0.44	0.34	-0.26	0.13	0.73	0.52	0.07	0.27

Fuente: cálculos propios a partir de la base de datos posestratificada MCS 2015.

Pobreza multidimensional de acuerdo con la base de datos posestratificada del MCS 2015

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Aguascalientes	26.05%	24.50%	1.54%	33.28%	7.82%	32.85%
Baja California	24.18%	22.46%	1.72%	39.02%	6.01%	30.80%
Baja California Sur	22.82%	21.15%	1.67%	40.46%	4.56%	32.15%
Campeche	34.68%	30.16%	4.51%	39.90%	2.90%	22.53%
Coahuila de Z.	19.64%	18.15%	1.49%	30.76%	10.61%	38.99%
Colima	26.49%	24.58%	1.92%	41.12%	4.95%	27.43%
Chiapas	65.95%	40.04%	25.91%	23.89%	1.99%	8.18%
Chihuahua	21.48%	20.61%	0.87%	34.56%	8.21%	35.76%
Cd. de México	26.07%	25.13%	0.94%	29.18%	7.05%	37.70%
Durango	32.97%	30.63%	2.34%	34.75%	7.75%	24.52%
Guanajuato	30.36%	28.66%	1.70%	44.04%	4.65%	20.95%
Guerrero	60.18%	41.81%	18.36%	31.93%	1.84%	6.05%
Hidalgo	41.83%	34.42%	7.42%	39.44%	3.12%	15.61%
Jalisco	25.53%	22.97%	2.56%	38.62%	5.24%	30.61%
México	41.37%	36.89%	4.48%	27.68%	8.77%	22.18%
Michoacán de O.	45.13%	36.22%	8.91%	38.30%	2.99%	13.58%
Morelos	37.51%	31.78%	5.73%	38.35%	4.00%	20.15%
Nayarit	32.07%	27.66%	4.41%	38.58%	5.47%	23.87%
Nuevo León	17.88%	16.89%	0.99%	30.38%	8.14%	43.60%
Oaxaca	60.09%	41.00%	19.10%	28.42%	1.59%	9.89%
Puebla	50.59%	43.22%	7.36%	32.72%	4.33%	12.37%
Querétaro	24.74%	22.22%	2.52%	40.58%	4.97%	29.71%
Quintana Roo	30.38%	25.44%	4.94%	39.82%	4.79%	25.01%
San Luis Potosí	36.54%	30.65%	5.89%	34.49%	5.08%	23.89%
Sinaloa	28.12%	26.27%	1.85%	40.67%	4.93%	26.29%

Tabla 8

Concluye

Pobreza multidimensional de acuerdo con la base de datos posestratificada del MCS 2015

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Sonora	22.81%	20.69%	2.12%	37.06%	5.70%	34.42%
Tabasco	36.89%	31.27%	5.62%	50.58%	1.61%	10.92%
Tamaulipas	32.32%	28.98%	3.34%	31.79%	6.89%	29.00%
Tlaxcala	43.97%	41.17%	2.80%	33.46%	6.37%	16.20%
Veracruz de I. de la Ll.	51.01%	40.25%	10.75%	33.37%	2.80%	12.82%
Yucatán	34.90%	29.07%	5.83%	36.89%	4.76%	23.44%
Zacatecas	45.08%	40.39%	4.69%	30.45%	4.99%	19.48%
Total	37.19%	31.19%	6.00%	33.95%	5.47%	23.38%

Tabla 9

Continúa

Población con carencias de acuerdo con la base de datos posestratificada del MCS 2015

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Indicador de carencia por rezago educativo	Indicador de carencia por acceso a servicios de salud	Indicador de carencia por acceso a la seguridad social	Indicador de carencia por calidad y espacios de la vivienda	Indicador de carencia de acceso a servicios básicos de la vivienda	Indicador de carencia por acceso a la alimentación
Aguascalientes	59.32%	8.77%	13.73%	13.67%	43.36%	5.13%	4.63%	19.44%
Baja California	63.19%	12.73%	13.55%	18.19%	46.62%	9.68%	12.63%	15.17%
Baja California Sur	63.28%	14.81%	13.99%	12.67%	43.15%	15.10%	11.63%	21.91%
Campeche	74.57%	27.84%	19.90%	12.59%	58.32%	18.90%	34.17%	23.43%
Coahuila de Z.	50.41%	7.89%	12.12%	14.01%	31.49%	5.85%	4.63%	16.92%
Colima	67.61%	14.14%	17.80%	13.67%	52.88%	9.53%	9.77%	19.77%
Chiapas	89.84%	47.42%	29.52%	18.89%	81.73%	29.72%	54.55%	25.25%
Chihuahua	56.04%	7.48%	15.53%	12.10%	39.92%	5.22%	4.21%	13.13%
Cd. de México	55.25%	6.04%	8.86%	19.20%	45.85%	5.24%	1.77%	10.50%
Durango	67.72%	11.36%	15.45%	15.46%	51.19%	5.96%	11.87%	17.99%
Guanajuato	74.40%	16.95%	20.06%	14.40%	57.73%	9.28%	12.99%	26.57%

Población con carencias de acuerdo con la base de datos posestratificada del MCS 2015

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Indicador de carencia por rezago educativo	Indicador de carencia por acceso a servicios de salud	Indicador de carencia por acceso a la seguridad social	Indicador de carencia por calidad y espacios de la vivienda	Indicador de carencia de acceso a servicios básicos de la vivienda	Indicador de carencia por acceso a la alimentación
Guerrero	92.10%	50.84%	24.71%	16.55%	77.55%	34.67%	59.77%	40.87%
Hidalgo	81.27%	21.60%	17.15%	18.44%	69.99%	9.99%	24.76%	21.54%
Jalisco	64.15%	13.35%	18.74%	17.14%	49.91%	8.09%	9.59%	15.69%
México	69.05%	16.51%	13.14%	20.13%	58.44%	10.52%	10.32%	20.55%
Michoacán de O.	83.43%	33.36%	26.49%	23.29%	71.02%	16.82%	26.41%	32.55%
Morelos	75.86%	22.41%	16.95%	16.18%	62.16%	14.65%	22.77%	26.52%
Nayarit	70.65%	15.75%	15.66%	16.33%	57.48%	7.83%	16.74%	19.97%
Nuevo León	48.26%	6.81%	11.42%	12.54%	31.64%	5.11%	4.24%	15.64%
Oaxaca	88.52%	46.08%	27.10%	17.01%	77.77%	21.39%	59.54%	32.84%
Puebla	83.31%	27.12%	21.58%	19.17%	72.08%	14.25%	27.00%	24.08%
Querétaro	65.32%	15.36%	15.05%	12.91%	51.64%	9.37%	18.00%	18.00%
Quintana Roo	70.20%	21.33%	15.13%	16.76%	52.51%	18.66%	18.81%	22.39%
San Luis Potosí	71.03%	20.86%	17.05%	9.94%	56.82%	14.26%	29.73%	16.91%
Sinaloa	68.79%	14.66%	15.25%	14.81%	50.45%	9.09%	14.58%	23.96%
Sonora	59.88%	12.24%	12.46%	13.34%	38.36%	9.34%	10.06%	24.00%
Tabasco	87.47%	35.51%	17.32%	17.21%	71.07%	12.32%	46.19%	39.98%
Tamaulipas	64.10%	12.81%	15.76%	15.35%	46.80%	9.20%	10.30%	19.88%
Tlaxcala	77.43%	15.16%	13.58%	14.99%	67.50%	11.34%	11.02%	21.64%
Veracruz de I. de la Ll.	84.38%	36.30%	25.54%	19.42%	69.30%	20.20%	41.33%	30.74%
Yucatán	71.80%	28.15%	21.49%	15.02%	53.48%	19.37%	39.13%	17.20%
Zacatecas	75.53%	13.32%	18.42%	12.46%	62.43%	4.48%	14.49%	20.30%
Total	71.14%	20.95%	17.68%	17.02%	57.36%	12.58%	20.67%	22.13%

Ajuste demográfico por imputación

Demographic adjustment **by imputation**

Benito Durán Romo*

Un aumento inusual en el número de hogares reportado por el Módulo de Condiciones Socioeconómicas (MCS) 2015 se explica en parte por la contracción del tamaño del hogar y el consiguiente ajuste de factores de expansión para aumentar el volumen de la población con el objetivo de acercarse a las proyecciones del Consejo Nacional de Población (CONAPO). Esto también tuvo un efecto: aumentó el volumen de hogares, pero en este caso superó el número proyectado por el CONAPO (32 298 321, por debajo de los 33 218 037 reportados por el MCS 2015). Si en el MCS 2015 la contracción se debió a una omisión poblacional dentro del hogar durante la fase de encuestado, puede restaurarse el tamaño de los hogares con un cuidadoso proceso de imputación automática de individuos en ciertos hogares con base en la técnica

An unusual increase in the number of households reported by the Socio-Economic Conditions Module 2015 (MCS 2015 by its Spanish acronym) is explained in part by the contraction of the household size and the consequent adjustment of weights to increase the volume of the population with the aim of approaching to the projections of National Council of Population (CONAPO by its Spanish acronym). This also had an effect: increased the volume of households, but in this case exceeded the number projected by CONAPO (32 298 321, below the 33 218 037 reported by the MCS 2015). If in the MCS 2015 the contraction was due to a population omission within the household during the survey phase, we can restore the size of the households with a careful process of automatic imputation

* Instituto Nacional de Estadística y Geografía, benito.duran@inegi.org.mx

Nota: el autor agradece la participación de Adriana Pérez Amador, Olinca Páez Domínguez y Lilia Guadalupe Luna Ramírez en el proceso de investigación y de integración de este documento.

de imputación *Cold-Deck*. Para lograrlo, se creó un gran grupo (*pool*) de donantes con las personas registradas en los MCS 2012 hasta 2015. Como resultado final, el proceso produce más individuos que en la base de datos original, requiriendo un reajuste de los factores de expansión, de acuerdo con la nueva estructura de población resultante de la imputación. En general, esto significa que los factores de expansión no tienen que ser tan grandes como los originales y, por lo tanto, se obtienen menos hogares. Con este procedimiento de imputación del MCS 2015, el número de hogares se redujo en poco más de 500 mil, con un tamaño del hogar de 3.71 y no 3.65 correspondiente al proceso original. Al comparar el ingreso actual con el MCS 2014, el monto total crece 14.1% (1.53% menos que el original), aunque el promedio por hogar crece 12.25% (0.32% más que el original).

Palabras clave: *Cold-Deck*; imputación; individuos; hogares; encuesta; factores de expansión.

of individuals in certain households, based on the *Cold-Deck* imputation technique. To achieve this, a large pool of donors was created with the people recorded in the MCS 2012 until 2015. As a final result, the process produces more individuals than in the original database, requiring a readjustment of the weights, according to the new structure of the population resulting from imputation. In general, this means that the weights do not have to be as large as the original ones and therefore, fewer households are obtained. With this imputation procedure of MCS 2015, the number of households was reduced by just over 500 thousand, with a household size of 3.71 and not 3.65 corresponding to the original process. When comparing current income with MCS 2014, the total amount grows 14.1% (1.53% less than the original), although the average per household grows 12.25% (0.32% more than the original).

Key words: *Cold-Deck*; imputation; individuals; households; survey; weights.



Guélaguetza festivity in Mexico / Eduardo González/CON/Colaborador/Getty Images

1. Introducción

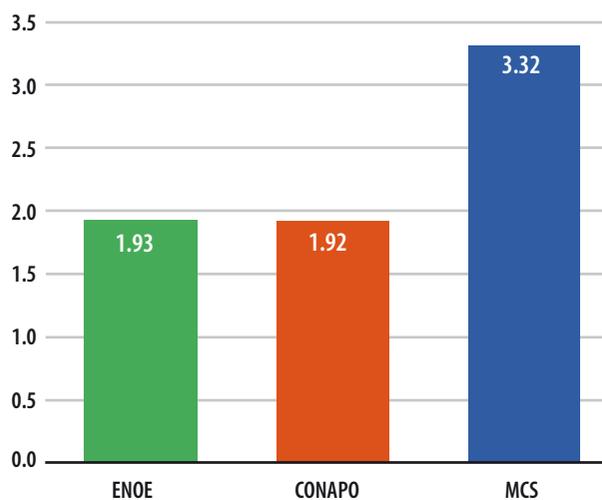
El Módulo de Condiciones Socioeconómicas (MCS) 2015 reportó un total de hogares casi al nivel del proyectado por el Consejo Nacional de Población (CONAPO) para el 2017, un poco más de 900 mil de lo que se tenía previsto para el 2015. Esto se debió a que el tamaño de hogar reportado por el MCS 2015 es menor de lo esperado y a que los factores de expansión de las encuestas son ajustados para llegar a un total de población, dejando de lado el número de hogares.

En la gráfica 1.1 se pueden observar las tasas de crecimiento de los hogares para distintas fuentes. En ésta notamos que mientras algunas encuestas del Instituto Nacional de Estadística y Geografía (INEGI) y las proyecciones del CONAPO reportaron tasas de crecimiento de los hogares menores a 2% entre el 2014 y 2015, el MCS la estimó 3.32% para el mismo periodo, es decir, 1.4% más que la del CONAPO, lo cual es una clara evidencia del crecimiento inusual de los hogares en el MCS 2015.

Ante esto, el objetivo de este ejercicio es incrementar el tamaño del hogar del MCS 2015 y, como consecuencia, acercar el total de hogares a lo proyectado por el CONAPO correspondiente a ese año.

Gráfica 1.1

Tasa de crecimiento de los hogares 2014-2015 según distintas fuentes



Para llevarlo a cabo, se diseñó un procedimiento de imputación de individuos a cierto número de hogares. De esta forma, se incrementa la población muestral del evento y, dado ese efecto, se reajustan los factores de expansión consiguiendo incrementar el tamaño del hogar y disminuir el total de hogares.

Cabe destacar que este proceso de imputación solo pretende hacer un ajuste demográfico al MCS 2015, lo cual no necesariamente se traduce en un ajuste en la distribución del ingreso y otras variables relacionadas con éste.

La sección *Marco teórico* se divide en dos partes, en la primera se presenta una de las posibles explicaciones de la baja en el tamaño del hogar del MCS 2015 y en la segunda parte se muestran algunas definiciones del proceso de imputación y su clasificación. Después, en *Metodología* se detalla el proceso completo de imputación aplicado a este caso, desde la forma de determinar el número de personas a imputar hasta llegar a la imputación como tal. Los resultados del procedimiento de imputación se dan en la sección correspondiente y, por último, se presentan comentarios respecto a la metodología y algunas conclusiones.

2. Marco teórico

2.1 Módulo de Condiciones Socioeconómicas

El MCS surge en el 2008 como el insumo para la medición multidimensional de la pobreza que por mandato de la *Ley General de Desarrollo Social (LGDS)* tiene que hacer el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) y que, a la vez, obliga al INEGI a llevar a cabo los censos, conteos y encuestas necesarios para su medición. Desde ese año y hasta el 2014 los MCS tuvieron una periodicidad bienal (en sintonía con la LGDS, debido a que mandata que la pobreza categorizada por entidad federativa debe medirse como mínimo cada dos años) y acompañaron a la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), de manera que

una parte de la muestra la cubría la propia ENIGH (muestra ENIGH), captando información sobre ingreso y gasto; y otra el MCS (muestra MCS) captando solo información de ingreso. En el 2015 toda la muestra fue cubierta por el operativo del MCS, sin contar con el acompañamiento de la ENIGH. Esto pudo haber influido en que se observaran algunas diferencias en la información con respecto a los eventos anteriores.

Entre las diferencias que se observan en el MCS 2015 respecto a sus antecesores está un cambio en la tasa de crecimiento del total de hogares, que no es consistente con la tendencia observada desde el 2008.

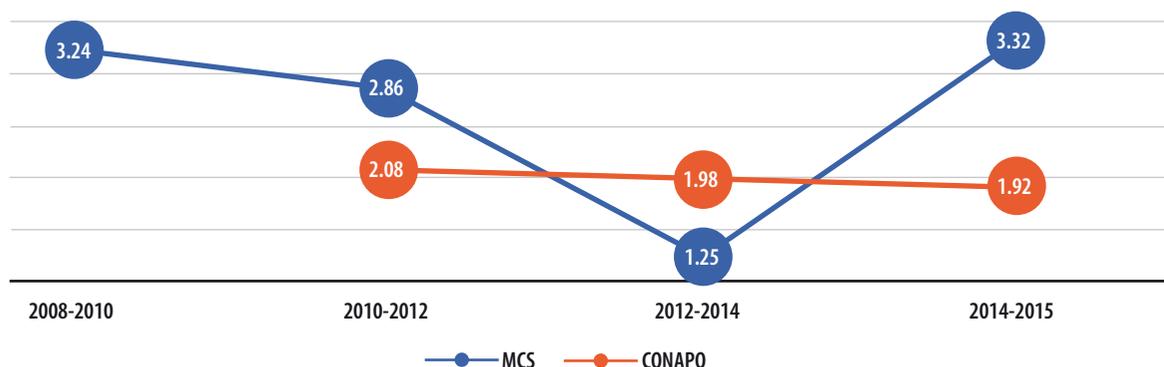
Como se observa en la gráfica 2.1, del 2014 al 2015 la tasa de crecimiento promedio anual de los hogares según los MCS es de 3.3%, mientras que la que resulta de las proyecciones de población del CONAPO es de solo 1.9 por ciento. Este aumento en

el crecimiento del número de hogares se debe, en parte, a que el tamaño del hogar reportado por el MCS 2015 es menor a lo proyectado y, como puede verse en la gráfica 2.2, muestra una ligera caída respecto a lo observado en el 2012 y 2014.

Como sucede en todas las encuestas sociodemográficas en las que la unidad de selección es la vivienda, un hogar en muestra representa a otros hogares con las mismas características socioeconómicas y demográficas. Así, los datos muestrales se expanden, con base en el inverso de su probabilidad de inclusión en una muestra, para referir no únicamente a la unidad muestral, sino a ésta más las que ella representa en la población de la que deriva dicha muestra. Sin embargo, debe considerarse que los factores de expansión de las encuestas sociodemográficas son ajustados con el fin de alcanzar el monto total de población proyectado por el CONAPO, independientemente del resultado en el número total de los hogares.

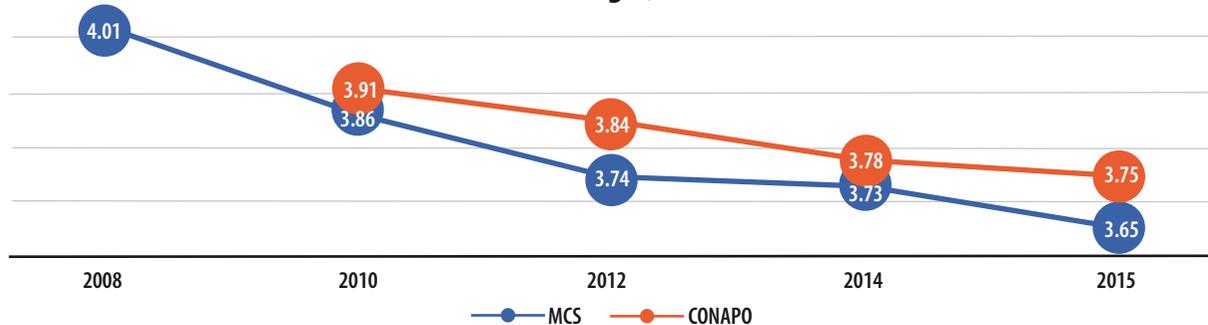
Gráfica 2.1

Tasa de crecimiento media anual de los hogares 2008-2015



Gráfica 2.2

Tamaño del hogar, 2008-2015



Considerando los factores naturales, es decir, los factores de expansión sin la corrección por la no respuesta ni la calibración por proyección de población, se observa que, en comparación con eventos previos, el tamaño promedio del hogar es menor en el MCS 2015 y, por lo tanto, el total de hogares en muestra arroja un menor registro de integrantes. Esto obliga a un ajuste adicional derivado del proceso de ajuste para alcanzar la cifra de población que indica la proyección del CONAPO; de lo contrario, el monto de población quedaría por debajo de lo proyectado para el año del levantamiento. El efecto final después del ajuste es que no solo aumenta el número de personas, sino el volumen de los hogares en los que habitan dichas personas, generando un efecto de mayor crecimiento. Esto significa que un hogar en muestra del MCS 2015 representa más hogares de los que representaría en los levantamientos previos que fueron acompañados con la ENIGH.

La baja tan pronunciada en el tamaño del hogar que arroja el MCS 2015 puede deberse a la omisión de personas, en términos del operativo de campo, debido a que en ocasiones el informante olvida reportarlos en la primera visita, sin embargo, se logra integrarlos en alguna de las visitas posteriores; esto es más viable para el levantamiento de la muestra ENIGH, dado que el hogar es visitado durante siete días consecutivos. En la muestra MCS es más factible que ese olvido persista, debido a que los cuestionarios pueden ser llenados en tan solo tres días y si en ese periodo no se consigue

integrarlos, se pierde su registro. Esto lo podemos constatar en la gráfica 2.3, donde se observa que históricamente la muestra MCS reporta tamaños de hogar más bajos que la muestra ENIGH. Además, dado que el MCS 2015 solo tuvo muestra MCS, el tamaño del hogar sigue su tendencia a la baja. En cierta forma, se puede decir que el tamaño del hogar de la muestra ENIGH jala hacia arriba el tamaño del hogar de toda la muestra, permitiendo que no sea tan bajo. En el MCS 2015 esto no fue posible por la falta de muestra ENIGH.

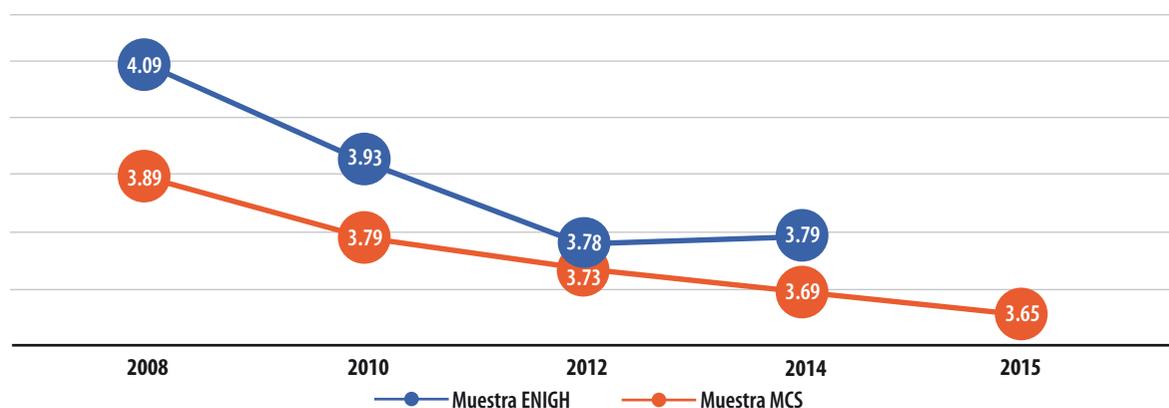
2.2 Métodos de imputación

La mayoría de los sistemas de información requiere que le proporcionen datos (entrada o captura de datos) para su funcionamiento, los cuales necesitan ser validados.

En la entrada, los datos deben cubrir ciertas características para poder ser aceptados como correctos al momento de ser capturados. El proceso que permite verificar que esos datos cumplen con las características necesarias para ser aceptados se le llama validación. Para llevar a cabo este proceso, es necesario, primero, identificar el tipo de variable a tratar, ésta puede ser cualitativa (por ejemplo, el género y el nivel educativo de una persona) o cuantitativa (por ejemplo, la edad e ingreso de la persona); y segundo, construir reglas de validación, las cuales pueden involucrar una o más variables.

Gráfica 2.3

Tamaño del hogar por origen de muestra del MCS, 2008-2015



Por otro lado, no solo los sistemas de información requieren que la entrada de datos sea validada, por ejemplo, el INEGI cada año desarrolla un número importante de proyectos de naturaleza estadística, como censos y encuestas, en los cuales se implementa la validación en varias etapas: comenzando por la captación de los datos, siguiendo con la entrada de éstos a los sistemas de captura y finalizando con la validación de las bases de datos. A medida que el proceso avanza en las etapas, la validación se va haciendo más compleja, sobre todo porque se incrementa el número de variables involucradas en una regla de validación.

Aunado a lo anterior, al llevar a cabo los procesos de validación, pueden detectarse algunas variables con ausencia de información, lo cual puede darse por dos situaciones diferentes; en un primer caso, este escenario puede ser correcto: una persona en la variable *edad* reporta 3 años y el *nivel de instrucción* está vacío; en un segundo caso, esa ausencia puede deberse a una inconsistencia: una persona que reporta en *edad* 15 años y el *nivel de instrucción* está vacío (debería aparecer un código: por lo menos *sin instrucción*). En el segundo caso es necesario imputar (asignar) un dato a la variable *nivel de instrucción* para que pase el proceso de validación.

Al respecto veamos algunas definiciones y/o comentarios.

Palacios y González (2004) nos dice que: "...por imputación entenderemos la asignación de nuevos valores a aquellas variables, en uno o más registros, que necesitan ser corregidos para pasar el proceso de validación".

De acuerdo con Schafer (2006): "...la imputación es la práctica de rellenar datos faltantes con valores probables, es una forma atractiva de analizar datos incompletos. Esto aparentemente resuelve el problema de análisis de datos incompletos. Sin embargo, un nuevo o poco escrupuloso método de imputación podría crear más problemas que los que resuelve, distorsionando estimaciones, errores estándar y pruebas de hipótesis".

Según Juárez (2004): "...las imputaciones son extracciones de una distribución predictiva de los datos faltantes y requiere de un método para crear una distribución predictiva para la imputación basada en datos observados. La imputación es un método general y flexible para manejar problemas de datos faltantes".

Dempster y Rubin (1983) consideran que: "La idea de la imputación es seductora y peligrosa. Seductora porque puede adormecer al usuario en un estado placentero de creer que los datos están completos después de todo y esto es peligroso porque junta situaciones en donde el problema es suficientemente menor que puede ser legítimamente manejado en esta forma y situaciones en donde estimadores estándares aplicados a los datos reales e imputados tienen sesgos substanciales".

En pocas palabras, imputar es un procedimiento para asignar un valor a una variable con no respuesta para que el cuestionario pase el proceso de validación. Además, como señalan los autores mencionados, la imputación puede ser una buena solución al trabajar con datos incompletos, pero también tiene sus riesgos, ya que de no hacerse este procedimiento con el mayor de los cuidados, los datos finales pueden quedar muy sesgados y, por lo tanto, no mostrar la realidad que se pretende descubrir por medio de una encuesta.

Las imputaciones se pueden hacer de forma manual o automática. En el primer caso, el analista examina los datos y él mismo asigna un valor a la variable en cuestión. En el segundo caso, y de acuerdo con la UNECE (2000), las imputaciones automáticas se pueden clasificar en:

- Imputación determinística. Donde solo existe un valor correcto, por ejemplo, la falta de un total. Un valor como éste es determinado a partir de otros valores en el mismo cuestionario.
- Imputación basada en modelo. Usa una media, mediana, un modelo de regresión, etc., para imputar el valor.
- Imputación *Deck*. Se determina un cuestionario donador para suplir el valor faltante:

- Imputación *Hot-Deck*. El donador es encontrado en la misma encuesta. Pueden distinguirse dos variantes, *el determinístico (vecino más cercano)*, en el que un único donador se identifica basado en una métrica y los valores se imputan de ese caso; y *el aleatorio*, en el cual el donante se selecciona aleatoriamente de un conjunto de posibles donadores.
- Imputación *Cold-Deck*. Con ésta técnica, el donador se encuentra en la misma encuesta, pero de un evento previo. Las dos variantes del *Hot-Deck* pueden ser aplicadas en este caso.
- Imputación Mixta. Utiliza una combinación de métodos. Primero se hace la imputación determinística y si ésta falla, se intenta una *Hot-Deck* y si aún sigue fallando lo hace con una imputación basada en modelo, y si de plano todas fallan, entonces se hace una manual.
- Sistemas expertos y redes neuronales. Se pueden desarrollar sistemas expertos o redes neuronales para imputar datos.

Entre los distintos métodos de imputación automática, el método *Cold-Deck*¹ se considera una opción viable para el problema antes planteado debido a que no se trata de imputar valores ausentes en una variable cualitativa debido a una omisión en campo o captura, que pudiera rescatarse a partir de otros valores presentes en el mismo cuestionario, y para lo cual una *imputación determinística* resultaría suficiente. Tampoco se trata de predecir valores de una variable cuantitativa a partir del valor dado por la media, mediana, o por un modelo de regresión, y para lo cual la *imputación basada en modelo* sería la adecuada. En este caso se trata de corregir la omisión de individuos al interior de los hogares y, para ello, es posible aprovechar la información que previamente ha sido generada por el propio INEGI mediante los MCS anteriores. Así, a través de la técnica del *Cold-*

¹ Kim y Fuller (2004) mencionan que la imputación *Hot-Deck* aleatoria preserva las propiedades distributivas del conjunto de datos imputado, además de introducir variabilidad al seleccionar aleatoriamente al donante. Esto es aplicable al método *Cold-Deck* aleatorio.

Deck aleatorio es posible encontrar un donador en la misma encuesta, pero de un evento previo. A diferencia de la *imputación Hot-Deck*, donde el donador proviene del mismo evento, la técnica *Cold-Deck* permite ampliar la base de donadores con la finalidad de mejorar las precisiones estadísticas.

3. Metodología

3.1. Su aplicación

El procedimiento de imputación consistió de tres etapas: 1) determinar el volumen de población expandida a imputar, 2) definir las características sociodemográficas básicas del individuo a imputar y 3) el proceso de imputación como tal.

3.1.1 ¿Cuántos imputar?

Una vez definido el problema, surge la pregunta, ¿cuántos individuos imputar?

Se pueden tomar algunas encuestas del INEGI como punto de partida, de las cuales vale la pena revisar el dato que arrojan y sus características:

- Anclarse a la Encuesta Intercensal (EIC) 2015. Con un tamaño del hogar de 3.74 y una definición de hogar diferente a la de otras encuestas del mismo Instituto, provocando que los hogares no sean comparables debido a que la EIC reporta un hogar por vivienda, mientras que las demás pueden detectar más de uno.
- Basarse en la Encuesta Nacional de Ocupación y Empleo (ENOE). Con un tamaño del hogar de 3.79 (tercer trimestre del 2015) y tasas de variación irregulares. Esto es normal para la ENOE, ya que está diseñada para proporcionar información de la población de 15 años y más de edad sobre ocupación y empleo, como su nombre lo dice; por lo tanto, la información que se pueda obtener de los hogares es secundaria.

- La Encuesta Nacional de la Dinámica Demográfica (ENADID) 2014, con un tamaño del hogar de 3.71.

Al revisar estas encuestas no se dejó de lado el conocimiento de que el tamaño del hogar del MCS 2014 es de 3.73 y que el del MCS 2015 es menor que éste (3.65) y que parte del objetivo del ejercicio es acercarlo mas no dejarlo por arriba; por lo tanto, se descartó considerarlas para el ejercicio.

Otra opción consistía en tomar como referencia las proyecciones de población y hogares del CONAPO, específicamente las tasas de variación implícitas del tamaño del hogar. Que al final fue la vía que se tomó.

Entonces, se decidió que para ajustar demográficamente el MCS 2015, el tamaño del hogar de este evento debía ser menor que el del MCS 2014 para todas las entidades federativas, dada la tendencia a la baja que a través del tiempo muestran las proyecciones del CONAPO (ver tabla 3.1).

Con base en lo anterior, tomamos la tasa de variación implícita 2014-2015 del tamaño del hogar que reportan las proyecciones del CONAPO por entidad federativa ($TCh_i, i = 1, \dots, 32$); así, por ejemplo, la tasa de variación de Aguascalientes es de -0.91% ($TCh_1 = -0.0091$).

Y calculamos el tamaño del hogar esperado en la i -ésima entidad para el 2015 como sigue:

$$TamHOG_{15,i}^{(es)} = TamHOG_{14,i} * (1 + TCh_i),$$

donde $TamHOG_{14,i}$ es el tamaño del hogar reportado por el MCS 2014.

Por ejemplo, para Aguascalientes que presenta $TamHOG_{14,i} = 3.83$ se tiene que:

$$TamHOG_{15,1}^{(es)} = 3.83 * (1 + (-0.0091))$$

$$TamHOG_{15,1}^{(es)} = 3.79$$

Con este dato, se calcula el total de hogares esperado para MCS 2015 en la i -ésima entidad:

$$TotHOG_{15,i}^{(es)} = \frac{TotPOB_{15,i}}{TamHOG_{15,i}^{(es)}},$$

siendo $TotPOB_{15,i}$ la población total de la i -ésima entidad del MCS 2015.

Siguiendo con el ejemplo para Aguascalientes, la población reportada por el MCS 2015 fue $TotPOB_{15,1} = 1\ 292\ 721$, entonces, el total de hogares esperado es:

$$TotHOG_{15,1}^{(es)} = \frac{1\ 292\ 721}{3.79}$$

$$TotHOG_{15,1}^{(es)} = 340\ 989$$

Sin embargo, con el total de hogares esperado y el tamaño del hogar del MCS 2015 para la mayoría de las entidades, no es posible llegar al total de población que reporta este evento, por lo que la población preliminar por entidad está dada por:

$$TotPOB_{15,i}^{(*)} = TotHOG_{15,i}^{(es)} * TamHOG_{15,i}$$

donde $TamHog_{15,i}$ representa el tamaño del hogar que arroja el MCS 2015 para la i -ésima entidad.

Luego, para Aguascalientes:

$$TotPOB_{15,1}^{(*)} = 340\ 989 * 3.78$$

$$TotPOB_{15,1}^{(*)} = 1\ 290\ 080$$

Por lo tanto, es necesario imputar población de forma expandida, que se obtiene por:

$$ImpPOB_{15,i}^{(*)} = TotPOB_{15,i} - TotPOB_{15,i}^{(*)}$$

Siguiendo el caso de Aguascalientes, nos da:

$$ImpPOB_{15,1}^{(*)} = 1\ 292\ 721 - 1\ 290\ 080$$

$$ImpPOB_{15,1}^{(*)} = 2\ 641$$

Tabla 3.1

Tasa de crecimiento media anual del tamaño del hogar

Entidad	2010-2012	2012-2014	2014-2015
Aguascalientes	-0.93	-0.92	-0.91
Baja California	-1.08	-0.99	-0.95
Baja California Sur	-0.71	-0.68	-0.66
Campeche	-0.83	-0.84	-0.83
Coahuila de Zaragoza	-0.78	-0.76	-0.75
Colima	-0.81	-0.78	-0.76
Chiapas	-0.87	-0.93	-0.97
Chihuahua	-0.73	-0.66	-0.63
Ciudad de México	-1.00	-0.94	-0.90
Durango	-0.90	-0.87	-0.87
Guanajuato	-0.87	-0.88	-0.88
Guerrero	-0.73	-0.81	-0.85
Hidalgo	-0.79	-0.80	-0.81
Jalisco	-0.89	-0.87	-0.84
México	-1.00	-0.99	-0.96
Michoacán de Ocampo	-0.76	-0.77	-0.75
Morelos	-0.74	-0.74	-0.71
Nayarit	-0.57	-0.54	-0.54
Nuevo León	-0.85	-0.80	-0.77
Oaxaca	-0.53	-0.58	-0.62
Puebla	-0.75	-0.79	-0.81
Querétaro	-1.07	-1.05	-1.02
Quintana Roo	-1.06	-0.92	-0.84
San Luis Potosí	-0.65	-0.70	-0.73
Sinaloa	-0.93	-0.87	-0.85
Sonora	-0.75	-0.72	-0.71
Tabasco	-1.00	-1.02	-1.00
Tamaulipas	-0.80	-0.76	-0.75
Tlaxcala	-0.85	-0.86	-0.87
Veracruz de Ignacio de la Llave	-0.75	-0.76	-0.76
Yucatán	-0.78	-0.75	-0.73
Zacatecas	-0.66	-0.69	-0.70

Fuente: cálculos propios con base en las proyecciones de población y hogares del CONAPO.

Esos 2 641 casos representan la cobertura de individuos que se deben imputar de forma expandida para Aguascalientes; cabe aclarar que en algunas entidades el resultado de este ajuste fue negativo; ante ello, se tomó la decisión de no hacer nada en esos casos (ver tabla 3.2).

3.1.2 ¿Qué y a quién imputar?

Una vez que se tiene el número de pobladores a imputar en cada entidad, ahora sigue determinar las características de los individuos donadores. Para ello, requerimos cuidar ciertos aspectos:

- La proporción de hombres y mujeres en las entidades federativas.
- La proporción de los grupos de edad (0 a 11, 12 a 64 y 65 y más), los tamaños de localidad y los estratos socioeconómicos tanto en hombres como en mujeres, por entidad federativa.

Al cuidar esas proporciones buscamos evitar que el proceso de imputación sesgue los resultados hacia una o varias categorías de esas variables, es decir, que al final del proceso haya más hombres que mujeres, por ejemplo.

Para obtener las proporciones, usamos información de la EIC y la ENOE, no con el fin de tratar de acercarse a su estructura poblacional (el MCS 2015 ya trae la propia y modificarla sería demasiado ostentoso) sino, como ya se dijo, para evitar el sesgo hacia alguna categoría.

Para decidir sobre las características que debe tener el donador, se usaron las proporciones ya mencionadas. En el caso del sexo se utilizaron las de hombres y mujeres por entidad federativa. Para grupos de edad, tamaño de localidad y estrato socioeconómico, se construyeron intervalos de decisión para cada categoría de manera que fuera de una longitud igual a la proporción observada ya

Tabla 3.2

Coberturas de imputación por entidad federativa

Entidad	Cobertura
Total	1 956 473
Aguascalientes	2 641
Baja California	165 844
Baja California Sur	23 583
Campeche	0
Coahuila de Zaragoza	35 612
Colima	20 877
Chiapas	0
Chihuahua	40 888
Ciudad de México	200 560
Durango	29 222
Guanajuato	0
Guerrero	157 513
Hidalgo	0
Jalisco	0
México	307 343
Michoacán de Ocampo	182 359

Entidad	Cobertura
Morelos	53 027
Nayarit	9 639
Nuevo León	119 344
Oaxaca	207 278
Puebla	92 396
Querétaro	46 780
Quintana Roo	48 339
San Luis Potosí	78 436
Sinaloa	25 861
Sonora	412
Tabasco	0
Tamaulipas	50 154
Tlaxcala	9 777
Veracruz de Ignacio de la Llave	0
Yucatán	48 587
Zacatecas	0

sea en hombres o bien en mujeres (ver tablas 3.3 a 3.5 para el caso de Aguascalientes).

La segregación de sexo por las variables *entidad federativa, grupos de edad, tamaño de localidad y estrato socioeconómico*, realizada de forma independiente para obtener las proporciones, se hizo pensando que al hacer inferencia estadística a esos niveles se obtienen coeficientes de variación aceptables para un gran número de variables. Hacerlo de forma diluida, es decir, desagregando sexo den-

tro de grupos de edad, estrato socioeconómico, tamaño de localidad y entidad federativa complicaría el procedimiento (dada la alta posibilidad de no encontrar donadores), además de que la inferencia estadística a esos niveles de desagregación arroja coeficientes de variación no satisfactorios.

El proceso de asignación de características se lleva a cabo entidad por entidad, generando números aleatorios de la distribución uniforme, siendo de la siguiente forma:

Tabla 3.3

Intervalos de decisión para grupos de edad. Aguascalientes

Grupos de edad	Hombres			Mujeres		
	Proporción	LI	LS	Proporción	LI	LS
1. De 0 a 11 años	0.241	0.000	0.241	0.224	0.000	0.224
2. De 12 a 64 años	0.705	0.241	0.946	0.714	0.224	0.938
3. De 65 y más años	0.054	0.946	1.000	0.062	0.938	1.000

Fuente: INEGI. Encuesta Intercensal 2015.

Tabla 3.4

Intervalos de decisión para tamaño de localidad. Aguascalientes

Tamaño de localidad	Hombres			Mujeres		
	Proporción	LI	LS	Proporción	LI	LS
1. De 100 mil y más habitantes	0.595	0.000	0.595	0.602	0.000	0.602
2. De 15 mil a 99 999 habitantes	0.113	0.595	0.708	0.113	0.602	0.715
3. De 2 500 a 14 999 habitantes	0.089	0.708	0.797	0.088	0.715	0.803
4. De menos de 2 500 habitantes	0.203	0.797	1.000	0.197	0.803	1.000

Fuente: INEGI. Encuesta Intercensal 2015.

Tabla 3.5

Intervalos de decisión para estrato económico. Aguascalientes

Estrato socioeconómico	Hombres			Mujeres		
	Proporción	LI	LS	Proporción	LI	LS
1. Bajo						
2. Medio bajo	0.553	0.000	0.553	0.540	0.000	0.540
3. Medio alto	0.327	0.553	0.880	0.334	0.540	0.874
4. Alto	0.120	0.880	1.000	0.126	0.874	1.000

Fuente: INEGI. Encuesta Nacional de Ocupación y Empleo. Tercer trimestre del 2015.

1. Para determinar el sexo, se genera un primer número aleatorio, si este es menor o igual a la proporción de hombres, se asigna un hombre, de lo contrario, se asigna una mujer.
2. Ya que se tiene el sexo del individuo, se genera un segundo número aleatorio para asignarle el grupo de edad condicionado al sexo determinado con anterioridad. Para decidir sobre éste, se verifica en qué intervalo de decisión se incluye.
3. Para saber qué tamaño de localidad y estrato socioeconómico es asignado, se generan un tercer y cuarto números aleatorios, y se revisa el intervalo de decisión en el que se incluyen, condicionado también al sexo.

3.1.3 Proceso de imputación

El proceso de imputación se hace también entidad por entidad.

- Para seleccionar a los hogares receptores de individuos, se usa la técnica *Permanent Random Numbers (PRN)*. Ésta consiste en asignar un número aleatorio (de la distribución uniforme) a cada hogar y después ordenarlos descendientemente por ese aleatorio.
- Una vez ordenados los hogares por el número aleatorio, se busca el primer hogar que cumpla con las condiciones de mismo tamaño de localidad y mismo estrato socioeconómico como los ya designados.
- Se asigna el individuo a ese hogar y se acumula su factor de expansión, luego el siguiente y acumulando su factor de expansión y así hasta que el factor acumulado supere la cobertura de la entidad, como se muestran en la tabla 3.1.

Una vez que se determinó lo que ha de imputarse y se seleccionaron los receptores, se procede al proceso de imputación de valores para el resto de variables aún no determinadas.

Para asegurarnos de encontrar posibles donadores, se construyó un agregado (*pool*) de pobladores

de los MCS 2012 a 2015 con el fin de tener un conjunto de individuos mucho más grande de dónde obtener dichos donadores. En la tabla 3.6 se presentan el total de imputados y su origen.

Tabla 3.6

Total de imputados por MCS de origen

MCS	Imputados	%
Total	3 759	100.0
2012	1 039	27.6
2014	1 139	30.3
2015	1 581	42.1

El método de imputación usado fue *Cold-Deck* aleatorio, considerando las siguientes variables de empate para encontrar posibles donadores:

- Entidad federativa.
- Tamaño de localidad.
- Estrato socioeconómico.
- Clase de hogar (nuclear, ampliado y compuesto).
- Sexo, edad y nivel educativo del jefe del hogar.
- Sexo y grupo de edad del individuo a imputar.

A pesar de que se pudo haber elegido un conjunto más amplio de variables, no se hizo, puesto que con dicha decisión se corría el riesgo de que en un gran número de casos no encontrarán posibles donadores.

Es importante mencionar que aunque la clase de hogar se incluyó como variable de empate, se determinó no modificarla. Para ello, se filtraron los parentescos que pueden ser parte de cada clase; por ejemplo, para hogar nuclear solo se buscaron donadores cuyo parentesco fuera hija o hijo, para hogares ampliados solo se buscaron a hijas o hijos y otros parientes, y para compuestos se buscó a hijas o hijos, otros parientes y no parientes. En la tabla 3.7 se puede observar el total de imputados por la clase de hogar y parentesco.

Además, para los parentescos se aplicaron ciertos filtros:

Tabla 3.7

Total de imputados por clase de hogar y parentesco

Clase de hogar	Parentesco	Imputados	%
Total		3 759	100.0
Nuclear	Total	2 662	70.8
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	2 574	68.5
	Hijo(a) adoptivo(a)	1	0.0
	Hijastro(a), entenado(a)	87	2.3
Ampliado	Total	1 062	28.3
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	667	17.7
	Hijastro(a), entenado(a)	10	0.3
	Madre, padre	68	1.8
	Padrastra, madrastra	2	0.1
	Hermano(a)	60	1.6
	Abuelo(a)	3	0.1
	Nieto(a)	215	5.7
	Bisnieto(a)	2	0.1
	Tío(a)	7	0.2
	Sobrino(a)	23	0.6
	Primo(a)	5	0.1
Compuesto	Total	35	0.9
	Hijo(a), hijo(a) consanguíneo(a), hijo(a) reconocido(a)	22	0.6
	Hijastro(a), entenado(a)	1	0.0
	No tiene parentesco	7	0.2
	Nieto(a)	2	0.1
	Bisnieto(a)	1	0.0
	Sobrino(a)	1	0.0
	Primo(a)	1	0.0

- Si el donador era hija o hijo, se cuidó que la edad de la jefa o del jefe del hogar receptor fuera 13 o más años mayor a dicho donador, con el fin de evitar imputar un hijo mayor que el jefe.
- Si el donador era hija o hijo y el hogar receptor era encabezado por una mujer, se buscó que la diferencia máxima de edades entre la jefa o el jefe y el donador fuera de 45 años, para evitar imputar hijas o hijos menores de 12 años a jefas de hogar de la tercera edad.
- Si el donador era nieta o nieto, bisnieta o bisnieto o bien tataranieta o tataranieto, se cuidó que la edad de la jefa o el jefe del hogar

receptor fuera 26, 39 y 52 años mayor que aquéllos (como mínimo), respectivamente.

- Si el donador era abuela o abuelo, bisabuela o bisabuelo o bien tatarabuela o tatarabuelo, se cuidó que la edad del jefe del hogar receptor fuera 26, 39 y 52 años menor que aquéllos (como mínimo), respectivamente.

Además, cabe aclarar que solo se donó la información sociodemográfica del individuo, excluyendo toda aquella relacionada con ingresos, gastos y trabajos. Así que, cuando se imputó a un individuo cuyo donador estaba clasificado como población no económicamente activa (PNEA) o buscador de trabajo, se le respetó la categoría. Sin embargo, cuando el donador era una persona con trabajo, se cambió la condición de actividad a PNEA o buscador de trabajo del individuo agregado. Lo anterior en el entendido de no incrementar el número de trabajadores.

Los siguientes puntos enuncian la forma en que quedaron clasificados aquellos registros en los que sus donadores eran parte de la población económicamente activa (PEA):

- Como estudiante si asistía a la escuela.
- Dedicado a quehaceres domésticos si era menor de 15 años de edad y no asistía a la escuela.
- Buscador de trabajo si era hombre de 15 años y más y no asistía a la escuela.
- 46% de las mujeres se clasificaron como dedicadas a quehaceres domésticos si eran de 15 años y más y no asistía a la escuela; las restantes (44%) se catalogaron como buscadoras de trabajo (en virtud de los porcentajes de mujeres de 15 años y más que se clasifican como PNEA y PEA²).

Debido a que son pocas las características de empate entre la persona que se busca y los posibles donadores, al ir por estos últimos se pueden encontrar más de uno. Si esto ocurre, se vuelve a usar la técnica PRN para elegir solo uno.

² INEGI. *Encuesta Nacional de Ocupación y Empleo*. Tercer trimestre del 2015.

Una vez concluido el proceso de imputación, dado que hay más pobladores, se reajustaron los factores de expansión (ajuste de razón por entidad y tamaño de localidad) usando como base los factores de expansión del propio MCS 2015 (como fueron publicados) para que el total de la población expandida fuera congruente con las proyecciones del CONAPO para el 2015. En general, ello significa que los factores de expansión no tienen que ser tan grandes como los originales y, por ende, se obtienen menos hogares.

3.2. Nota técnica

3.2.1 Software usado

Visual FoxPro 9.0 de Microsoft fue el *software* usado para llevar a cabo el procedimiento de imputación. Esta pieza de *software* es un lenguaje de programación procedural y orientado a objetos que incluye un Sistema Gestor de Bases de Datos (DBMS, por sus siglas en inglés).

Para llevar a cabo el trabajo, se hizo uso tanto del lenguaje de programación como del SQL que este *software* incluye.

3.2.2 Fuentes utilizadas

La primera fuente de información usada fueron las proyecciones del CONAPO con el fin de obtener las variaciones 2014-2015 del tamaño del hogar. Una segunda fue la *Encuesta Intercensal 2015*, de la cual se consiguieron las proporciones de hombres y mujeres por entidad federativa, las proporciones de hombres por grupos de edad y tamaño de localidad, así como las proporciones de mujeres por grupos de edad y tamaño de localidad, en cada entidad federativa. Otra más fue la *Encuesta Nacional de Ocupación y Empleo*, de la que se obtuvieron las proporciones de hombres por estrato socioeconómico y mujeres por los mismos estratos en cada entidad federativa.

Ahora bien, para incrementar las posibilidades de encontrar donadores para todos los casos a imputar,

se construyó un *pool* de pobladores con los MCS 2012 a 2015, donde se incluyeron todas las variables sociodemográficas (solo la tabla *Población* de la base de datos).

Entonces, enlistando las fuentes, serían:

- INEGI. *Módulo de Condiciones Socioeconómicas 2012*.
- INEGI. *Módulo de Condiciones Socioeconómicas 2014*.
- INEGI. *Módulo de Condiciones Socioeconómicas 2015*.
- INEGI. *Encuesta Intercensal 2015*.

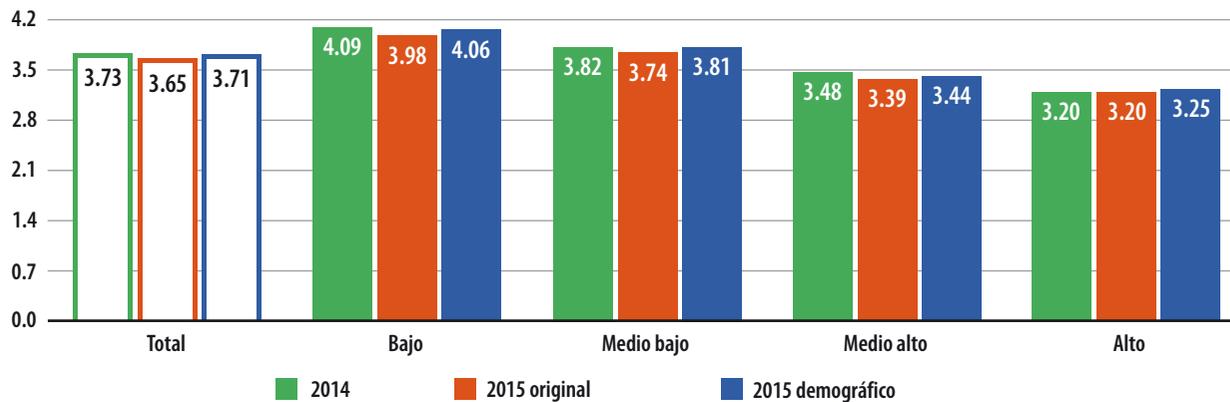
- INEGI. *Encuesta de Ocupación y Empleo*. Tercer trimestre del 2015.
- CONAPO. Proyecciones de población nacional y entidades federativas, 2010-2030.
- CONAPO. Proyecciones de los hogares en México y las entidades federativas, 2010-2030.

4. Resultados

4.1 Resultados en el ámbito nacional, por entidad federativa y por decil con la nueva base del MCS 2015

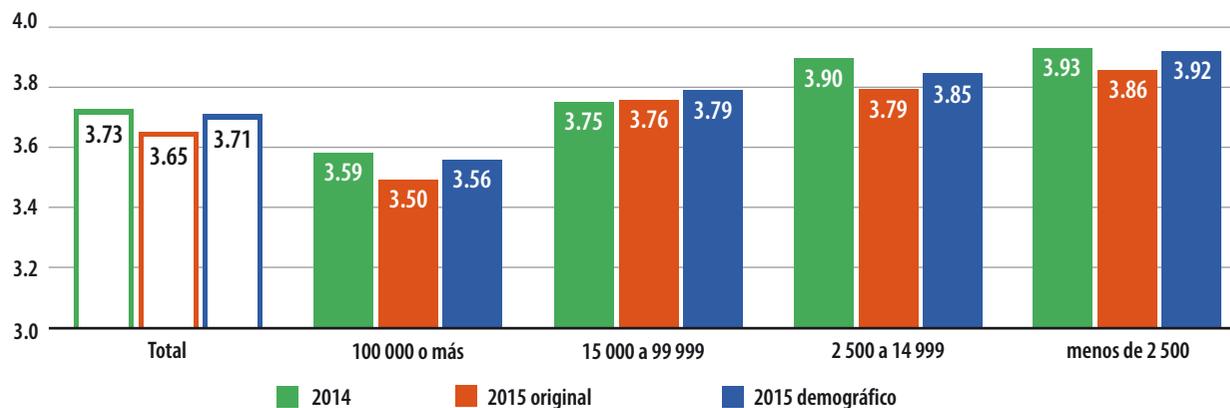
Gráfica 4.1

Tamaño del hogar por estrato socioeconómico



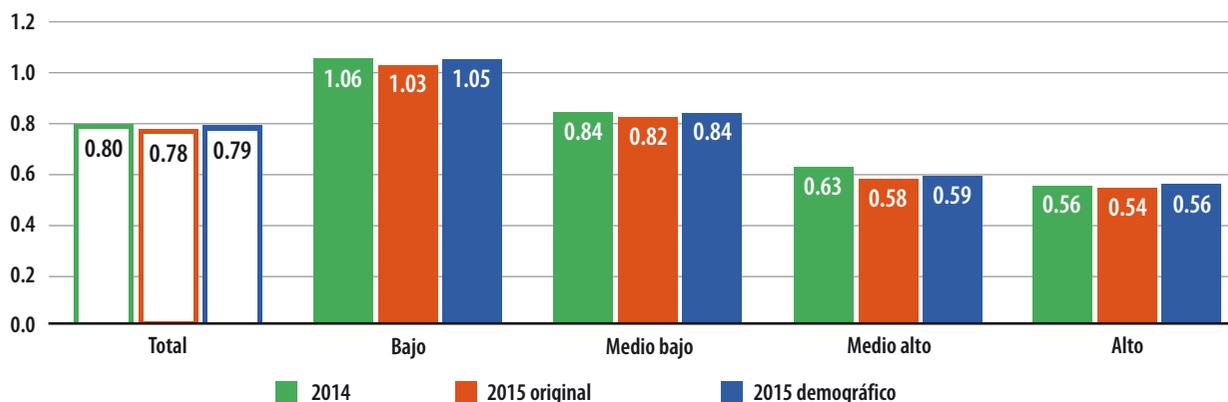
Gráfica 4.2

Tamaño del hogar por tamaño de localidad



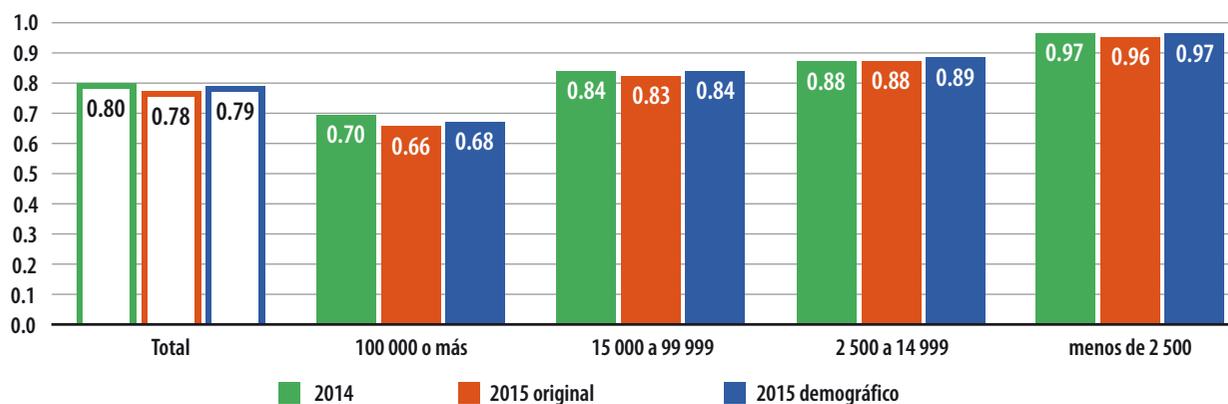
Gráfica 4.3

Promedio de integrantes del hogar menores de 12 años por estrato socioeconómico



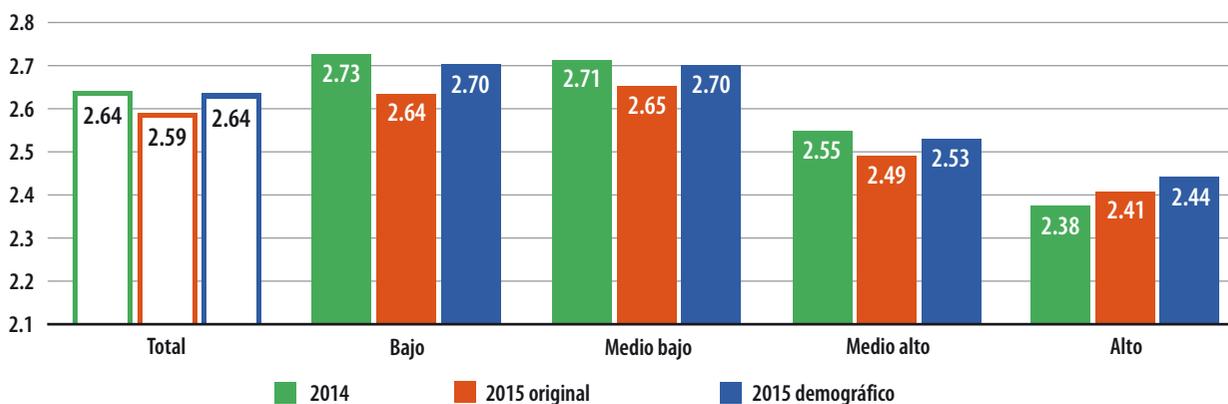
Gráfica 4.4

Promedio de integrantes del hogar menores de 12 años por tamaño de localidad



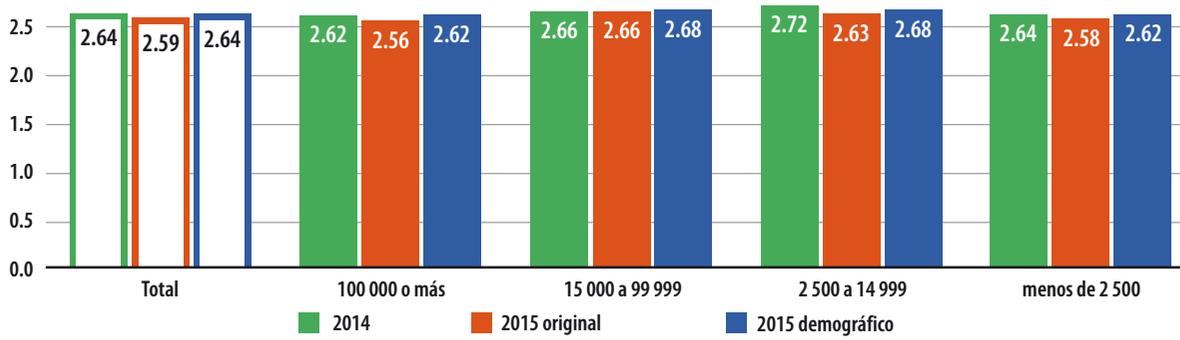
Gráfica 4.5

Promedio de integrantes del hogar de 12 a 64 años por estrato socioeconómico



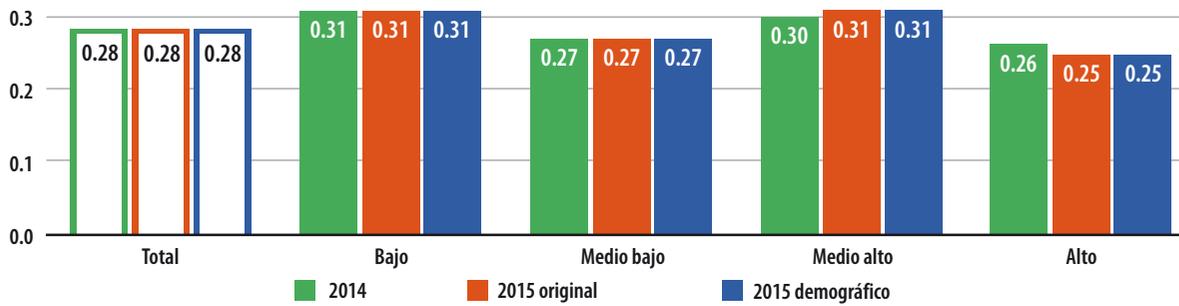
Gráfica 4.6

Promedio de integrantes del hogar de 12 a 64 años por tamaño de localidad



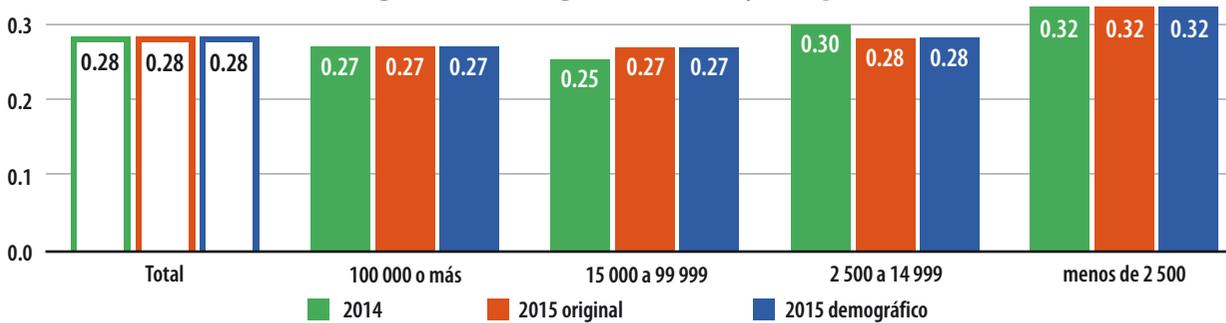
Gráfica 4.7

Promedio de integrantes del hogar de 65 años y más por estrato socioeconómico



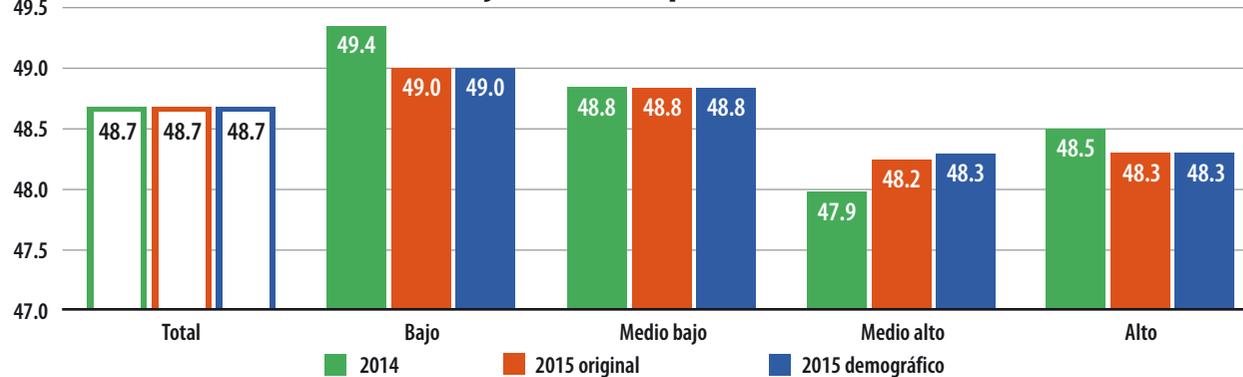
Gráfica 4.8

Promedio de integrantes del hogar de 65 años y más por tamaño de localidad



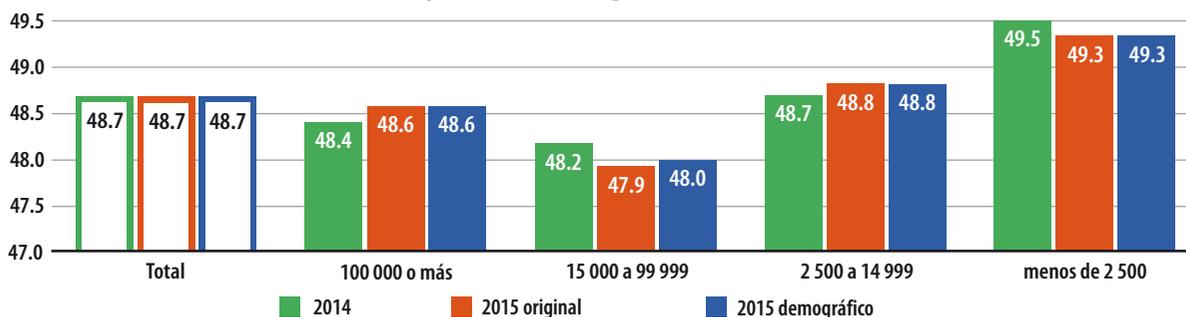
Gráfica 4.9

Porcentaje de hombres por estrato socioeconómico



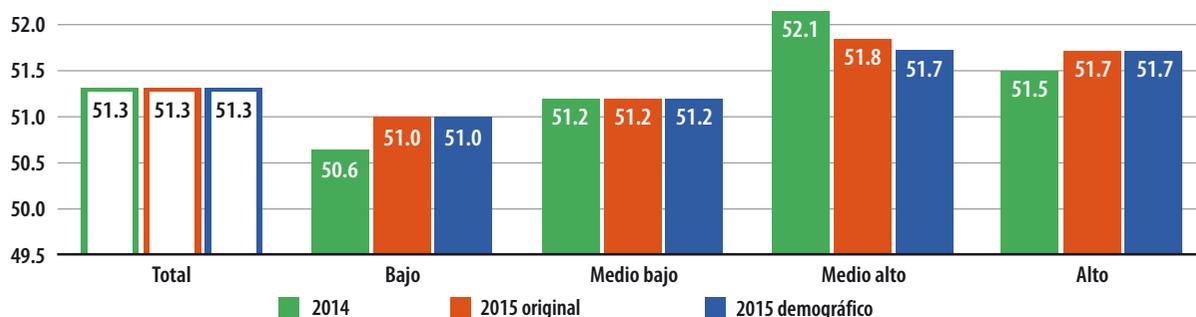
Gráfica 4.10

Porcentaje de hombres por tamaño de localidad



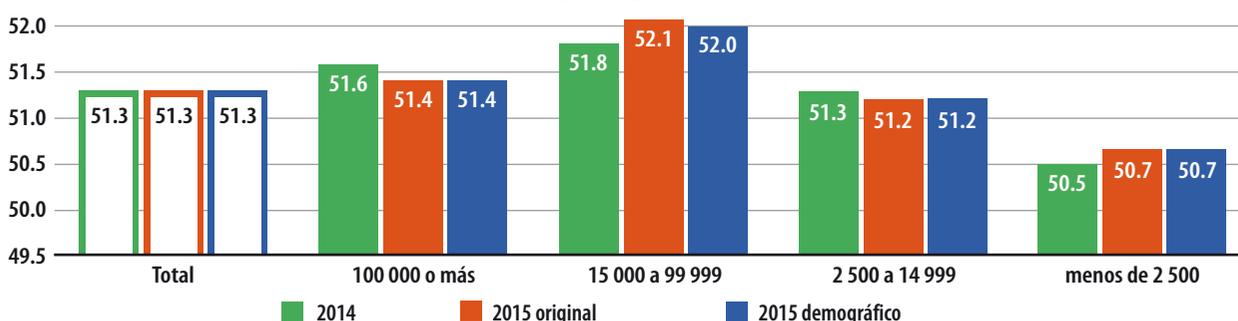
Gráfica 4.11

Porcentaje de mujeres por estrato socioeconómico



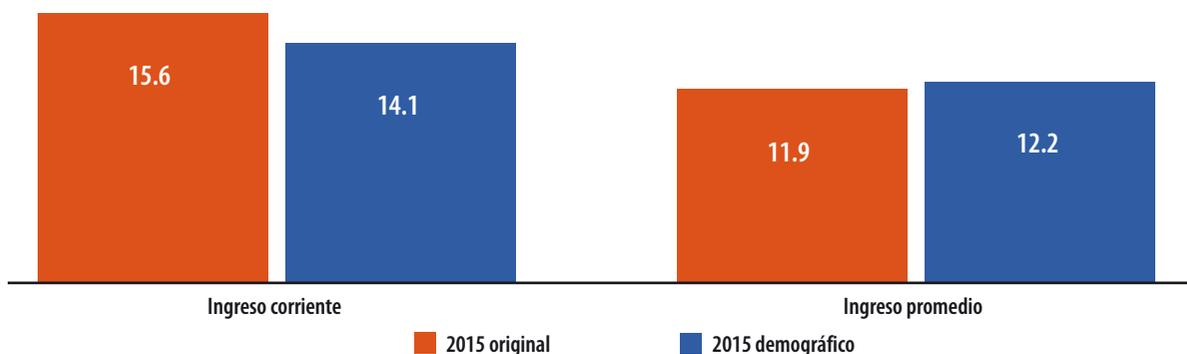
Gráfica 4.12

Porcentaje de mujeres por tamaño de localidad



Gráfica 4.13

Variación porcentual del ingreso corriente 2014-2015



4.2 Comparación de los resultados con los publicados para el MCS 2015 y ENIGH-MCS 2010, 2012 y 2014. A nivel nacional, por entidad federativa y por decil

Gráfica 4.14

Promedio por hogar del ingreso corriente total 2010-2015

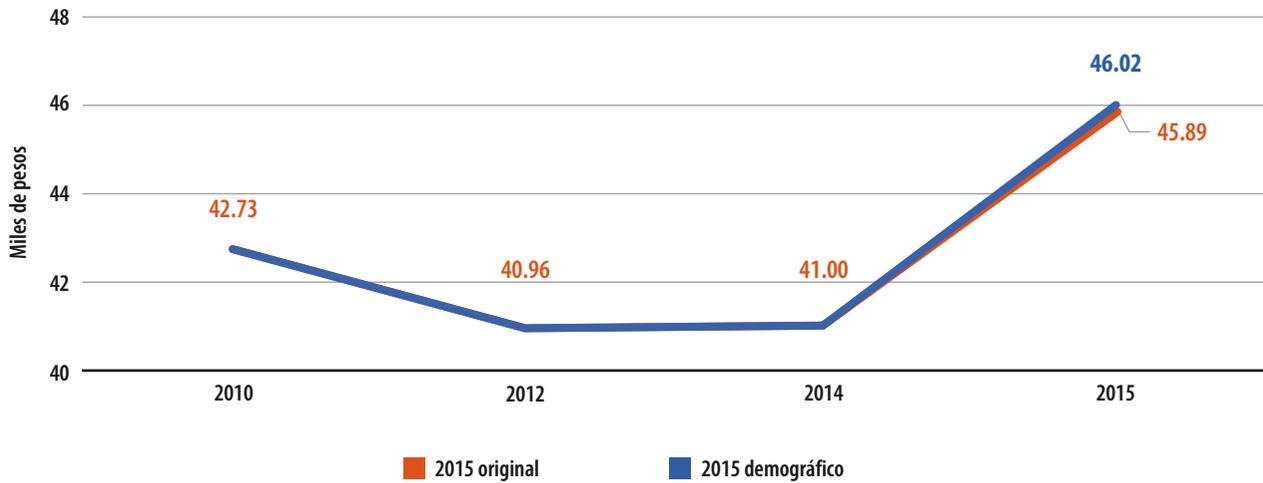


Tabla 4.1

Continúa

Variaciones porcentuales del tamaño de los hogares, del ingreso total y del ingreso promedio por hogar para cada entidad federativa

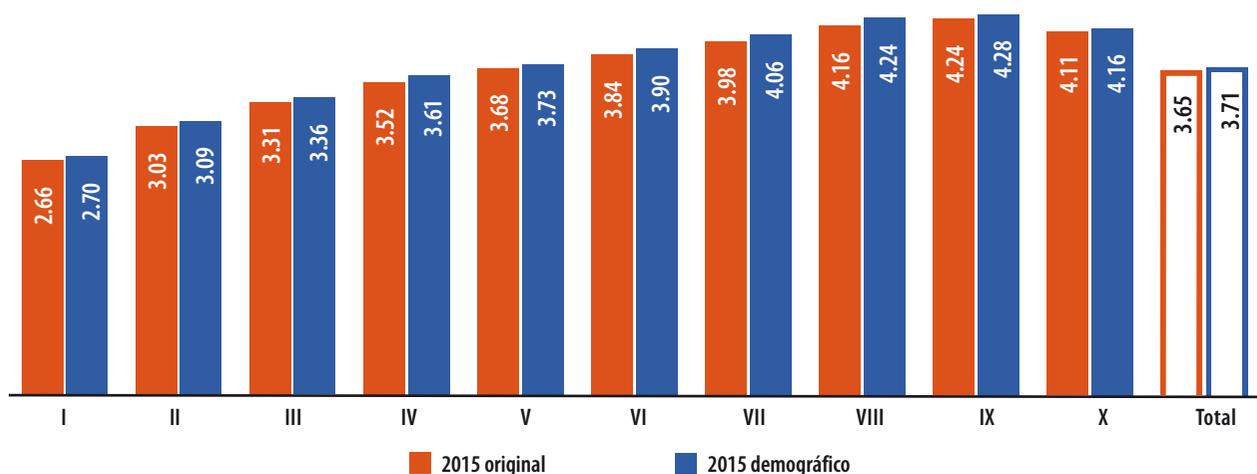
Entidad	2014-2015 original			2014-2015 demográfico		
	Hogares	Ingreso corriente	Ingreso promedio por hogar	Hogares	Ingreso corriente	Ingreso promedio por hogar
Estados Unidos Mexicanos	3.32	15.64	11.92	1.65	14.10	12.25
Aguascalientes	2.61	28.31	25.05	2.45	27.74	24.68
Baja California	7.67	19.96	11.42	2.23	13.96	11.47
Baja California Sur	7.02	33.13	24.40	3.50	28.64	24.29
Campeche	1.42	16.03	14.40	1.39	15.88	14.30
Coahuila de Zaragoza	3.17	6.98	3.68	2.10	7.44	5.23
Colima	5.81	12.45	6.27	2.54	9.00	6.29
Chiapas	0.31	32.66	32.26	0.33	32.75	32.31
Chihuahua	2.84	48.55	44.45	1.50	46.38	44.22
Ciudad de México	3.56	-1.68	-5.06	1.05	-4.16	-5.15
Durango	3.64	29.56	25.01	1.84	27.20	24.90
Guanajuato	1.31	31.69	29.99	1.14	33.86	32.35
Guerrero	6.21	9.39	2.99	1.18	3.85	2.64
Hidalgo	1.80	17.05	14.99	1.96	16.53	14.29

Variaciones porcentuales del tamaño de los hogares, del ingreso total y del ingreso promedio por hogar para cada entidad federativa

Entidad	2014-2015 original			2014-2015 demográfico		
	Hogares	Ingreso corriente	Ingreso promedio por hogar	Hogares	Ingreso corriente	Ingreso promedio por hogar
Jalisco	-0.78	15.03	15.94	-0.54	16.58	17.21
México	4.44	10.33	5.64	2.35	8.27	5.79
Michoacán de Ocampo	5.67	31.56	24.50	1.56	26.77	24.83
Morelos	4.94	26.64	20.67	1.81	25.28	23.05
Nayarit	3.29	22.19	18.30	2.41	20.99	18.15
Nuevo León	4.91	14.24	8.90	2.37	11.56	8.98
Oaxaca	6.91	24.65	16.59	1.25	18.40	16.93
Puebla	3.64	4.92	1.24	1.92	2.79	0.85
Querétaro	5.24	21.88	15.82	2.75	20.05	16.83
Quintana Roo	7.16	1.62	-5.17	3.85	-0.71	-4.39
San Luís Potosí	5.01	26.02	20.00	1.80	22.10	19.94
Sinaloa	2.68	35.95	32.40	1.67	34.68	32.46
Sonora	2.20	14.63	12.16	2.12	14.57	12.18
Tabasco	2.11	20.27	17.79	2.09	20.05	17.59
Tamaulipas	3.52	11.01	7.23	2.06	9.38	7.17
Tlaxcala	3.18	26.63	22.74	2.52	26.29	23.19
Veracruz de Ignacio de la Llave	0.49	8.65	8.11	2.10	11.74	9.43
Yucatán	4.94	23.11	17.32	2.23	19.90	17.29
Zacatecas	-1.32	13.57	15.09	-1.27	13.90	15.36

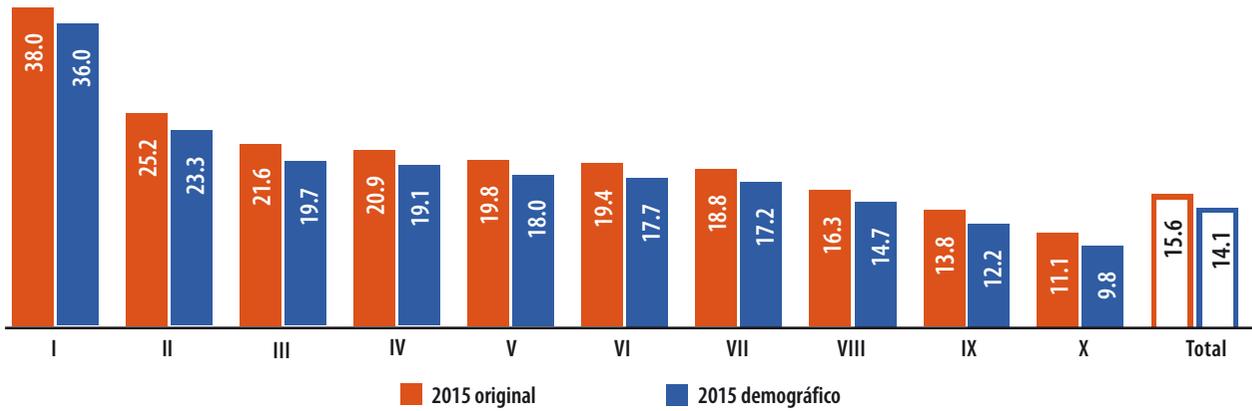
Gráfica 4.15

Tamaño promedio del hogar por deciles



Gráfica 4.16

Volumen de ingreso corriente total por deciles. Variación porcentual 2014-2015



Gráfica 4.17

Promedio por hogar de ingreso corriente total por deciles. Variación porcentual 2014-2015

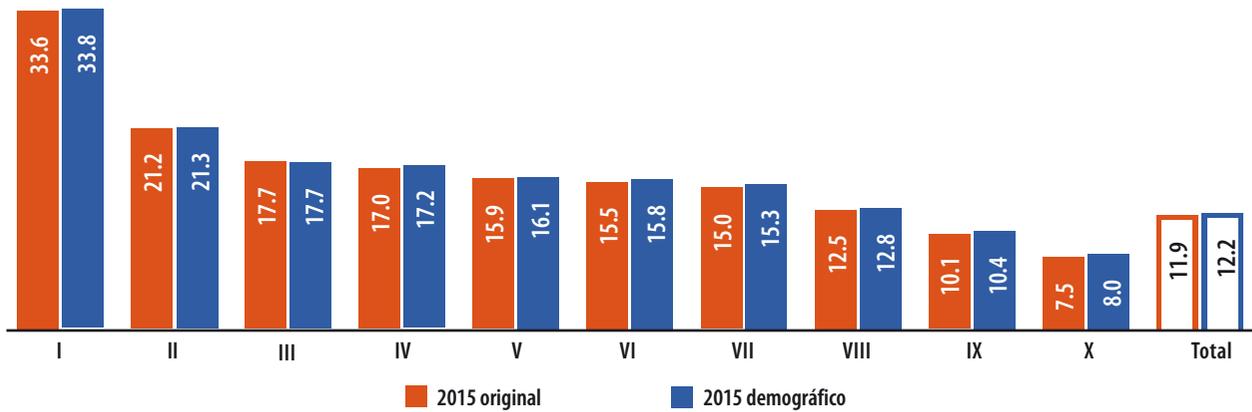


Tabla 4.2

Descomposición del ingreso corriente total (ICT)

Decil	2015 original				2015 demográfico			
	ICT ₁₅ / ICT ₁₄	IPP	PH	Hog	ICT ₁₅ / ICT ₁₄	IPP	PH	Hog
Total	1.156	1.077	1.039	1.033	1.141	1.082	1.037	1.017
I	1.380	1.313	1.017	1.033	1.360	1.318	1.015	1.017
II	1.252	1.243	0.976	1.033	1.233	1.243	0.976	1.017
III	1.216	1.155	1.019	1.033	1.197	1.155	1.019	1.017
IV	1.209	1.146	1.021	1.033	1.191	1.147	1.022	1.017
V	1.198	1.105	1.049	1.033	1.180	1.113	1.044	1.017
VI	1.194	1.123	1.028	1.033	1.177	1.127	1.028	1.017
VII	1.188	1.094	1.051	1.033	1.172	1.097	1.051	1.017
VIII	1.163	1.072	1.049	1.033	1.147	1.076	1.049	1.017
IX	1.138	1.030	1.069	1.033	1.122	1.036	1.066	1.017
X	1.111	0.993	1.082	1.033	1.098	1.000	1.080	1.017

Nota: IPP = variación de ingreso corriente por perceptor, PH = variación de perceptores por hogar y Hog = variación de hogares.

Tabla 4.3

Pobreza multidimensional en MCS 2015 original

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Total	36.69%	30.85%	5.85%	33.94%	5.45%	23.91%
Aguascalientes	25.88%	24.38%	1.50%	33.05%	7.72%	33.35%
Baja California	24.19%	22.52%	1.67%	38.80%	5.94%	31.08%
Baja California Sur	22.63%	20.92%	1.71%	40.11%	4.60%	32.66%
Campeche	34.16%	29.66%	4.50%	39.89%	2.89%	23.06%
Coahuila de Zaragoza	19.01%	17.62%	1.40%	30.92%	10.49%	39.57%
Colima	25.99%	24.08%	1.91%	41.03%	5.03%	27.95%
Chiapas	65.54%	40.22%	25.32%	24.11%	1.94%	8.41%
Chihuahua	21.14%	20.25%	0.89%	34.73%	8.07%	36.07%
Ciudad de México	25.54%	24.68%	0.86%	28.92%	6.98%	38.56%
Durango	32.37%	30.12%	2.25%	35.01%	7.59%	25.03%
Guanajuato	30.04%	28.31%	1.73%	43.74%	4.65%	21.56%
Guerrero	59.35%	41.44%	17.91%	32.51%	1.80%	6.33%
Hidalgo	41.41%	34.09%	7.32%	39.34%	3.20%	16.06%
Jalisco	24.91%	22.50%	2.41%	38.70%	5.35%	31.04%
México	40.69%	36.37%	4.31%	27.86%	8.62%	22.83%
Michoacán de Ocampo	44.42%	35.79%	8.64%	38.64%	2.98%	13.96%
Morelos	36.78%	31.36%	5.42%	38.47%	3.93%	20.82%
Nayarit	31.40%	27.12%	4.28%	38.81%	5.47%	24.32%
Nuevo León	17.49%	16.52%	0.96%	30.29%	8.05%	44.18%
Oaxaca	59.73%	40.74%	18.99%	28.54%	1.60%	10.14%
Puebla	50.11%	42.87%	7.24%	32.76%	4.30%	12.83%
Querétaro	24.03%	21.67%	2.36%	40.37%	4.95%	30.65%
Quintana Roo	30.01%	25.01%	4.99%	39.76%	4.80%	25.43%
San Luis Potosí	36.50%	30.47%	6.03%	34.40%	4.88%	24.22%
Sinaloa	27.47%	25.72%	1.74%	40.37%	4.94%	27.22%
Sonora	22.46%	20.39%	2.07%	36.84%	5.68%	35.02%
Tabasco	36.68%	31.01%	5.67%	50.50%	1.64%	11.17%
Tamaulipas	31.99%	28.70%	3.29%	31.75%	6.92%	29.34%
Tlaxcala	43.39%	40.59%	2.80%	33.70%	6.39%	16.52%
Veracruz de Ignacio de la Llave	50.72%	40.26%	10.46%	33.28%	2.89%	13.11%
Yucatán	34.36%	28.59%	5.76%	36.94%	4.80%	23.91%
Zacatecas	44.69%	40.09%	4.60%	30.49%	5.01%	19.81%

Tabla 4.4

Pobreza multidimensional en MCS 2015 demográfico

Entidad federativa	Pobreza	Pobreza moderada	Pobreza extrema	Población vulnerable por carencias	Población vulnerable por ingresos	Población no pobre y no vulnerable
Total	37.32%	31.34%	5.98%	33.25%	5.63%	23.80%
Aguascalientes	26.12%	24.61%	1.51%	32.75%	7.81%	33.32%
Baja California	26.33%	24.67%	1.66%	36.63%	7.00%	30.04%
Baja California Sur	24.37%	22.45%	1.93%	38.47%	5.35%	31.81%
Campeche	34.20%	29.67%	4.53%	39.92%	2.89%	22.99%
Coahuila de Zaragoza	18.96%	17.57%	1.39%	30.65%	10.11%	40.28%
Colima	27.95%	25.63%	2.32%	39.29%	5.84%	26.92%
Chiapas	65.51%	40.21%	25.30%	24.18%	1.93%	8.38%
Chihuahua	21.57%	20.70%	0.88%	34.43%	8.03%	35.97%
Ciudad de México	26.33%	25.37%	0.96%	28.23%	7.50%	37.94%
Durango	32.62%	30.22%	2.40%	34.70%	7.76%	24.92%
Guanajuato	29.65%	27.93%	1.72%	44.20%	4.56%	21.59%
Guerrero	61.77%	42.52%	19.24%	30.29%	1.88%	6.06%
Hidalgo	41.13%	33.79%	7.34%	38.93%	3.54%	16.40%
Jalisco	24.66%	22.34%	2.33%	38.78%	5.26%	31.30%
México	41.63%	37.17%	4.45%	27.10%	8.78%	22.49%
Michoacán de Ocampo	46.66%	37.66%	9.00%	36.14%	3.56%	13.64%
Morelos	37.89%	32.22%	5.68%	36.92%	4.27%	20.91%
Nayarit	31.75%	27.39%	4.37%	38.52%	5.50%	24.23%
Nuevo León	18.16%	16.92%	1.24%	29.85%	8.62%	43.37%
Oaxaca	61.74%	41.60%	20.15%	26.44%	1.82%	9.99%
Puebla	51.00%	43.62%	7.38%	31.94%	4.36%	12.70%
Querétaro	24.91%	22.41%	2.50%	38.96%	5.29%	30.84%
Quintana Roo	31.35%	26.33%	5.02%	38.16%	5.48%	25.00%
San Luis Potosí	37.57%	31.33%	6.24%	33.60%	5.36%	23.46%
Sinaloa	28.04%	26.15%	1.89%	39.86%	5.06%	27.05%
Sonora	22.50%	20.42%	2.08%	36.86%	5.67%	34.96%
Tabasco	36.75%	31.06%	5.69%	50.47%	1.65%	11.13%
Tamaulipas	32.55%	29.33%	3.22%	31.28%	7.06%	29.11%
Tlaxcala	43.83%	40.95%	2.87%	33.10%	6.48%	16.60%
Veracruz de Ignacio de la Llave	49.82%	39.73%	10.09%	33.00%	2.74%	14.44%
Yucatán	35.78%	29.70%	6.08%	35.74%	4.89%	23.58%
Zacatecas	44.63%	40.03%	4.60%	30.51%	5.00%	19.86%

Tabla 4.5

Pobreza multidimensional. Población con carencias en MCS 2015 original

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Total	70.64%	20.53%	17.95%	16.87%	56.59%	12.01%	20.48%	21.75%
Aguascalientes	58.93%	8.56%	13.76%	13.77%	42.93%	4.90%	4.70%	18.79%
Baja California	62.99%	12.45%	13.57%	18.10%	46.39%	9.34%	12.59%	14.95%
Baja California Sur	62.74%	14.61%	14.17%	12.76%	43.01%	14.12%	11.57%	21.72%
Campeche	74.05%	27.45%	20.14%	12.35%	57.60%	18.36%	34.04%	23.17%
Coahuila de Zaragoza	49.93%	7.50%	12.29%	13.75%	31.04%	5.31%	4.59%	16.33%
Colima	67.02%	13.68%	17.98%	13.36%	51.94%	9.09%	9.56%	19.53%
Chiapas	89.66%	47.06%	29.96%	18.84%	81.21%	28.94%	54.57%	25.04%
Chihuahua	55.86%	7.36%	15.80%	11.94%	39.64%	4.96%	4.21%	12.92%
Ciudad de México	54.46%	5.79%	8.91%	19.07%	45.19%	4.85%	1.67%	9.92%
Durango	67.37%	11.29%	16.01%	15.37%	50.49%	5.70%	11.67%	17.81%
Guanajuato	73.79%	16.52%	20.28%	14.17%	56.88%	8.77%	12.77%	26.14%
Guerrero	91.86%	50.22%	25.40%	16.44%	76.68%	33.44%	59.36%	40.65%
Hidalgo	80.75%	21.47%	17.60%	18.29%	69.22%	9.71%	24.49%	21.57%
Jalisco	63.61%	12.98%	18.78%	17.13%	49.06%	7.74%	9.55%	15.23%
México	68.55%	16.00%	13.19%	19.91%	57.70%	9.96%	10.21%	20.19%
Michoacán de Ocampo	83.07%	32.84%	27.20%	23.12%	69.97%	16.24%	26.47%	32.04%
Morelos	75.25%	21.76%	17.18%	16.06%	61.21%	13.75%	22.16%	26.19%
Nayarit	70.21%	15.43%	15.95%	15.97%	56.57%	7.54%	16.84%	19.50%
Nuevo León	47.78%	6.57%	11.55%	12.34%	30.94%	4.74%	4.22%	15.55%
Oaxaca	88.27%	45.78%	27.92%	16.86%	76.94%	20.89%	59.46%	32.49%
Puebla	82.87%	26.90%	21.91%	18.92%	71.25%	13.83%	26.96%	23.95%
Querétaro	64.40%	14.63%	14.92%	12.87%	50.51%	8.62%	17.66%	17.46%
Quintana Roo	69.77%	21.00%	15.31%	16.48%	52.11%	18.06%	18.85%	22.24%
San Luis Potosí	70.90%	20.93%	17.80%	9.83%	56.22%	14.09%	29.92%	16.85%
Sinaloa	67.84%	14.06%	15.78%	14.69%	49.11%	8.23%	14.05%	23.31%
Sonora	59.30%	11.87%	12.55%	13.33%	37.71%	8.78%	9.93%	23.67%
Tabasco	87.18%	35.17%	17.80%	17.10%	70.47%	11.64%	46.13%	39.79%
Tamaulipas	63.73%	12.61%	16.21%	15.02%	46.12%	8.94%	10.44%	19.61%
Tlaxcala	77.09%	14.92%	13.84%	14.84%	66.95%	10.76%	11.06%	21.51%
Veracruz de Ignacio de la Llave	84.00%	35.60%	26.03%	19.21%	68.33%	19.48%	40.88%	30.09%
Yucatán	71.30%	27.66%	21.81%	14.91%	52.70%	18.24%	38.86%	16.87%
Zacatecas	75.18%	13.22%	19.10%	12.37%	61.59%	4.29%	14.43%	20.03%

Tabla 4.6

Pobreza multidimensional. Población con carencias en MCS 2015 demográfico

Entidad federativa	Población con al menos una carencia	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Total	70.57%	20.43%	17.77%	16.87%	56.55%	12.05%	20.38%	21.61%
Aguascalientes	58.86%	8.52%	13.74%	13.68%	42.90%	4.84%	4.69%	18.80%
Baja California	62.96%	12.25%	13.29%	17.94%	46.58%	9.26%	12.15%	14.71%
Baja California Sur	62.84%	14.54%	14.00%	12.84%	43.17%	14.07%	11.49%	21.76%
Campeche	74.12%	27.54%	20.17%	12.35%	57.69%	18.39%	34.17%	23.23%
Coahuila de Zaragoza	49.61%	7.35%	12.08%	13.62%	30.87%	5.09%	4.49%	16.51%
Colima	67.24%	13.77%	17.61%	13.57%	52.35%	9.17%	9.63%	19.53%
Chiapas	89.69%	47.07%	29.95%	18.87%	81.25%	28.98%	54.56%	25.03%
Chihuahua	56.00%	7.32%	15.72%	11.94%	39.75%	4.94%	4.21%	12.97%
Ciudad de México	54.56%	5.82%	8.86%	19.09%	45.33%	4.85%	1.72%	9.85%
Durango	67.32%	11.31%	15.93%	15.38%	50.54%	5.67%	11.63%	17.79%
Guanajuato	73.85%	16.56%	20.34%	14.49%	57.00%	8.94%	12.61%	25.85%
Guerrero	92.06%	50.59%	25.08%	16.60%	77.01%	33.63%	60.07%	40.56%
Hidalgo	80.06%	21.34%	17.42%	18.33%	68.66%	9.56%	24.29%	21.40%
Jalisco	63.44%	12.84%	18.67%	17.14%	48.91%	7.69%	9.33%	15.25%
México	68.73%	15.95%	13.16%	19.82%	57.86%	10.09%	10.14%	20.22%
Michoacán de Ocampo	82.80%	32.13%	26.45%	22.99%	69.81%	15.98%	26.04%	31.44%
Morelos	74.82%	21.52%	16.73%	16.21%	60.88%	13.59%	21.49%	26.07%
Nayarit	70.27%	15.46%	15.92%	15.92%	56.68%	7.58%	16.87%	19.55%
Nuevo León	48.01%	6.55%	11.33%	12.55%	31.46%	4.49%	4.35%	15.50%
Oaxaca	88.18%	45.51%	27.22%	17.04%	76.98%	20.69%	59.20%	32.23%
Puebla	82.94%	26.89%	21.73%	18.87%	71.26%	13.82%	27.30%	23.91%
Querétaro	63.87%	14.47%	14.71%	13.03%	50.14%	8.47%	16.95%	17.54%
Quintana Roo	69.52%	20.78%	14.97%	16.35%	52.18%	18.00%	18.75%	21.75%
San Luis Potosí	71.18%	20.91%	17.68%	10.17%	56.45%	14.05%	29.90%	16.86%
Sinaloa	67.90%	14.04%	15.71%	14.64%	49.19%	8.26%	14.05%	23.24%
Sonora	59.36%	11.94%	12.57%	13.37%	37.77%	8.79%	9.98%	23.71%
Tabasco	87.22%	35.21%	17.81%	17.08%	70.48%	11.66%	46.21%	39.82%
Tamaulipas	63.83%	12.52%	16.12%	14.96%	46.22%	8.92%	10.44%	19.57%
Tlaxcala	76.92%	14.81%	13.72%	14.86%	66.81%	10.61%	11.02%	21.30%
Veracruz de Ignacio de la Llave	82.82%	34.97%	25.44%	18.91%	66.95%	20.33%	40.14%	29.06%
Yucatán	71.52%	27.83%	21.77%	14.84%	53.05%	18.33%	39.22%	16.87%
Zacatecas	75.13%	13.21%	19.08%	12.39%	61.55%	4.28%	14.44%	20.01%

5. Validación y evaluación de la metodología

Aunque el ajuste demográfico no tiene gran impacto en la disminución del ingreso y aumento de la pobreza del MCS 2015, sí elimina el efecto que el problema demográfico pudiera tener en éstos, es decir, este ejercicio elimina los posibles efectos que el comportamiento demográfico podría representar para que el MCS 2015 arrojará ingresos más altos que lo que se venían captando en los eventos anteriores. Así, pone el piso base para que se realicen ejercicios de alineación a los eventos anteriores.

Sin duda, los métodos de imputación *Deck* arrojan resultados satisfactorios cuando existe un conjunto muy grande de posibles donadores como en este ejercicio. Sin embargo, cuando el conjunto de posibles donadores es pequeño, para un gran número de casos existe la posibilidad de que se encuentren el o los mismos donadores, eliminando poco a poco la variabilidad en las observaciones y como consecuencia compactar las varianzas.

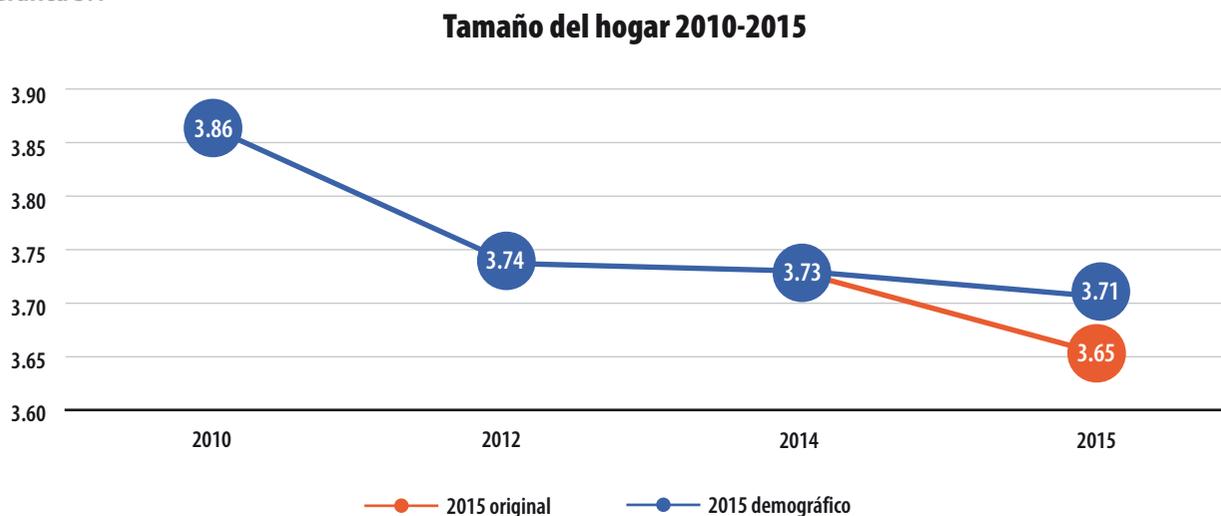
Ahora bien, para este ejercicio en específico, los resultados fueron muy satisfactorios. Si observamos la gráfica 5.1, notamos que se cumple con el objetivo de incrementar el tamaño del hogar del MCS 2015, al pasar de 3.65 a 3.71. Además, el total de ho-

gares lo deja en 32 681 856 y no los 33 218 037 del MCS 2015 original (536 181 hogares menos), dejando la tasa de crecimiento de éstos entre el 2014 y 2015 en 1.65, muy por debajo del 3.3 del original.

Por otro lado, en la gráfica 5.2 se presenta la descomposición del tamaño del hogar en grandes grupos de edad, pudiendo observar que el ejercicio se acerca a la tendencia histórica del MCS, sobre todo en el grupo de 12 a 64 años, que es donde se presenta el desajuste del MCS 2015 original. También se comporta mejor que la EIC 2015 y que un ejercicio de posestratificación del MCS 2015 con *Stata* (teniendo como base las variables total de integrantes del hogar, edad y sexo de la jefa o del jefe del hogar que reportó la EIC 2015), ya que estos dos últimos rejuvenecen la población.

Así como en el ingreso, en las carencias no hay grandes efectos de que preocuparse. Como se observa en el cuadro 5.3, las diferencias en las entidades federativas entre el MCS 2015 original y el MCS 2015 demográfico van desde -1.38 hasta 0.86 por ciento. Aunque estas diferencias mínimas no necesariamente pueden ser aplicables a todas las variables, para conocer los efectos a profundidad del ajuste demográfico es necesario hacer un análisis sobre aquellas que se puedan considerar las más importantes del MCS 2015.

Gráfica 5.1



Gráfica 5.2

Descomposición del tamaño del hogar por grandes grupos de edad 2008-2015

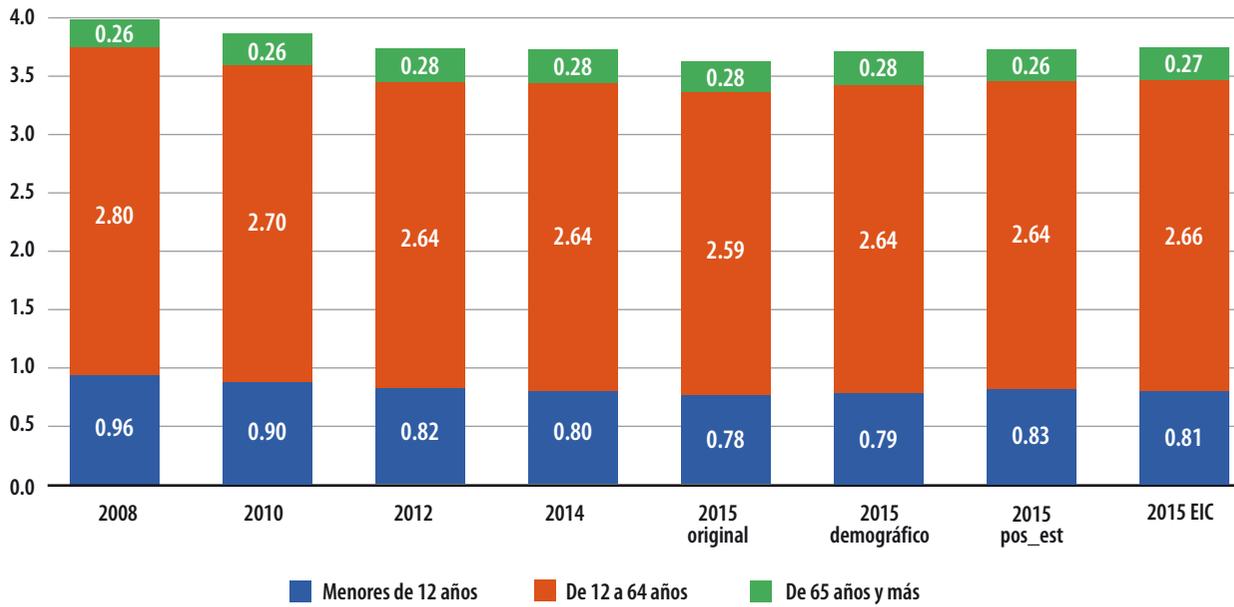


Tabla 5.1

Pobreza multidimensional. Diferencias 2015 demográfico vs. 2015 original, medidas entre las entidades federativas

Medida	Población con al menos una carencias	Población con al menos tres carencias	Población con carencia por rezago educativo	Población con carencia por acceso a servicios de salud	Población con carencia por acceso a la seguridad social	Población con carencia por calidad y espacios de la vivienda	Población con carencia de acceso a servicios básicos de la vivienda	Población con carencia por acceso a la alimentación
Mínimo	-1.18	-0.71	-0.74	-0.30	-1.38	-0.26	-0.73	-1.02
Máximo	0.27	0.37	0.06	0.33	0.52	0.86	0.71	0.18
Media	-0.06	-0.07	-0.18	0.02	0.00	-0.01	-0.07	-0.10
Desviación estándar	0.30	0.20	0.20	0.14	0.33	0.19	0.29	0.23
Mediana	0.05	-0.03	-0.11	0.01	0.06	-0.03	-0.02	-0.02

6. Conclusión y comentarios finales

Los resultados del ejercicio logran el objetivo de incrementar el tamaño del hogar de acuerdo con las tasas de variación implícitas calculadas a partir de las proyecciones de hogares y población del CONAPO y como consecuencia disminuir el total de hogares que reportó el MCS 2015.

Los efectos en el ingreso y la pobreza son marginales, ya que logra disminuir muy poco el monto del ingreso corriente y aumentar, también muy poco, al promedio de ingreso por hogar (debido a la disminución de los hogares) y la pobreza, esto respecto al original.

Las restricciones en cuanto a sexo, edad, tamaño de localidad y estrato socioeconómico, aunado a la aleatoriedad del ejercicio, logran una distribución consistente de la población para esas variables.

Probablemente existan otros métodos de imputación más efectivos que los *Deck*, pero es un hecho que éstos son los más simples y fáciles de im-

plementar, y sus resultados se pueden considerar satisfactorios. Sin embargo, es posible explorar en la literatura otros métodos más efectivos y compararlos con éstos, en estudios futuros.

7. Referencias

- Dempster, A. P. y D. B. Rubin. *Incomplete data in sample surveys*. Academic Press, New York, NY, pp.3-10, 1983.
- Juárez Alonso, Carlos Alberto. *Fusión de datos: imputación y validación*. Tesis doctoral, Universidad Politécnica de Cataluña, 2004.
- Kim, J. K. y W. Fuller. "Fractional hot deck imputation", en: *Biometrika*. 91(3):559-578, 2004.
- Palacios Ostria, Margot Alejandra y Edgar Javier González Liceaga. *Metodología de Fellegi y Holt: validación e imputación de datos*. Tesis profesional, UNAM, 2004.
- Schafer, J. L. *The multiple imputation FAQ page*. (DE) consultada el 9 de enero de 2017 en http://www.stat.ufl.edu/~athienit/STA6167/Missing%20Data/MI_FAQ.pdf
- United Nations Statistical Commission and Economic Commission for Europe. *Glossary of terms on statistical data editing*, UNECE, 2000. (DE) consultada el 9 de enero de 2017 en https://webgate.ec.europa.eu/fpfis/mwikis/essvalidserv/images/3/37/UN_editing_glossary.pdf

8. Anexos

Tabla 8.1

VARIABLES DEMOGRÁFICAS

Continúa

	Total	Estrato				Tamaño de localidad			
		Bajo	Medio bajo	Medio alto	Alto	100 mil o más	15 mil a 99 999	2 500 a 14 999	Menos de 2 500
Integrantes del hogar									
2014	119 897 652	23 889 861	60 786 821	24 819 266	10 401 704	57 207 724	17 627 470	17 260 987	27 801 471
2015 original	121 381 730	24 017 970	62 122 995	24 673 826	10 566 939	57 932 027	17 866 706	17 470 197	28 112 800
2015 demográfico	121 381 494	24 025 184	61 861 654	24 823 977	10 670 679	57 931 828	17 866 097	17 470 402	28 113 167

Tabla 8.1

Concluye

Variables demográficas

	Total	Estrato				Tamaño de localidad			
		Bajo	Medio bajo	Medio alto	Alto	100 mil o más	15 mil a 99 999	2 500 a 14 999	Menos de 2 500
Menores de 12 años									
2014	25 865 041	6 159 128	13 433 923	4 463 723	1 808 267	11 140 730	3 961 036	3 883 818	6 879 457
2015 original	25 874 659	6 222 485	13 607 517	4 253 593	1 791 064	10 948 270	3 928 672	4 036 002	6 961 715
2015 demográfico	25 951 155	6 211 846	13 625 007	4 290 008	1 824 294	10 980 741	3 981 272	4 032 474	6 956 668
12 a 64 años									
2014	84 970 122	15 944 673	43 088 074	18 200 206	7 737 169	41 785 936	12 469 412	12 065 644	18 649 130
2015 original	86 079 732	15 926 893	44 018 306	18 170 193	7 964 340	42 510 060	12 642 329	12 129 427	18 797 916
2015 demográfico	86 138 251	15 968 263	43 839 374	18 299 244	8 031 370	42 520 713	12 631 671	12 154 212	18 831 655
65 años y más									
2014	9 062 489	1 786 060	4 264 824	2 155 337	856 268	4 281 058	1 197 022	1 311 525	2 272 884
2015 original	9 427 339	1 868 592	4 497 172	2 250 040	811 535	4 473 697	1 295 705	1 304 768	2 353 169
2015 demográfico	9 292 088	1 845 075	4 397 273	2 234 725	815 015	4 430 374	1 253 154	1 283 716	2 324 844
Hombres									
2014	58 366 324	11 790 126	29 653 461	11 879 887	5 042 850	27 695 664	8 492 877	8 406 539	13 771 244
2015 original	59 099 914	11 773 368	30 319 001	11 904 056	5 103 489	28 136 511	8 563 786	8 529 336	13 870 281
2015 demográfico	59 119 189	11 768 571	30 211 132	11 986 030	5 153 456	28 146 732	8 577 120	8 524 750	13 870 587
Mujeres									
2014	61 531 328	12 099 735	31 133 360	12 939 379	5 358 854	29 512 060	9 134 593	8 854 448	14 030 227
2015 original	62 281 816	12 244 602	31 803 994	12 769 770	5 463 450	29 795 516	9 302 920	8 940 861	14 242 519
2015 demográfico	62 262 305	12 256 613	31 650 522	12 837 947	5 517 223	29 785 096	9 288 977	8 945 652	14 242 580

Tabla 8.2a

Insumos para la descomposición del ICT

Decil	Ingreso corriente total ^a			Hogares (Hog)			Perceptores			Ingreso por perceptor (IPP)			Perceptores por hogar (PH)		
	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón	2014	2015 original	Razón
	Total	1 318 162	1 524 263	1.156	32 150 400	33 218 037	1.033	75 381 871	80 911 445	1.073	17 486	18 839	1.077	2.34	2.44
I	19 662	27 137	1.380	3 215 040	3 321 804	1.033	5 925 336	6 226 511	1.051	3 318	4 358	1.313	1.84	1.87	1.017
II	36 680	45 939	1.252	3 215 040	3 321 804	1.033	6 725 440	6 778 853	1.008	5 454	6 777	1.243	2.09	2.04	0.976
III	50 224	61 079	1.216	3 215 040	3 321 804	1.033	6 856 385	7 219 234	1.053	7 325	8 461	1.155	2.13	2.17	1.019
IV	63 588	76 890	1.209	3 215 040	3 321 804	1.033	7 166 736	7 563 063	1.055	8 873	10 167	1.146	2.23	2.28	1.021
V	78 976	94 598	1.198	3 215 040	3 321 804	1.033	7 320 377	7 934 187	1.084	10 789	11 923	1.105	2.28	2.39	1.049
VI	97 024	115 818	1.194	3 215 040	3 321 804	1.033	7 720 394	8 203 636	1.063	12 567	14 118	1.123	2.40	2.47	1.028
VII	120 349	142 938	1.188	3 215 040	3 321 804	1.033	8 046 727	8 739 751	1.086	14 956	16 355	1.094	2.50	2.63	1.051
VIII	154 661	179 838	1.163	3 215 040	3 321 804	1.033	8 492 589	9 207 690	1.084	18 211	19 551	1.072	2.64	2.77	1.049
IX	213 831	243 329	1.138	3 215 040	3 321 804	1.033	8 650 216	9 558 109	1.105	24 720	25 458	1.030	2.69	2.88	1.069
X	483 167	536 696	1.111	3 215 040	3 321 804	1.033	8 477 671	9 480 410	1.118	56 993	56 611	0.993	2.64	2.85	1.082

Nota: se incluye solo a los perceptores de ingreso monetario, excluyendo a quienes reportan ingreso no monetario.

^a Millones de pesos constantes a precios del 2015.

Tabla 8.2b

Insumos para la descomposición del ICT

Decil	Ingreso corriente total ^a			Hogares (Hog)			Perceptores			Ingreso por perceptor (IPP)			Perceptores por hogar (PH)		
	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón	2014	2015 demográfico	Razón
	Total	1 318 162	1 504 051	1.141	32 150 400	32 681 856	1.017	75 381 871	79 496 030	1.055	17 486	18 920	1.082	2.34	2.43
I	19 662	26 747	1.360	3 215 040	3 268 186	1.017	5 925 336	6 115 158	1.032	3 318	4 374	1.318	1.84	1.87	1.015
II	36 680	45 238	1.233	3 215 040	3 268 186	1.017	6 725 440	6 674 012	0.992	5 454	6 778	1.243	2.09	2.04	0.976
III	50 224	60 103	1.197	3 215 040	3 268 186	1.017	6 856 385	7 101 789	1.036	7 325	8 463	1.155	2.13	2.17	1.019
IV	63 588	75 756	1.191	3 215 040	3 268 186	1.017	7 166 736	7 445 925	1.039	8 873	10 174	1.147	2.23	2.28	1.022
V	78 976	93 221	1.180	3 215 040	3 268 186	1.017	7 320 377	7 765 773	1.061	10 789	12 004	1.113	2.28	2.38	1.044
VI	97 024	114 213	1.177	3 215 040	3 268 186	1.017	7 720 394	8 064 370	1.045	12 567	14 163	1.127	2.40	2.47	1.028
VII	120 349	141 016	1.172	3 215 040	3 268 186	1.017	8 046 727	8 595 184	1.068	14 956	16 406	1.097	2.50	2.63	1.051
VIII	154 661	177 390	1.147	3 215 040	3 268 186	1.017	8 492 589	9 051 806	1.066	18 211	19 597	1.076	2.64	2.77	1.049
IX	213 831	240 004	1.122	3 215 040	3 268 186	1.017	8 650 216	9 372 792	1.084	24 720	25 606	1.036	2.69	2.87	1.066
X	483 167	530 363	1.098	3 215 040	3 268 186	1.017	8 477 671	9 309 222	1.098	56 993	56 972	1.000	2.64	2.85	1.080

Nota: se incluye solo a los perceptores de ingreso monetario, excluyendo a quienes reportan ingreso no monetario.

^a Millones de pesos constantes a precios del 2015.

Ajuste del MCS 2015 por alineación de deciles

Adjustment of MCS 2015
by means of decile
alignment

Rodrigo Negrete Prieto* y Benito Durán Romo**



Bar chart made if sandwiches / Tim Macpherson/Getty Images

Es un hecho que en los microdatos del Módulo de Condiciones Socioeconómicas (MCS) 2015 hubo una importante disminución de registros (hogares) con ingresos inferiores a 3 mil pesos trimestrales (mil mensuales). Ello tuvo un impacto en la comparación de los deciles de dicho levantamiento con el que le precedió al producir una composición híbrida de los mismos dado que, en la parte superior de cada decil del 2015 quedaron algunos que hubieran pertenecido al decil $n + 1$ de no haber ocurrido el colapso de los fractiles de ingresos más bajos en el primer decil (corrimiento hacia abajo de todos los fractiles). La metodología que aquí se propone tiene como punto de partida una compensación del primer decil (el cual experimentó una pérdida de piso respecto al 2014) mediante un procedimiento que acerca más la composición de todos los deciles MCS 2015 a los de 2014, obteniéndose con ello una moderación de los resultados del levantamiento del 2015.

El procedimiento deja una base de microdatos ajustada de donde se desprende una evolución del ingreso corriente total, así como de la pobreza multidimensional sensiblemente más cercana a la de la serie histórica, ello sin dejar de acusar que el levantamiento del 2015 presentó una mayor captación de ingreso corriente de lo que venía realizando, de modo que el resultado de moderación del 2015 no se parece a una mera proyección del pasado.

Palabras clave: deciles; hibridación de deciles; *decilización*; *peranzanas*; fractiles; colapso de frecuencias; Módulo de Condiciones Socioeconómicas 2015; ENIGH; factores de expansión; microdatos; ingreso corriente total; ajuste *Delta*; pobreza; pobreza multidimensional; pobreza extrema; pobreza moderada; carencias sociales; no pobres y no vulnerables.

It is a fact that on regard the Socioeconomic Conditions Module 2015 micro data took place an important diminishing amongst households with a quarterly current income less than 3th pesos. In turn this has an impact in terms of the whole decile composition making them hybrid so to speak on regard previous rounds' decile structure because the upper part of each ended up with a segment or fractile belonging the next decile, had not been taken place the collapse of the left side of the distribution. Taking this into account the methodology here advanced has as its starting point a compensation of the first decile which of all them was the one most affected in its inner composition. This is done by means a procedure at the end of which decile's composition of 2015 gets closer to the one of 2014 an with an outcome of moderation in terms of income growth between those rounds.

Even though from the new microdata thus obtained moderation effects are evident, some properties of the 2015's round signaling some improvement in collecting data are not lost, so the adjustment does not make that round a mere projection of the past.

Key words: decile hybridation; pearapples; fractiles; frequency collapse; Socioeconomic Conditions Module 2015; Income-Expenditure National Survey; weighting factors; microdata on current income; Delta.

* Instituto Nacional de Estadística y Geografía (INEGI), rodrigo.negrete@inegi.org.mx

** INEGI, benito.duran@inegi.org.mx

Nota: los autores agradecen la participación de Lilia Guadalupe Luna Ramírez y Ana Miriam Romo Anaya en el proceso de investigación e integración de este documento.

1. Introducción

De las mayores perplejidades que se desprenden del Módulo de Condiciones Socioeconómicas (MCS) 2015 no son solo los incrementos en la captación del ingreso corriente total (ICT) que presenta, sino que, además, estos incrementos resultan en particular acusados en los tres primeros deciles; el primero, más que ninguno. A lo largo del proceso de análisis y discusión de estos resultados se ha llegado a atribuir que este efecto no nada más obedece a que magnitudes pequeñas son más susceptibles de presentar variaciones mayores, sino que esas variaciones resultaron de esfuerzos encaminados a revisar la consistencia y justificaciones de esos ingresos y que, en consecuencia, extrajeron más información de ingreso corriente de tales hogares, empeño del que no fueron objeto, en la misma medida, el resto de ellos.

Una reflexión pertinente a la que no se le había puesto suficiente atención es al hecho de que los deciles del MCS 2015 no son exactamente comparables con los del levantamiento que le precedió, y no lo son porque en dicho levantamiento se perdieron o disminuyeron de manera sustantiva registros con los niveles más bajos de ingresos, como los que presentaban menos de 3 mil pesos al trimestre. En ello pudo influir, desde luego, el efecto conjunto de instrucciones, como las C14, C326 y C327 del *Manual del entrevistador* que obligaban al operativo a justificar satisfactoriamente los casos de ingresos tan bajos que la subsistencia misma del hogar así reportado se volvía difícil de comprender. Si se suma el efecto que tuvo una mayor captación de fuentes de ingreso por hogares, esto terminó afectando no solo al primer decil sino en sí a toda la *decilización* de los hogares en el 2015: algunos hogares que, en el 2014 hubieran quedado en el decil $n + 1$ se situaron en el 2015 en el decil anterior (n), introduciendo cambios estructurales. Así, por ejemplo, la mayor composición asalariada del decil I no deja de acusar la presencia de hogares que, en otras circunstancias, hubieran quedado en el decil II, pero ahora son parte de su formación (el primer decil normalmente es menos asalariado que el segundo y éste, que el tercero, ya que

el autoempleo, así como las transferencias, tienen mayor presencia a esos niveles en sus estrategias de supervivencia).

Esta hibridación de los deciles en el MCS 2015 hace entonces que, *vis a vis* con los del 2014, se estuvieran comparando, en alguna medida, *peras* (2014) con *manzanas* (2015) o, dicho más precisamente, con *peranzanas*: deciles donde una parte de los hogares es análoga a los del mismo decil en el 2014 (*peras*), pero que incluye en su composición hogares que hubieran correspondido al decil superior en tal año (*manzanas*).

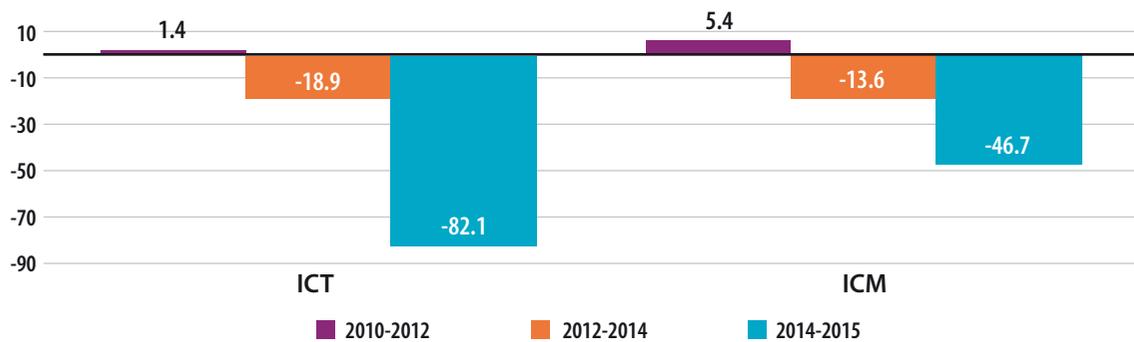
La metodología que aquí se plantea parte de neutralizar el tramo *manzana* de cada *peranzana* en el 2015, de modo tal que el primer decil sea el más compensado de todos. Cabe señalar que por neutralizar se entiende aquí dar un factor de expansión igual a 1, y compensar quiere decir que en los segmentos *pera* en dirección descendente se reparten los factores de expansión de esos registros neutralizados, ello bajo criterios de misma pertenencia a entidad, tamaño de localidad y estrato y de manera tal que, el primer decil se convierta en un receptor neto de expansión compensatoria.

En la sección *Marco teórico* se ilustrarán estas ideas y conceptos de deciles híbridos (*peranzanas*) que dieron lugar al procedimiento, mismo que se detalla en *Metodología*. Los resultados así obtenidos de los microdatos se presentan en la sección correspondiente en dos tiempos: primero, antes del ajuste demográfico al tamaño del hogar en función de las variaciones 14-15 que, al respecto, se desprenden de las proyecciones del Consejo Nacional de Población (CONAPO) —ver el documento de esta nueva serie que detalla el procedimiento de ajuste demográfico por la vía de la imputación— y, enseguida, el resultado que añade dicho ajuste. A efectos solo de este documento, al procedimiento de neutralización-compensación para colapsar el tramo *manzana* y sobreexpandir el tramo *pera* de los deciles se le llamará *Delta*. Así, en la primera sección de resultados se muestra solo la fase puramente *Delta* del ejercicio, mientras que en la segunda, donde se ilustran los datos finales que

Gráfica 2.1

Hogares con ingresos menores o iguales a 3 mil pesos trimestrales

(variación anual porcentual; datos muestrales a precios constantes del 2015)



arroja el ejercicio, corresponden a un ajuste *Delta* seguido de un ajuste demográfico por imputación.

2. Marco teórico

El punto de partida de todo el ejercicio es una reflexión sobre cómo cambia la *decilización* de un evento —en este caso del MCS 2015— cuando colapsa el número de registros (viviendas u hogares) que en otras circunstancias formarían parte de los fractiles iniciales, es decir, los de ingreso corriente total más bajo.

A este respecto, se debe tener presente que el número de registros muestrales en el 2015 con un ICT inferior a 3 mil pesos trimestrales disminuyó

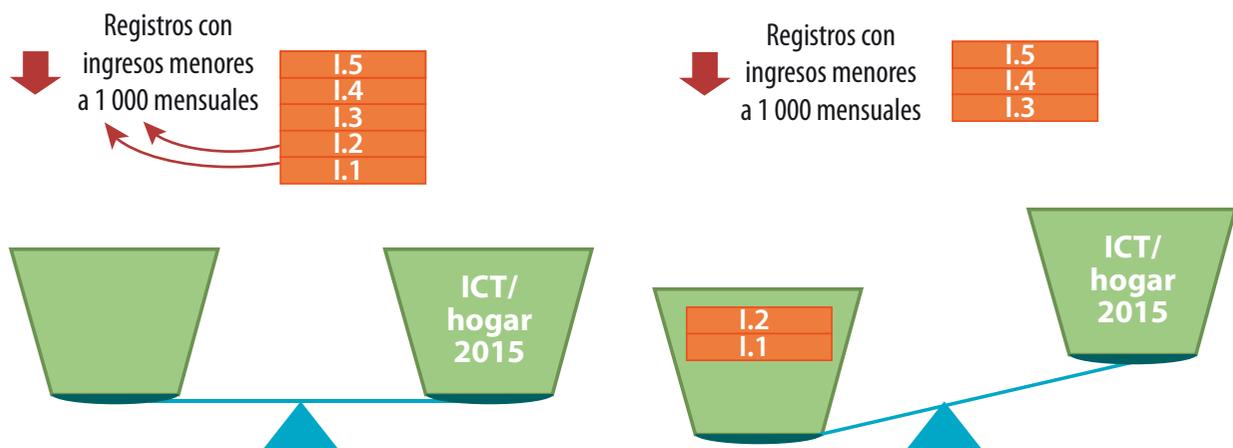
82.1% respecto al del 2014 (de 520 a 93 registros) y 46.7% en términos de ingreso corriente monetario (ICM), es decir, de 2 394 a 1 275. Ello en un marco en el que ya se venía registrando de suyo una tendencia a la disminución de estos casos, aunque no de una forma tan pronunciada como ocurrió en el 2015. Como sea, si esta pérdida introduce dificultades al comparar con el 2014, con mayor razón lo hace respecto a años previos (ver gráfica 2.1).

En la ilustración 2.1 se muestra cómo el colapso de lo que serían los primeros fractiles del primer decil tiene por efecto inmediato incrementar su ingreso promedio.

Empero, el efecto en el primer decil no se limita a eso, dada la ausencia de *piso*; en el 2015, para com-

Ilustración 2.1

Desplazamiento de fractiles y efecto en el ingreso promedio



pletarse, termina incorporando fractiles que, de otro modo, hubieran correspondido al decil II. Así, en términos de su comparabilidad con respecto al 2014, se forma un decil *peranzana* (ver ilustración 2.2).

Si bien el decil I fue el único afectado en forma doble (por abajo y por arriba) lo que, desde luego, le imprime una variabilidad muy brusca en su comparación respecto al 2014, todos los deciles subsiguientes también fueron afectados por este corrimiento de los fractiles; lo que se presenta en la ilustración 2.3 para los deciles III y VIII, al igual

de lo que se hizo para el decil I, es en proporciones aproximadas: los montos exactos se muestran en la primera matriz en la ilustración 3.1.

Una manera alternativa de expresar el desfaseamiento de los deciles sería visualizando la *deci-lización* del 2015 como si, en términos del ICT, no hubiera comenzado desde un piso, sino desde un punto más alto (ver en la ilustración 2.4 la diferencia entre las dos torres que corresponden al 2015): la de más a la derecha sería la torre 2014 montada sobre los factores exógenos, esto es, el contexto económico y/o las medidas tomadas de políticas

Ilustración 2.2

Representación de las *peranzanas* para el decil I

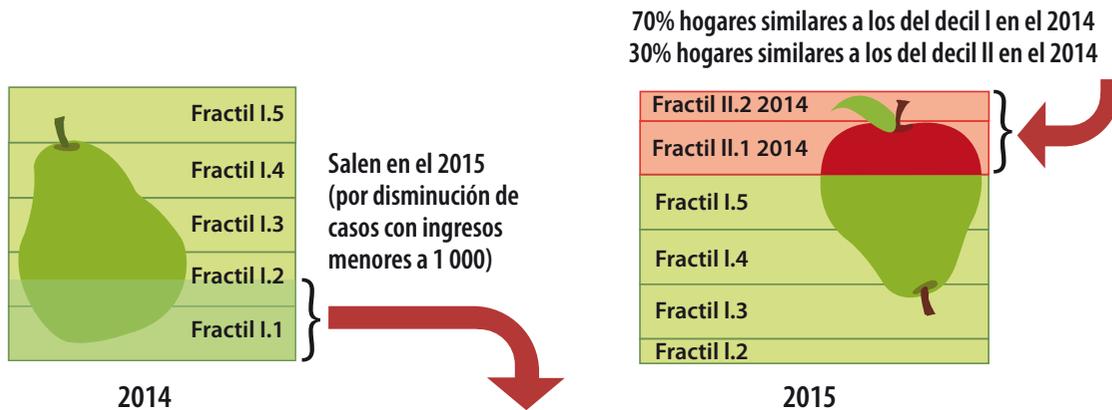
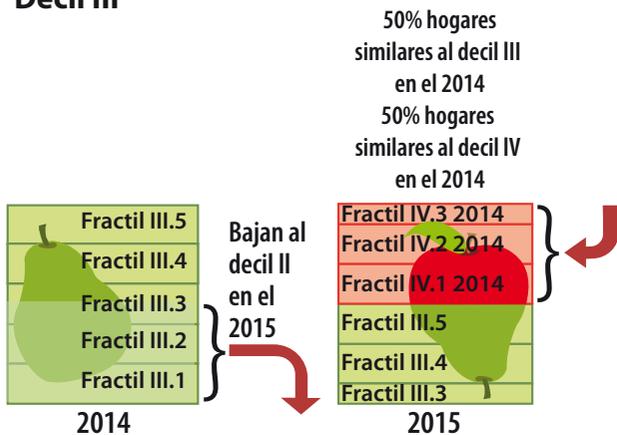


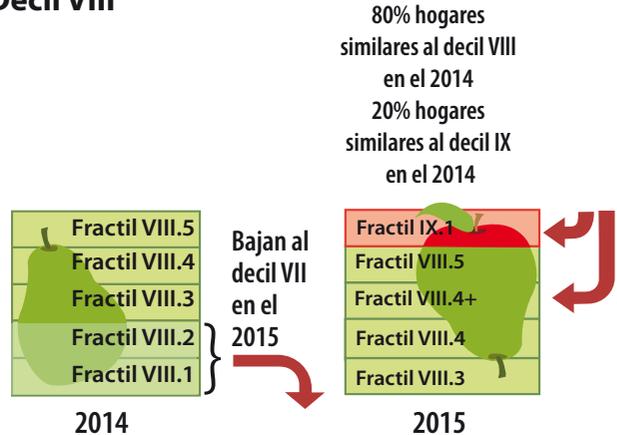
Ilustración 2.3

Representación de las *peranzanas* para los deciles III y VIII

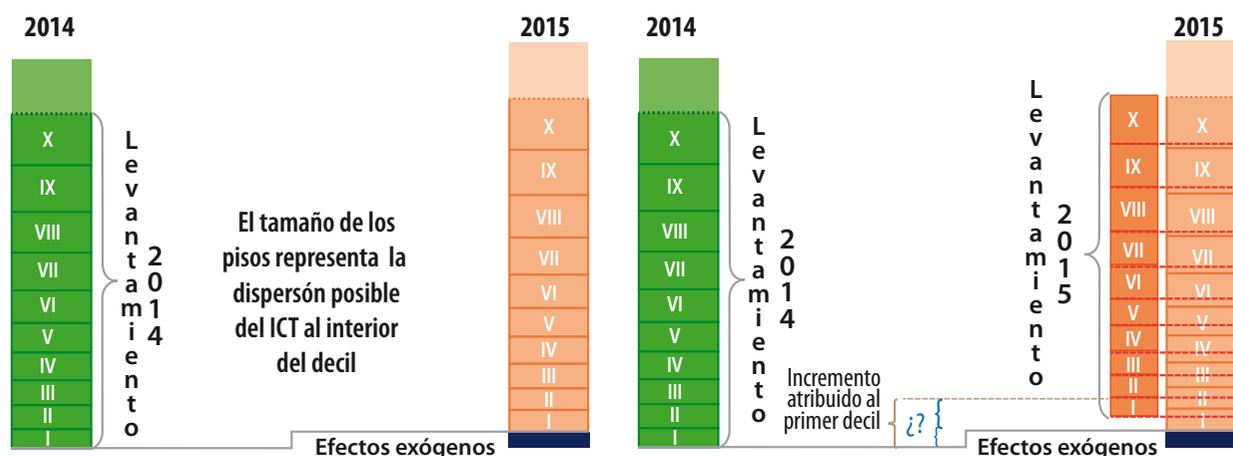
Decil III



Decil VIII



Esquema de comparación del ICT entre el MCS del 2014 y el 2015



públicas que hacen alguna diferencia entre ambos años y le dan un piso o escalón al MCS 2015, pero el problema es que la torre que le corresponde en realidad comienza en un punto más arriba de esa plataforma. Claramente, en la formación del primer decil quedan hogares que también hubieran sido del decil I en el 2014, pero incorpora otros que habrían pertenecido al decil superior y esto sucede, asimismo, en todos los demás deciles.

De lo que se trata, entonces, es darle de nuevo un piso a ese edificio; es decir, que su elevación respecto al 2014 obedezca, dentro de lo posible, solo a factores exógenos y no a otros que tengan que ver con el levantamiento del 2015 y/o las decisiones tomadas para ese levantamiento. Una manera de hacerlo es sobreexpandiendo los registros que no fueron afectados por esas circunstancias, esto es, los que su asignación de decil hubiera sido la misma en el 2014 y 2015.

A continuación se explica, en primer término, cómo es que se identifican los tramos desfasados (*manzana*) dentro de cada decil 2015; luego, cómo se procedió a neutralizarlos y, finalmente, cómo se expandió el resto de la muestra bajo un esquema que tiene como objetivo principal fortalecer en particular al primer decil, que fue el más afectado de todos.

3. Metodología

3.1 Su aplicación

El primer paso en la metodología de corrección que parte del diagnóstico arriba esbozado es tratar de identificar en los deciles del 2015 el tramo en ellos que hubiera quedado en el decil superior de no haber colapsado los fractiles iniciales del ICT más bajo, esto es, el tramo *manzana* dentro de la *peranzana*.

El problema, desde luego, es que la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)/MCS no es una de panel de hogares, es decir, una en la que las mismas viviendas que se entrevistaron en el 2014 se hayan entrevistado de nuevo en el 2015. La Encuesta fue panel en otro sentido: 90% de las unidades primarias de muestreo se repitieron de un levantamiento a otro, mas no en sí las viviendas en ellas contenidas. ¿Cómo entonces saber qué hubiera sucedido de no haber ocurrido el colapso de frecuencias en los niveles más bajos del ICT?

Aquí se tratará de estimar esto formando un *pool* de los registros del 2014 y 2015 como si hubieran formado una sola muestra (los ICT del 2014 se llevan a precios del 2015, de modo que el efecto inflacionario no sea un factor que introduzca una

diferencia entre los registros de uno y otro año). Al hacer esto, el 2014 aporta los registros del ICT más bajos que colapsaron en el 2015 para evitar el corrimiento descendente de todos los fractiles.

Decilizado este pool, lo que interesa en particular es comparar la ubicación de los registros pertenecientes al 2015 con su ubicación en la *decilización* del MCS 2015. Aquí se asume que los tramos *manzana* serán aquellos que en el MCS 2015 quedaron en el decil *n*, pero que en el *pool* terminan siendo ubicados en el decil *n + 1*. Es, pues, el tramo en el decil del 2015 que descendió de ubicación por haberse modificado ciertas condiciones de levantamiento u operación en dicho año respecto al anterior.

En lo que sigue, por medio de una matriz se mostrará la ubicación de los registros del 2015, contrastando la pertenencia a deciles del MCS 2015 con la pertenencia a deciles del *pool*. Los renglones corresponden a los deciles del 2015, en tanto que las columnas, a los deciles del *pool* 14-15. Es así que en el MCS 2015 para el primer decil se tiene un total de 5 880 viviendas; de éstas, 4 587 no cambian su asignación de decil, pues en el *pool* siguen cayendo en el 1; sin embargo, 1 293 de esas viviendas en el *pool* quedaron clasificadas en el II. Generalizando, en rojo se marca el tramo *manzana* y en verde, el *pera* de cada decil o *peranzana* del MCS 2015 (ver ilustración 3.1).

Una vez que se tienen detectados los registros que forman el tramo *manzana* en la *decilización* del MCS 2015, se procede a neutralizarlos con factor de expansión = 1, de modo que no representen en la muestra a otra cosa que a sí mismos.

El factor de expansión de esos registros se asigna al tramo *pera* del decil inferior (*n - 1*) a manera de secuencia en cascada. Donadores y receptores deben pertenecer a la misma entidad federativa, igual tamaño de localidad y estrato socioeconómico.¹

Así pues, la asignación procede de la siguiente forma:

- Se busca correspondencia en entidad, localidad y estrato en el tramo *pera* del decil inferior (*n - 1*); de haberla, se neutralizan con factor 1 los registros *manzana* del decil *n* que no tienen problema de asignación y se traslada su factor de expansión.
- Si no hay correspondencia en el tramo *pera* del decil *n - 1*, se busca entonces correspondencia en el tramo *pera* del propio decil *n*; de haberla, se neutralizan con factor 1 los

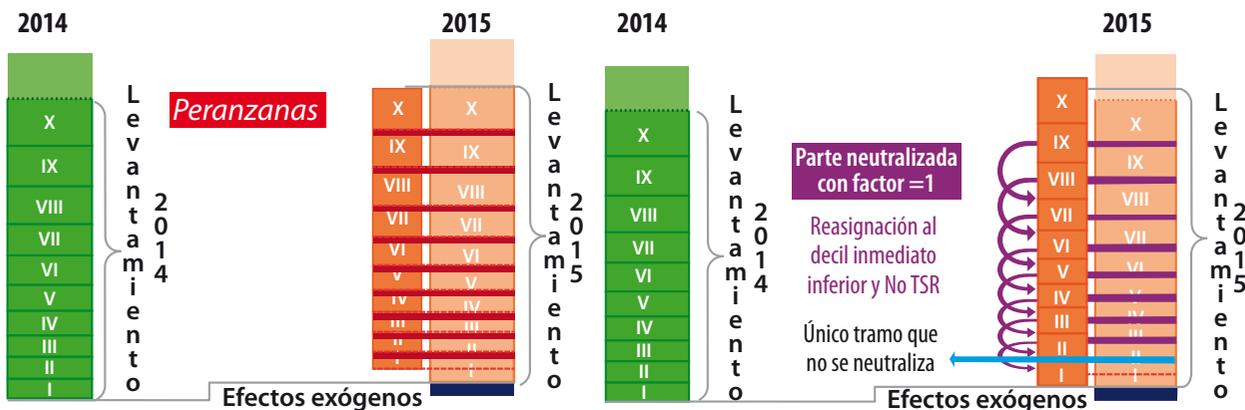
¹ Si bien no idéntico, el procedimiento es análogo a tratar como no respuesta a los registros neutralizados, pues sus factores de expansión se redistribuyen a tramos con los que guardan alguna afinidad, al menos en términos de entidad federativa, tamaño de localidad y estrato.

Ilustración 3.1

Matriz de formación de deciles

	Decil <i>pool</i> (MCS 2014 y MCS 2015)										Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	
I	4 587	1 293									5 880
II		4 061	1 607								5 668
III			3 763	1 924							5 687
IV				3 760	2 081						5 841
V					3 712	1 983					5 695
VI						3 900	1 902				5 802
VII							4 161	1 506			5 667
VIII								4 662	1 030		5 692
IX									5 269	497	5 766
X										5 496	5 496
Total	4 587	5 354	5 370	5 684	5 793	5 883	6 063	6 168	6 299	5 993	57 194

Esquema de comparación del ICT entre el MCS del 2014 y 2015



registros *manzana* del decil n que no tienen problema de asignación y se traslada el factor de expansión.

- Si no hay correspondencia ni en el tramo *manzana* del decil $n - 1$ ni tampoco en el del decil n , el o los registros dentro del tramo *manzana* no se neutralizan con factor 1, quedan con su factor original.

Bajo esa lógica de asignación, todos los deciles entre el II y el VIII donan factor de expansión hacia abajo y reciben de arriba; solo el IX es un donador neto y el I, un receptor neto (no hay que olvidar que este último decil en el MCS 2015 fue afectado de forma doble: por arriba pero también por abajo por el

colapso de los fractiles inferiores). El decil X, por su parte, ni dona, ni recibe (ver ilustración 3.2).

Realizando el procedimiento anterior, se *deciliza* de nuevo para comparar ahora cuántos registros de los microdatos del MCS 2015 (después de este primer ajuste) quedan en un decil $n + 1$ respecto a su *decilización* en el *pool* 2014-2015 (ilustración 3.3).

De la matriz mostrada, se observa que todavía quedaron registros desfasados (*manzanas*) en todos los deciles del I al IX.

Se itera de nuevo y se encuentra que los deciles II y III ya no muestran registros asignados al decil

Ilustración 3.3

Primera iteración

Decil MCS 2015	Decil <i>pool</i> (MCS 2014 y MCS 2015)										Total
	I	II	III	IV	V	VI	VII	VIII	IX	X	
I	3 942	375									4 317
II		4 258	96								4 354
III			5 119	210							5 329
IV				5 373	311						5 684
V					5 497	320					5 817
VI						5 660	418				6 078
VII							5 968	416			6 384
VIII								6 223	299		6 522
IX									6 466	531	6 997
X										5 712	5 712
Total	3 942	4 633	5 215	5 583	5 808	5 980	6 386	6 639	6 765	6 243	57 194

Ilustración 3.4

Segunda iteración

		Decil pool (MCS 2014 y MCS 2015)										Total
		I	II	III	IV	V	VI	VII	VIII	IX	X	
Decil MCS 2015	I	3 891	321									4 212
	II		4 131									4 131
	III		45	5 128								5 173
	IV				5 509	57						5 566
	V					5 676	41					5 717
	VI						5 904	117				6 021
	VII							6 315	143			6 458
	VIII								6 586	85		6 671
	IX									6 847	686	7 533
	X										5 712	5 712
Total		3 891	4 497	5 128	5 509	5 733	5 945	6 432	6 729	6 932	6 398	57 194

superior. Dado que ello interrumpe la secuencia en cascada (ya no pueden donar a su decil previo), se decide detener ahí el procedimiento (ilustración 3.4).

A todo este procedimiento de neutralización y asignación de factor de expansión de manera iterada se le denomina *Delta*, y es importante señalar que se realiza antes del ajuste demográfico. Ello es así porque no tiene caso alinear los crecimientos del tamaño del hogar a la proyección del CONAPO y luego desalinear lo así obtenido al redistribuir factores de expansión.

Por lo anterior, se presentan primero los resultados obtenidos por el procedimiento *Delta* antes del ajuste demográfico. Enseguida, se mostrarán los resultados a los que se añade tal ajuste.

3.2. Nota técnica

3.2.1 Notación matemática

A continuación, se presenta la notación matemática del procedimiento *Delta*.

Se declaran dos conjuntos muestras:

- ${}_{2015}H$ como el conjunto de hogares $\{1, 2, \dots, n_{2015}\}$ en la muestra de la base de datos 2015.
- ${}_{pool}H$ como el conjunto de hogares $\{1, 2, \dots, n_{pool}\}$ en la muestra de la base de datos construida a partir del agregado de las bases 2014 y 2015.

Y los siguientes índices:

Índice que denotará el decil: $j = 1, 2, \dots, 10$.

Índice que denotará la entidad: $k = 1, 2, \dots, 32$.

Índice que denotará la localidad: $l = 1, 2, 3, 4$.

Índice que denotará el estrato sociodemográfico: $e = 1, 2, 3, 4$.

Entonces:

${}_{2015}H_j^{kle}$ corresponde al conjunto de hogares del 2015 que están en el decil j , entidad k , localidad l y estrato e .

${}_{pool}H_j^{kle}$ corresponde al conjunto de hogares en el *pool*, que están en el decil j , entidad k , localidad l y estrato e .

Para definir conjuntos *pera* y *manzana* por decil se identifican las siguientes intersecciones: para cada $j = 1, 2, \dots, 10$; $k = 1, 2, \dots, 32$; $l = 1, 2, 3, 4$; $e = 1, 2, 3, 4$.

$P_j^{kle} = {}_{2015}H_j^{kle} \cap {}_{pool}H_j^{kle}$ tramo de hogares *pera* que pertenecen a la entidad k , localidad l y estrato e .

$M_j^{kle} = {}_{2015}H_j^{kle} \cap {}_{pool}H_{j+1}^{kle}$ tramo de hogares *manzana* que pertenecen a la entidad k , localidad l y estrato e .

De esta última definición se desprende que para $j = 10$ se tiene $M_{10} = \emptyset$.

Redistribución de pesos

Sean:

w_i = factor de expansión del hogar i ,

$$i \in \{1, 2, \dots, n_{2015}\}.$$

${}_{2015}w_i^*$ = nuevo factor de expansión del hogar i en la muestra 2015.

Para toda combinación $(j, k, l, e): j = 1, 2, \dots, 9; k = 1, 2, \dots, 32; l = 1, 2, 3, 4$ y $e = 1, 2, 3, 4$; mientras se cumpla que $M_{j+1}^{kle} \neq \emptyset$, los pesos se distribuirán de la siguiente manera:

Si $P_j^{kle} \neq \emptyset$

$${}_{2015}w_i^* = \begin{cases} w_i + \frac{w_i}{\sum_{m \in P_j^{kle}} w_m} * \sum_{s \in M_{j+1}^{kle}} (w_s - 1) & \forall i \in P_j^{kle} \\ 1 & \forall i \in M_{j+1}^{kle} \end{cases}$$

Si $P_j^{kle} = \emptyset$ y $P_{j+1}^{kle} \neq \emptyset$

$${}_{2015}w_i^* = \begin{cases} w_i + \frac{w_i}{\sum_{m \in P_{j+1}^{kle}} w_m} * \sum_{s \in M_{j+1}^{kle}} (w_s - 1) & \forall i \in P_{j+1}^{kle} \\ 1 & \forall i \in M_{j+1}^{kle} \end{cases}$$

Si $P_j^{kle} = \emptyset$ y $P_{j+1}^{kle} = \emptyset$, entonces:

$${}_{2015}w_i^* = w_i \quad \forall i \in M_{j+1}^{kle}$$

Una vez realizada la redistribución de pesos según el procedimiento anterior, se *decilizan* de nuevo las bases de datos 2015 para comparar con el *pool 2014-2015* y detectar el remanente de registros del 2015 en el decil superior de acuerdo con el *pool*.

El procedimiento se repite hasta que algún tramo *manzana* ya no pueda ser donador. En el presente ejercicio eso ocurrió en la segunda iteración, tal como se muestra en las matrices del apartado 3.1.

En este algoritmo queda implícito que para $j = 1$ su tramo *pera* puede recibir modificaciones en sus factores de expansión, sin embargo, es el único decil donde el tramo *manzana* no es dador de factores

de expansión y no es modificado de forma alguna, ya que se establece que iniciará el proceso cuando $M_{j+1}^{kle} \neq \emptyset$. Para el caso $j = 10$, su tramo *manzana* es vacío (esto se definió con anterioridad), por lo tanto, el algoritmo no implementa ningún cambio en su tramo *pera* pues j varía entre 1 y 9.

Un tramo *pera* (con excepción de los dos casos anteriores) puede ser receptor de factores de expansión de los tramos *manzana* tanto del decil $j + 1$ como del propio decil.

3.2.2 Software

El *software* utilizado en todo el procesamiento fue *Visual Fox Pro 9.0* por ser un lenguaje de programación procedural y orientado a objetos que incluye un Sistema Gestor de Bases de Datos (DBMS).

3.2.3 Fuentes

En todo el procedimiento no se utilizaron fuentes de información distintas a la ENIGH/MCS del 2010 al 2015, salvo en la utilización del índice nacional de precios al consumidor para indexar el ICT y sus fuentes a precios corrientes del 2015.

4. Resultados

En esta sección, como ya se había anticipado, se muestran los resultados primero para el ejercicio *Delta* sin ajuste demográfico seguido de los que lo incorporan.

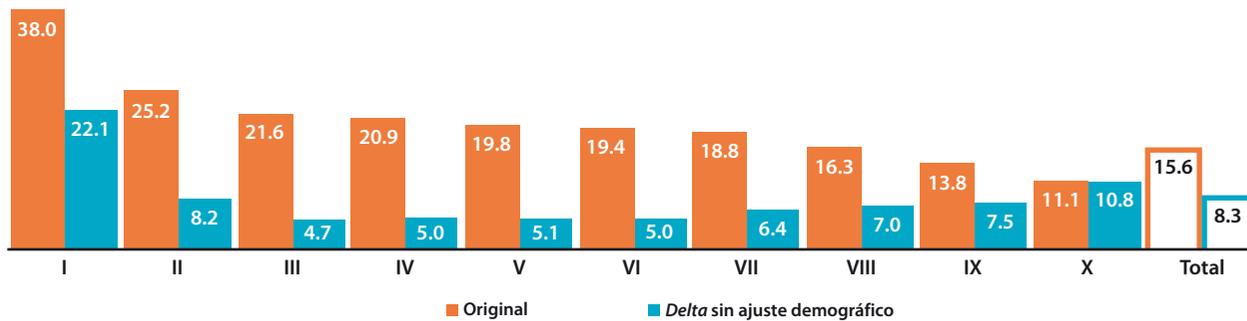
4.1. Resultados a nivel nacional, por entidad federativa y por decil con la nueva base del MCS 2015

4.1.1 Delta sin ajuste demográfico Ingreso corriente total

En términos de volumen, la variación del ICT a precios constantes que arrojó el MCS 2015 presentaba un incremento de 15.6% en el ámbito nacional.

Gráfica 4.1

Volumen de ingreso corriente total (variación porcentual 2014-2015)



En particular, para el primer decil se elevó hasta 38, 25.2 y 21.6%, respectivamente, para el II y el III. En el otro extremo (decil X), el incremento fue de 11.1% (barras de color naranja en la gráfica 4.1).

Delta sin ajuste demográfico (barras azules) disminuyó la variación total de 15.6 a 8.3%; la del primer decil, de 38 a 22.1%; la del segundo, de 25.2 a 8.2% y la del tercero, de 21.6 a 4.7 por ciento. La disminución sigue siendo palpable entre los deciles IV y IX, limitándose a una disminución de tres décimas en el X.

Una manera analíticamente relevante de describir el incremento en los montos del ICT del MCS 2015, tanto a nivel general como a nivel decil, es desagregando dicho incremento en tres factores contributivos: a) incremento de percepciones por perceptor (IPP), b) perceptores por hogar (PH) y c)

aumento de hogares (Hog). Los valores originales en el nacional de estos coeficientes fueron: 1.077 (7.7%), 1.039 (3.9%) y 1.033 (3.3%), respectivamente, que multiplicados dan el incremento total de 1.156 (15.6% referido líneas arriba). *Delta sin ajuste demográfico*, por definición, no afecta la contribución del incremento de hogares al volumen (dado que el método solo redistribuyó los factores de expansión a nivel hogar de manera que el total de hogares a nivel nacional es el mismo), lo que en cambio sí modifica es la contribución del incremento de los ingresos por perceptor, disminuyéndolo a 1.048 y prácticamente neutralizando la contribución del aumento de perceptores por hogar; este componente queda incluso como un factor de disminución (valores menores a 1) en los cuatro primeros deciles, así como en el VI. Por su parte, el decil X casi no experimenta cambio en su contribución de factores (ver tabla 4.1).

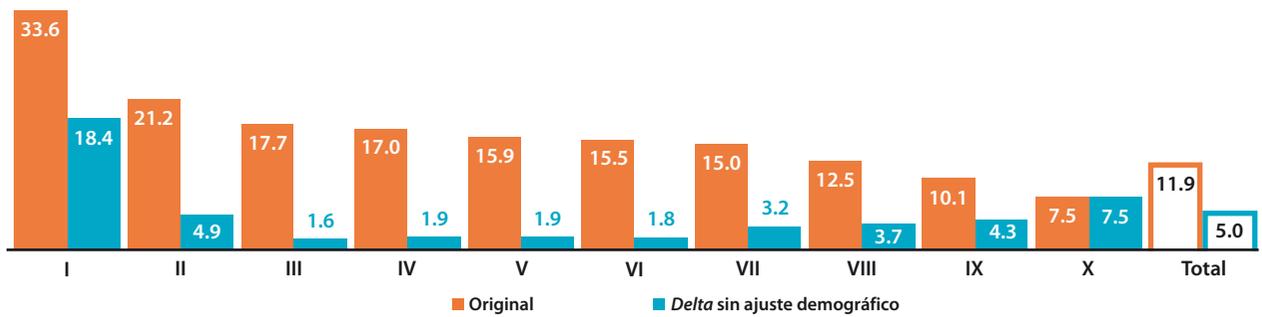
Tabla 4.1

Descomposición del ingreso corriente total

Decil	Original				<i>Delta sin ajuste demográfico</i>			
	ICT ₁₅ / ICT ₁₄	IPP	PH	Hog	ICT ₁₅ / ICT ₁₄	IPP	PH	Hog
Total	1.156	1.077	1.039	1.033	1.083	1.048	1.002	1.031
I	1.380	1.313	1.017	1.033	1.221	1.232	0.961	1.031
II	1.252	1.243	0.976	1.033	1.082	1.138	0.922	1.031
III	1.216	1.155	1.019	1.033	1.047	1.058	0.960	1.031
IV	1.209	1.146	1.021	1.033	1.050	1.053	0.967	1.031
V	1.198	1.105	1.049	1.033	1.051	1.013	1.006	1.031
VI	1.194	1.123	1.028	1.033	1.050	1.041	0.978	1.031
VII	1.188	1.094	1.051	1.033	1.064	1.003	1.029	1.031
VIII	1.163	1.072	1.049	1.033	1.070	1.029	1.008	1.031
IX	1.138	1.030	1.069	1.033	1.075	0.982	1.062	1.031
X	1.111	0.993	1.082	1.033	1.108	0.995	1.080	1.031

Gráfica 4.2

Promedio de ingreso corriente total (variación porcentual 2014-2015)



En cuanto a las variaciones del ICT promedio por hogar, la variación total a precios constantes pasó de 11.9 a 5% con *Delta*. El promedio del primer decil fue de un incremento de 33.6% a uno de 18.4%, de 21.2 a 4.9% en el segundo y de 17.7 a 1.6% en el tercero. Al igual que en términos de volumen, el ajuste es similar para los deciles IV a VI, mientras que para el decil X la variación es la misma (ver gráfica 4.2).

En lo que a pobreza extrema se refiere, el MCS 2015 la situó en 5.8% de la población frente a 9.5% que se desprendió del 2014, un descenso de 3.7 puntos. El ajuste *Delta* dejó el nivel en 7%, es decir, con una disminución de 2.5 puntos.

La pobreza moderada disminuyó 5.7 puntos en el MCS 2015 al pasar de 36.6% en el 2014 a 30.9 por ciento. Con el ajuste *Delta* la disminución es de 3.2 puntos para situarla en 33.4 por ciento.

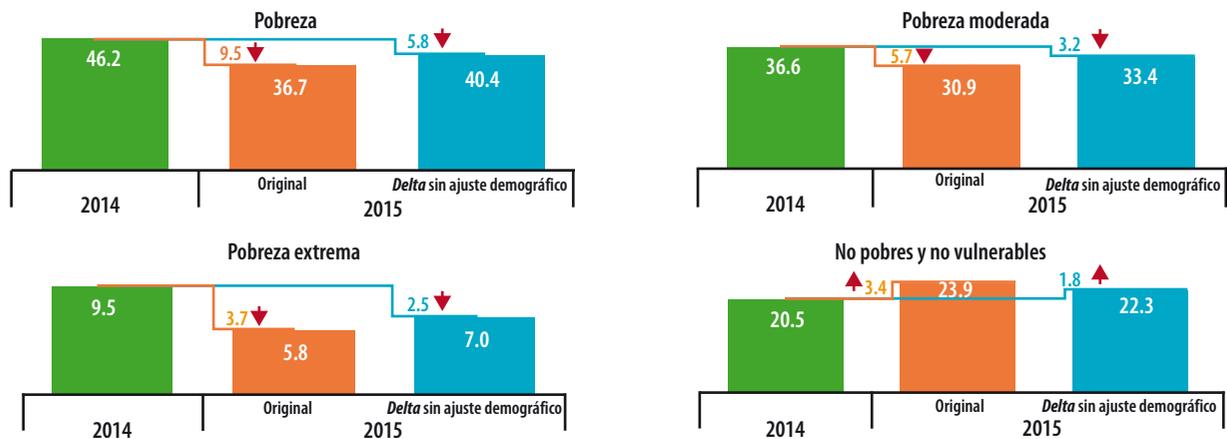
Niveles de pobreza y bienestar a nivel nacional

Del MCS 2015 se desprendió un nivel de pobreza que involucra a 36.7% de la población total frente a 46.2% del levantamiento del 2014, lo que significa un descenso de 9.5 puntos porcentuales. El ajuste *Delta* dejó el nivel de pobreza en 40.4%, un descenso de 5.8 puntos.

¿Y qué sucede con los que no son ni pobres ni vulnerables? Del Módulo de Condiciones Socioeconómicas 2015 se desprendió que este grupo de población se incrementó 3.4 puntos porcentuales pasando de significar 20.5% de la población en el 2014 a 23.9% en el 2015. El ajuste *Delta* moderó este aumento en 1.8 puntos, dejando el nivel en 22.3% de la población (ver gráfica 4.3).

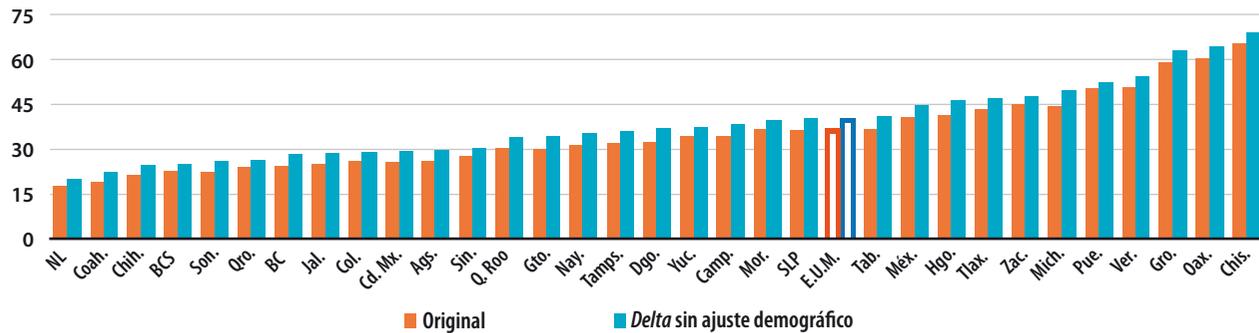
Gráfica 4.3

Porcentajes de pobreza multidimensional (categorías seleccionadas)



Gráfica 4.4

Porcentaje de pobreza por entidad federativa



Niveles y variación de la pobreza por entidad federativa

Como resultado de la aplicación del ajuste *Delta*, los niveles de pobreza de todas las entidades federativas se incrementaron respecto a los resultados que arrojó el MCS 2015 (ver gráfica 4.4).

Comparado ahora el ajuste en términos de variaciones respecto al 2014, se contrasta el hecho de que si bien en el ámbito nacional la pobreza multidimensional decreció menos (barras en color sólido de la gráfica 4.5) que el decrecimiento que se desprendió del MCS 2015 antes del ajuste (barras achuradas), hay un caso en el que el ajuste no solo disminuyó la variación sino que, de hecho, se tradujo en crecimientos de la pobreza: Ciudad de México.

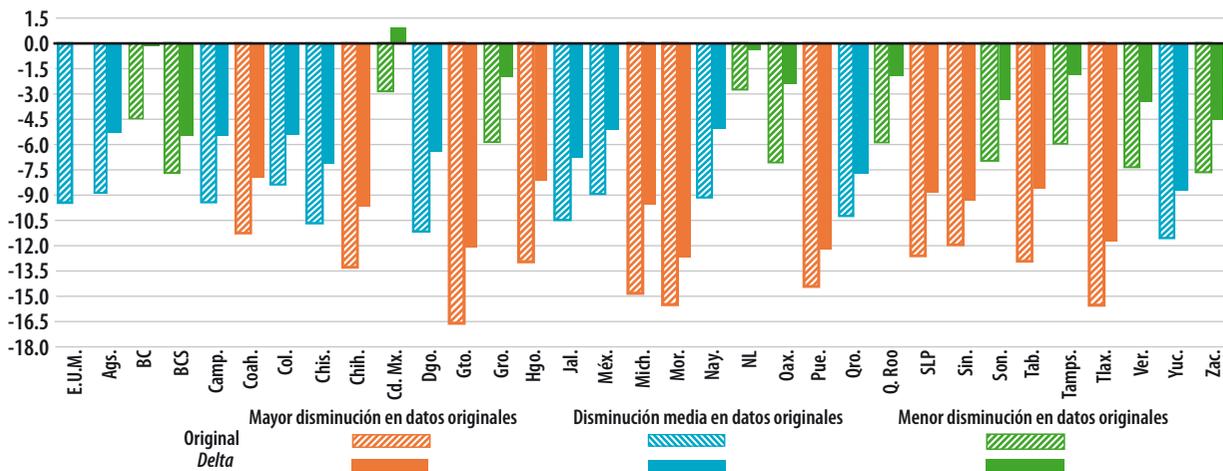
4.1.2 Delta con ajuste demográfico

Los resultados que se ilustran en este apartado corresponden al ajuste *Delta* seguido del ajuste demográfico, entendiendo por esto último el proceso de alineación del tamaño del hogar del MCS 2015 a los del levantamiento previo, tomando en cuenta la evolución del tamaño del hogar 14-15 que se desprende de las proyecciones del CONAPO (ver el documento en esta serie que corresponde a dicho ejercicio).

Es importante subrayar que, si bien cuando se tiene como unidad de observación al hogar, no se detectan grandes cambios en *Delta* sin y con ajuste demográfico en términos del ICT; en cambio, sí hay un contraste en cuanto a pobreza porque el ajuste demográfico la incrementa.

Gráfica 4.5

Variación porcentual 2014-2015 de la pobreza por entidad federativa



Lo anterior obedece al hecho de que la redistribución de factores de expansión del procedimiento *Delta* tuvo por efecto una disminución de población en la expansión de los microdatos. El ajuste demográfico, al recuperar esa población, impacta las mediciones de pobreza cuya unidad de observación no son los hogares sino los individuos.

En otras palabras, el ingreso per cápita en *Delta* antes del ajuste demográfico es sensiblemente mayor que después de él.

Ingreso corriente total

En la gráfica 4.6 se contrastan las variaciones en volumen que arroja el MCS 2015 original (barras co-

lor naranja) con las de *Delta* sin ajuste demográfico (azules) y *Delta* con dicho ajuste (amarillas). Se aprecia que el ajuste demográfico disminuyó el volumen a precios constantes a nivel total de 8.3 a 7.2%, es decir, algo más de un punto porcentual pero no mucho más allá. El efecto del ajuste fue reducir un punto porcentual más en los primeros cuatro deciles.

Visto desde la perspectiva de la descomposición de factores que inciden en el incremento del volumen del ICT (como el IPP y el PH), no cambian realmente de manera sensible respecto a lo descrito en el apartado anterior. Sin embargo, aquí sí tiene lugar una disminución en el factor de incremento de hogares (Hog) lo que determina que la variación en volumen del ICT a precios constantes ya no sea de 8.3% sino de 7.2 por ciento.

Gráfica 4.6

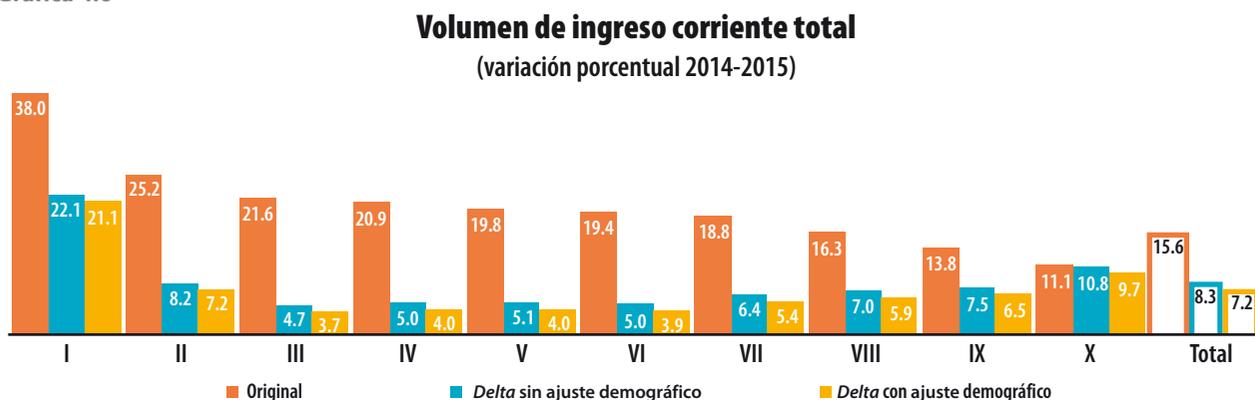


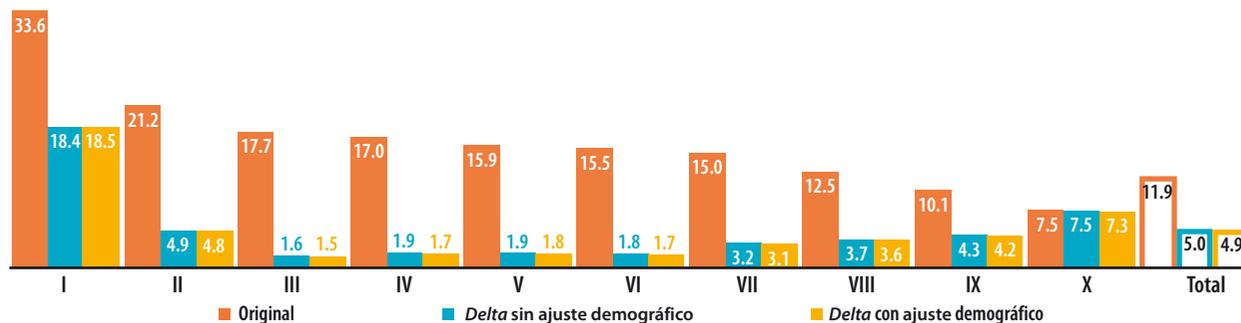
Tabla 4.2

Descomposición del ingreso corriente total

Decil	Original				Delta sin ajuste demográfico				Delta con ajuste demográfico			
	ICT ₁₅ /ICT ₁₄	IPP	PH	Hog	ICT ₁₅ /ICT ₁₄	IPP	PH	Hog	ICT ₁₅ /ICT ₁₄	IPP	PH	Hog
Total	1.156	1.077	1.039	1.033	1.083	1.048	1.002	1.031	1.072	1.046	1.002	1.022
I	1.380	1.313	1.017	1.033	1.221	1.232	0.961	1.031	1.211	1.233	0.961	1.022
II	1.252	1.243	0.976	1.033	1.082	1.138	0.922	1.031	1.072	1.136	0.923	1.022
III	1.216	1.155	1.019	1.033	1.047	1.058	0.960	1.031	1.037	1.058	0.959	1.022
IV	1.209	1.146	1.021	1.033	1.050	1.053	0.967	1.031	1.040	1.049	0.970	1.022
V	1.198	1.105	1.049	1.033	1.051	1.013	1.006	1.031	1.040	1.014	1.003	1.022
VI	1.194	1.123	1.028	1.033	1.050	1.041	0.978	1.031	1.039	1.036	0.982	1.022
VII	1.188	1.094	1.051	1.033	1.064	1.003	1.029	1.031	1.054	1.001	1.030	1.022
VIII	1.163	1.072	1.049	1.033	1.070	1.029	1.008	1.031	1.059	1.027	1.009	1.022
IX	1.138	1.030	1.069	1.033	1.075	0.982	1.062	1.031	1.065	0.980	1.063	1.022
X	1.111	0.993	1.082	1.033	1.108	0.995	1.080	1.031	1.097	0.993	1.081	1.022

Gráfica 4.7

Promedio de ingreso corriente total (variación porcentual 2014-2015)



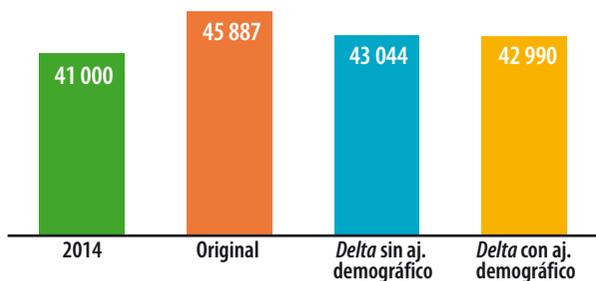
Las variaciones en términos del ICT promedio por hogar son todavía más suaves que en volumen con respecto a *Delta* antes de ajuste demográfico. Si bien con el ajuste hay disminución en las variaciones de los deciles I a IX en relación con *Delta* antes de dicho ajuste, se trata de variaciones realmente mínimas (ver gráfica 4.7).

Llegado a este punto, es importante tomar en cuenta que si bien en los ICT promedio por hogar

el ajuste demográfico a *Delta* no hace una gran diferencia en volumen —y, por ende, en términos de crecimiento— con respecto a *Delta* sin tal ajuste (ver gráficas 4.8 y 4.9), en términos de ingreso per cápita (por persona) este último no se alejó todo lo que debía de lo que arrojó el MCS 2015, dado que terminó con una merma o déficit poblacional respecto a la proyección del CONAPO. Una vez que se recalibraron los factores para alinearse a dicha proyección, se recuperó el monto poblacional

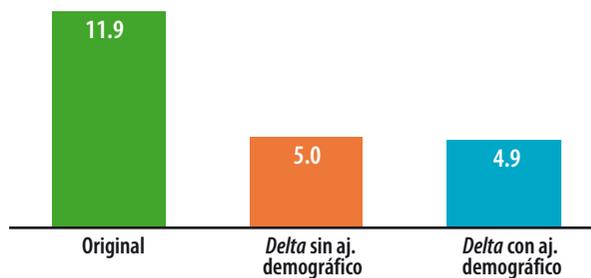
Gráfica 4.8

Promedio de ingreso corriente total por hogar



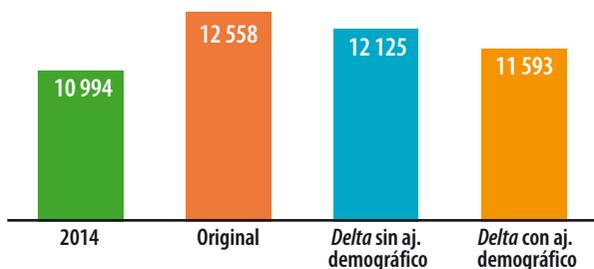
Gráfica 4.9

Promedio de ingreso corriente total por hogar (variación porcentual 2014-2015)



Gráfica 4.10

Promedio de ingreso corriente total por persona



Gráfica 4.11

Promedio de ingreso corriente total por persona (variación porcentual 2014-2015)



y, por consiguiente, el ingreso per cápita disminuyó más (ver gráficas 4.10 y 4.11).

Todo lo anterior tendrá un efecto en las mediciones de la pobreza que se basa en el ingreso corriente total per cápita. Así, en la siguiente sección se pondrá énfasis en el contraste entre *Delta* sin y con ajuste demográfico.

Niveles de pobreza y bienestar a nivel nacional

Con el método *Delta* sin ajuste demográfico el descenso de la pobreza multidimensional del CONEVAL se moderó de una caída de 9.5 puntos porcentuales respecto al 2014 que reportaba MCS 2015 a una de 5.8 puntos; pero una vez que se incorporó dicho ajuste, la disminución ahora es de 4.1 puntos, de modo que la pobreza multidimensional pasó de 46.2% de la población total en el 2014 a 42.1% (no 36.7% que arrojaba el MCS 2015 antes de estos procedimientos).

En lo que corresponde a pobreza extrema, *Delta* sin ajuste demográfico corrigió la disminución del MCS 2015 de 3.7 puntos a 2.5; pero una vez incorporado el ajuste, la disminución fue de 2, de modo

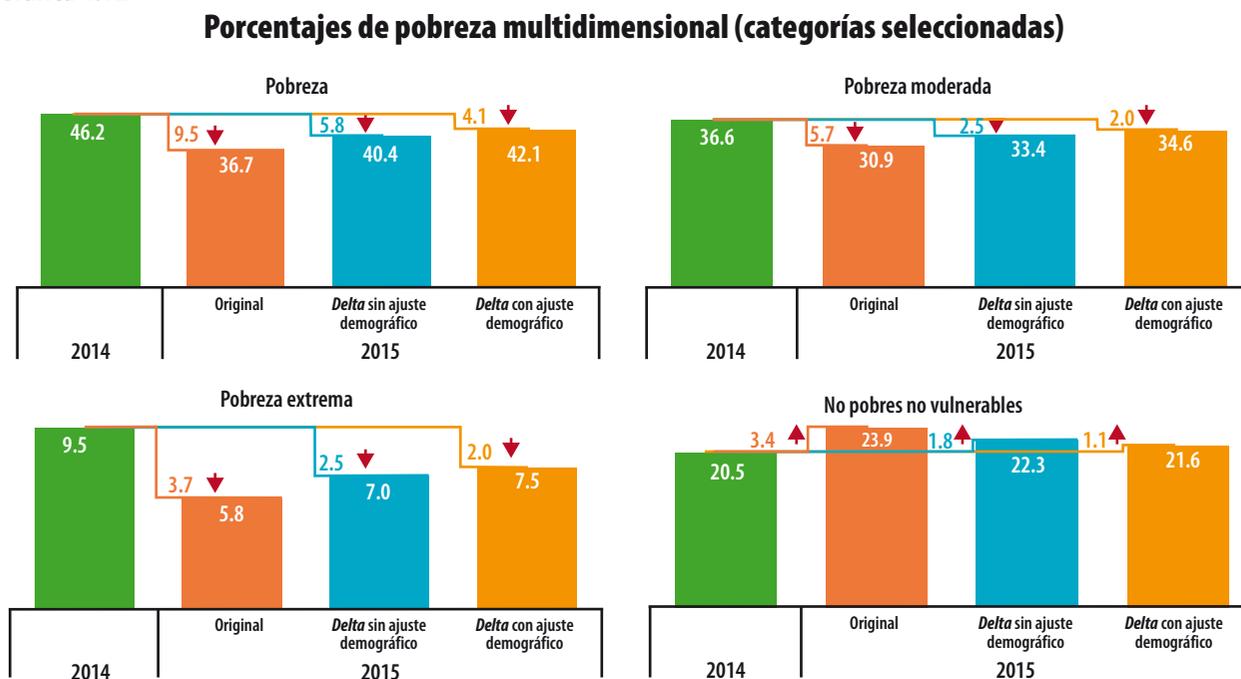
que este tipo de pobreza pasó de 9.5% de la población total en el 2014 a 7.5% un año después.

En cuanto a la pobreza moderada, pasó de una disminución de 5.7 puntos que le atribuía el MCS 2015 a 2.5 con *Delta* sin ajuste y a 2 con ajuste demográfico. El componente mayoritario de la pobreza evolucionó entonces de un nivel de 36.6% en el 2014 a 34.6% en el 2015.

Los no pobres y no vulnerables en vez de incrementarse 3.4 puntos en comparación con el 2014 en el MCS 2015 y en 1.8 puntos en *Delta* sin ajuste demográfico, lo hicieron ahora en 1.1 puntos porcentuales, pasando de un nivel de 20.5% en el 2014 a 21.6% (ver gráfica 4.12).

No está de más hacer notar que, siendo la estimación de porcentajes de población con carencias sociales algo no directamente determinado por el ingreso corriente monetario, el ajuste *Delta* si bien confirma que hubo un avance al respecto en todos los rubros respecto al 2014, éste resulta más moderado de lo que indica el MCS 2015 antes de someterse a los procedimientos descritos en esta nota (en la tabla 4.3 se designa como MCS original).

Gráfica 4.12



Niveles y variación de la pobreza por entidad federativa

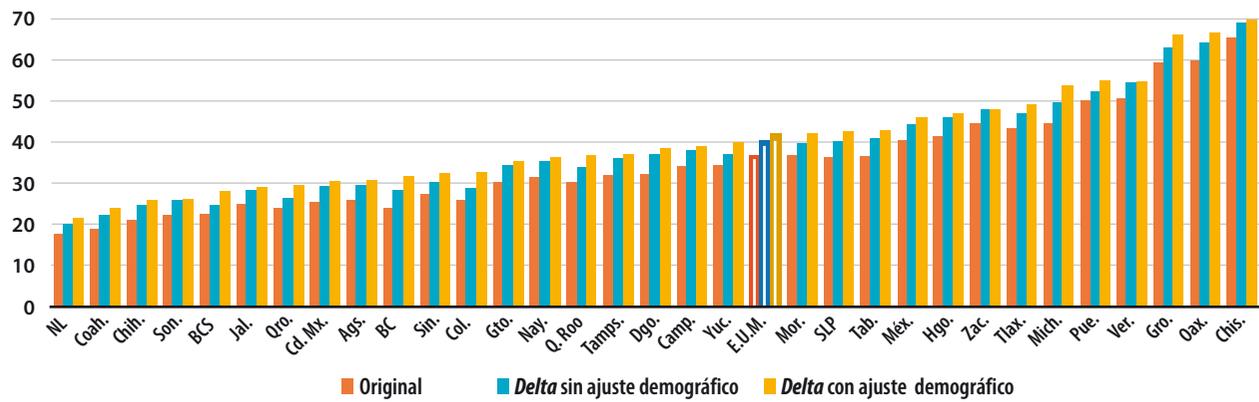
La incorporación del ajuste demográfico a *Delta* por entidad federativa (ver gráfica 4.13) se tradujo en mayores niveles de pobreza multidimensional que, sin dicho ajuste y comparado contra el MCS 2015 original, el mayor ajuste al alza fue para Michoacán de Ocampo (9.35 puntos), en tanto que el menor fue para Zacatecas (3.26). En la tabla 4.4 se resumen los valores máximos y mínimos de ajuste obtenidos de las entidades con respecto al MCS 2015 original para todos los agregados

de pobreza y carencias sociales que maneja el CONEVAL. Para consultar los datos por entidad, ver las tablas 7.5, 7.6 y 7.7.

Comparadas las variaciones respecto al 2014 con las que arroja el MCS 2015 (achurado en la gráfica 4.14), *Delta* con ajuste demográfico (barras sólidas) corroboró respecto al primer *Delta* que habría, de hecho, incremento en la pobreza en Ciudad de México, a lo que ahora se suman Baja California, Guerrero, Nuevo León y Quintana Roo. En el caso de Oaxaca lo que era una disminución de 7 puntos quedó como disminución de una décima.

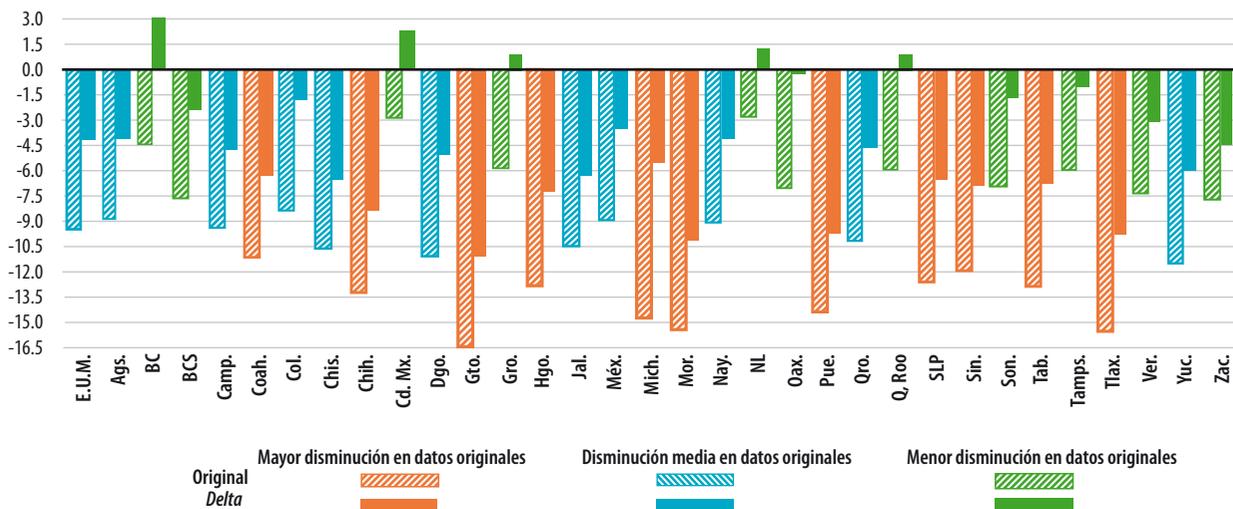
Gráfica 4.13

Porcentaje de pobreza por entidad federativa



Gráfica 4.14

Variación porcentual 2014-2015 de la pobreza por entidad federativa



Resumen de resultados

Tabla 4.3

Porcentaje de población por indicadores de pobreza, privación y carencias sociales

Indicador	2014	2015		Diferencia (Delta - original)
		Original	Delta	
Pobreza	46.17	36.69	42.09	5.40
Pobreza moderada	36.62	30.85	34.63	3.78
Pobreza extrema	9.55	5.85	7.47	1.62
Población vulnerable por carencias	26.26	33.94	29.76	-4.18
Población vulnerable por ingresos	7.07	5.45	6.58	1.13
Población no pobre y no vulnerable	20.50	23.91	21.56	-2.35
Privación social				
Población con al menos una carencia	72.43	70.64	71.85	1.22
Población con al menos tres carencias	22.12	20.53	21.45	0.92
Indicadores de carencia social				
Rezago educativo	18.66	17.95	18.52	0.58
Acceso a servicios de salud	18.16	16.87	16.99	0.12
Acceso a la seguridad social	58.47	56.59	57.83	1.24
Calidad y espacios de la vivienda	12.32	12.01	12.17	0.15
Acceso a servicios básicos de la vivienda	21.22	20.48	21.06	0.58
Acceso a la alimentación	23.35	21.75	22.66	0.91

Tabla 4.4

Valores estadísticos de las diferencias de los datos Delta menos MCS 2015 original de las entidades federativas

Indicador	Máxima	Mínima	Media	Mediana	Desviación estándar
Pobreza	9.35	3.26	5.52	5.37	1.14
Pobreza moderada	6.69	1.03	3.95	4.12	1.23
Pobreza extrema	4.54	0.36	1.57	1.26	1.04
Población vulnerable por carencias	-1.73	-8.50	-4.30	-4.17	1.36
Población vulnerable por ingresos	2.80	-0.29	1.12	1.14	0.73
Población no pobre y no vulnerable	-0.52	-4.47	-2.33	-2.40	1.00
Privación social					
Población con al menos una carencia	3.32	0.21	1.22	1.06	0.70
Población con al menos tres carencias	2.09	-0.43	0.84	0.83	0.50
Indicadores de carencia social					
Rezago educativo	1.37	-0.23	0.51	0.44	0.38
Acceso a servicios de salud	1.07	-0.92	0.28	0.37	0.48
Acceso a la seguridad social	3.41	-0.04	1.32	1.18	0.78
Calidad y espacios de la vivienda	1.10	-0.75	0.14	0.06	0.46
Acceso a servicios básicos de la vivienda	1.59	-0.63	0.48	0.26	0.60
Acceso a la alimentación	3.11	-0.32	0.98	0.92	0.74

4.2 Comparación de los resultados con los publicados para el MCS 2015 y ENIGH-MCS 2010, 2012 y 2014, a nivel nacional y por decil

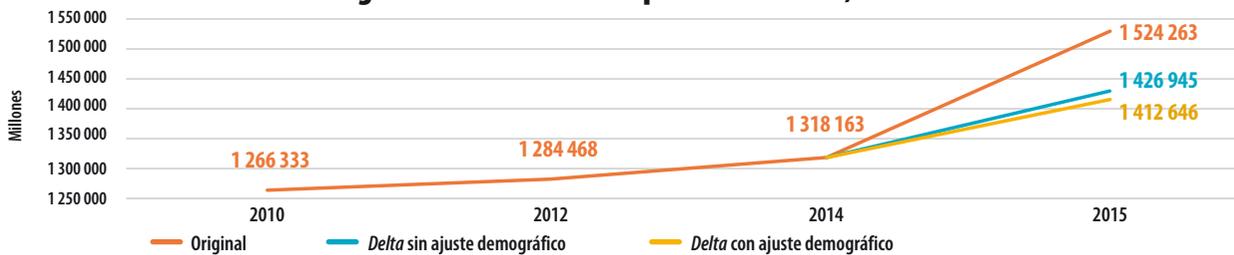
En todo el apartado anterior, los contrastes se establecían respecto al 2014. En esta sección, se ilustrará dentro de una perspectiva temporal más

amplia, de modo que se aprecie cómo se inscriben los datos obtenidos dentro de la trayectoria de la serie. La primera secuencia gráfica se referirá a la evolución del ICT en términos de volumen; la segunda, en términos de promedios. En una y otra se ilustra, asimismo, cómo queda la evolución para los tres primeros deciles.

Volumen del ICT a precios del 2015, serie del 2010 al 2015

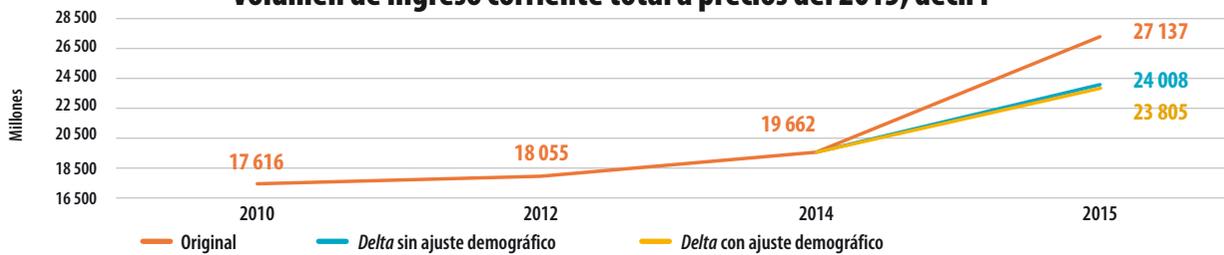
Gráfica 4.15

Volumen de ingreso corriente total a precios del 2015, ámbito nacional



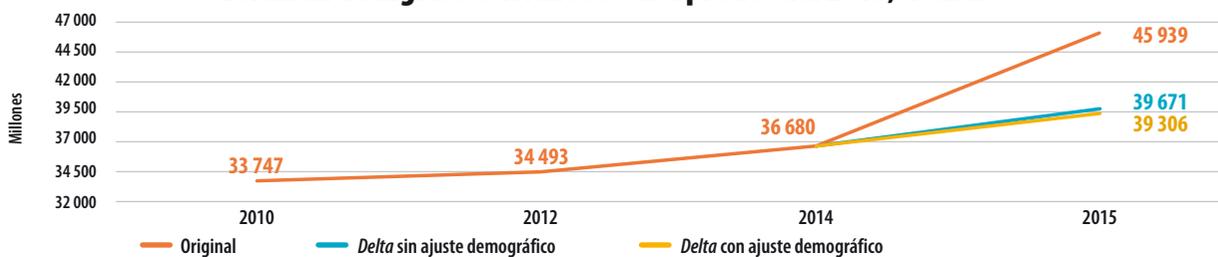
Gráfica 4.16

Volumen de ingreso corriente total a precios del 2015, decil I



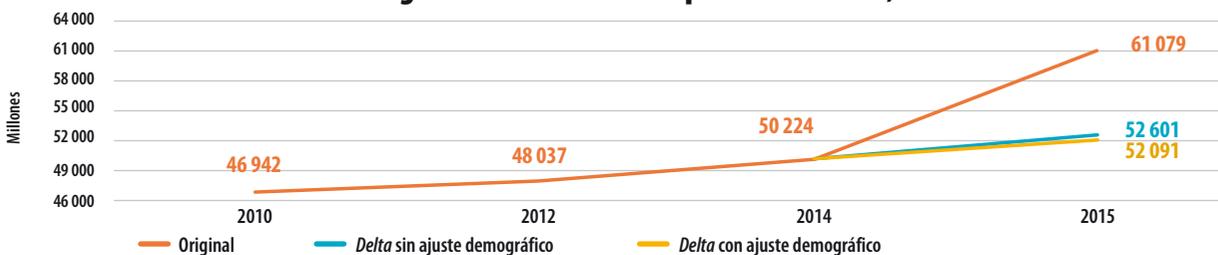
Gráfica 4.17

Volumen de ingreso corriente total a precios del 2015, decil II



Gráfica 4.18

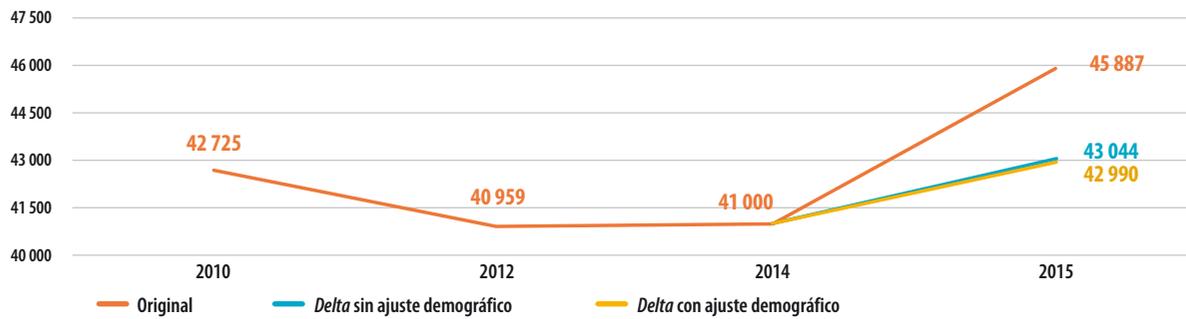
Volumen de ingreso corriente total a precios del 2015, decil III



Promedio del ICT por hogar a precios del 2015, serie del 2010 al 2015

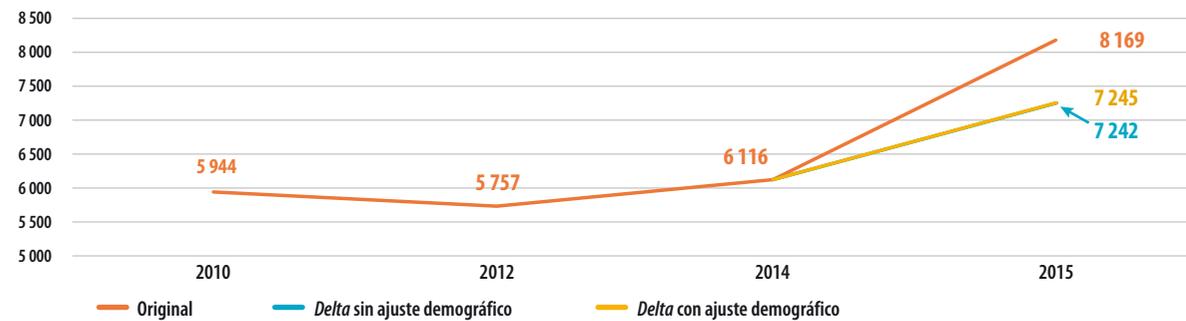
Gráfica 4.19

Promedio de ingreso corriente total por hogar a precios del 2015, ámbito nacional



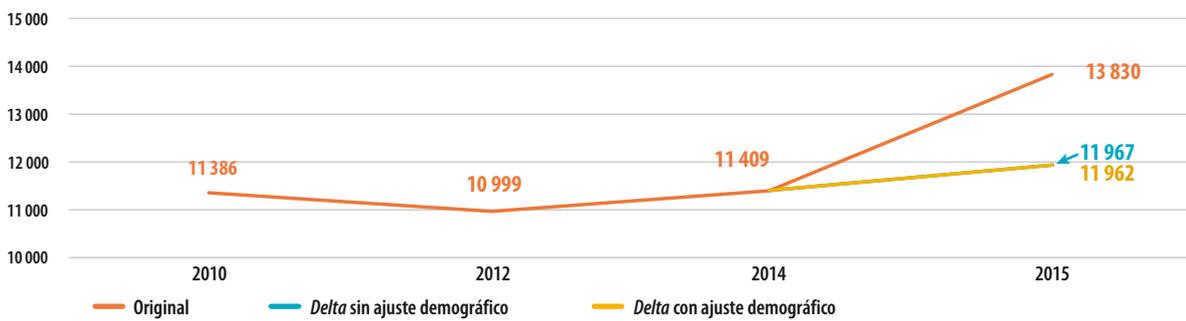
Gráfica 4.20

Promedio de ingreso corriente total por hogar a precios del 2015, decil I



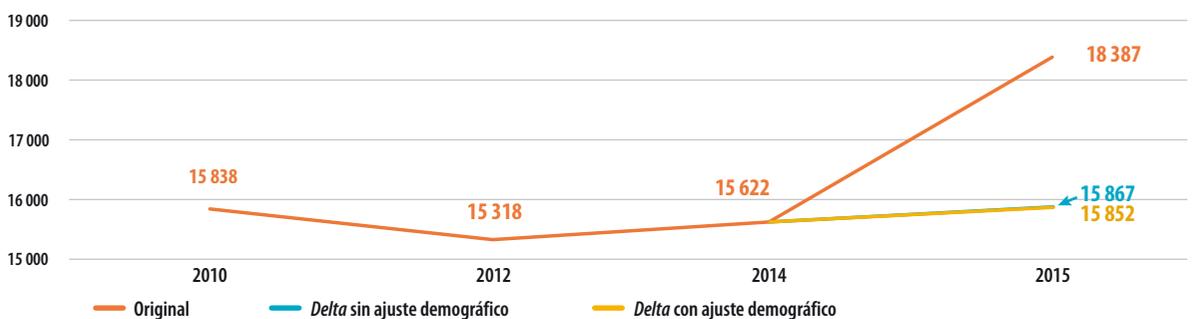
Gráfica 4.21

Promedio de ingreso corriente total por hogar a precios del 2015, decil II



Gráfica 4.22

Promedio de ingreso corriente total por hogar a precios del 2015, decil III

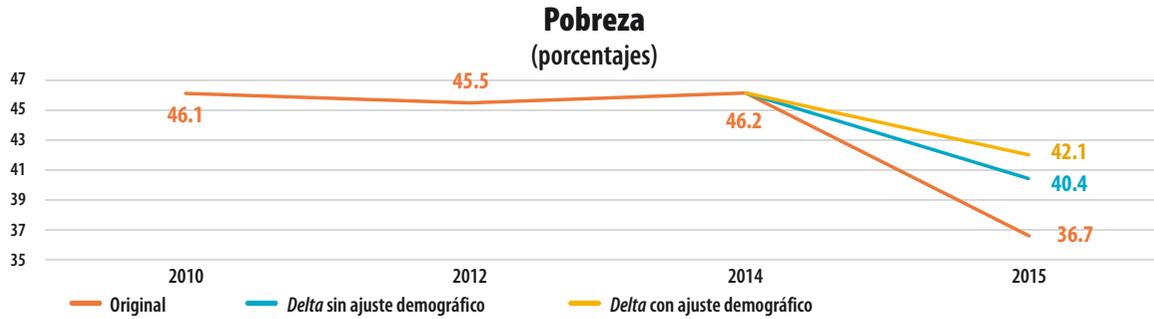


Pobreza multidimensional

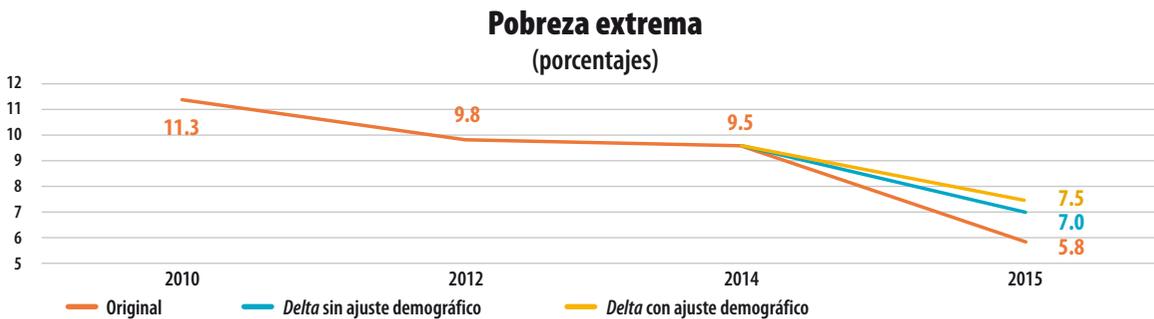
La tercera sección se reserva a las trayectorias en términos de la pobreza multidimensional y sus dos modalidades constitutivas, así como para las distintas modalidades de vulnerabilidad añ-

diendo, por último, la trayectoria para el porcentaje de población que escapa tanto a pobreza como a vulnerabilidad. Los datos correspondientes por entidad federativa no serán mostrados en esta sección gráfica sino que se reservan para el anexo estadístico.

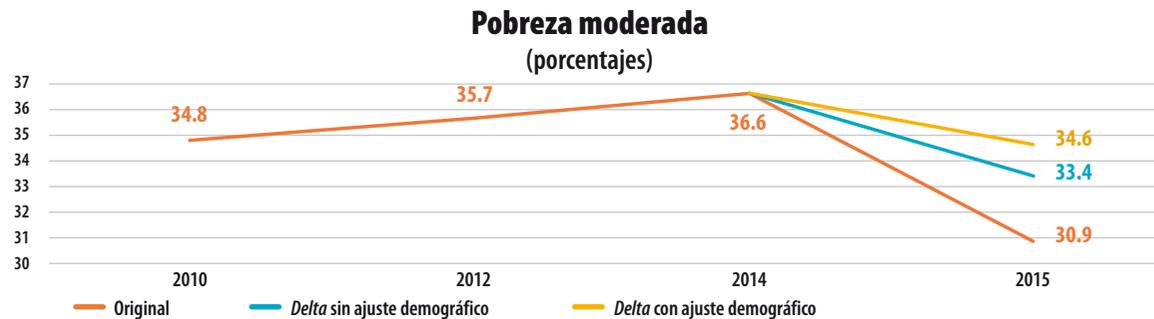
Gráfica 4.23



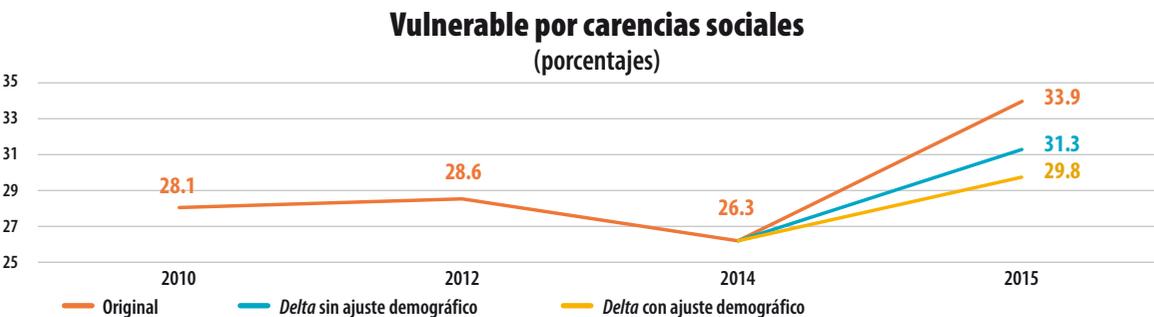
Gráfica 4.24



Gráfica 4.25



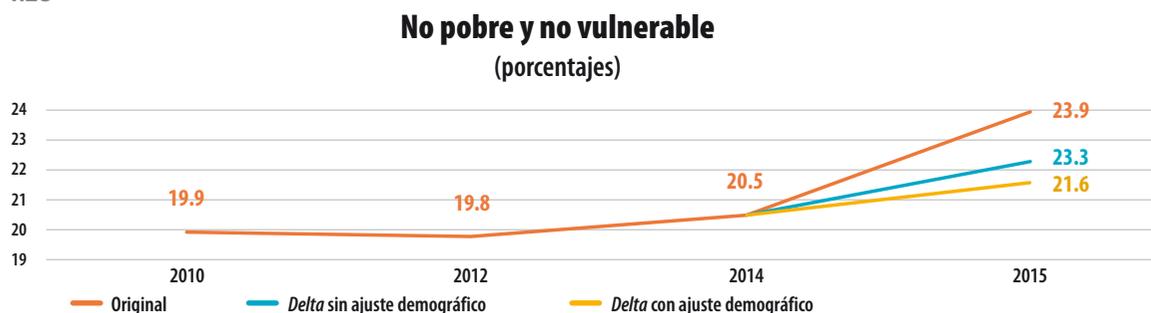
Gráfica 4.26



Gráfica 4.27



Gráfica 4.28



5. Validación y evaluación de la metodología

La pertinencia de este ejercicio puede verse primero en términos de que, si bien suaviza o modera los datos del MCS 2015 respecto a lo que le antecedió, no lo hace proyectando una evolución esperada ya sea de los ingresos o de la pobreza. Deja un margen en donde tiene cabida el hecho de que el MCS haya captado mejor ciertas cosas, como las fuentes de ingreso independientemente del decil de que se trate o acusar el efecto de factores exógenos que impactaron de manera específica al 2015, como la reforma fiscal. Es así que este ejercicio en su expresión final conserva dos hallazgos o propiedades del MCS 2015 en el ámbito nacional: sigue habiendo más ingresos por perceptor y más perceptores por hogar (factores mayores a 1 en el bloque de la derecha de la tabla 4.2). Dicho margen es posible, entre otras cosas, porque el procedimiento en ningún momento modifica los datos de ingreso tal como fueron declarados por los informantes. Lo captado por las entrevistas se respeta por completo.

La validez de este ejercicio supone, entonces, que la captación del ingreso es tal que si bien la medición de la pobreza no disminuye como se desprendería de la base original del MCS 2015, no dejó de hacerlo en alguna medida, aunque más modesta. En ese sentido, los datos de este ejercicio deben inscribirse en un punto intermedio entre lo que se obtuvo del 2014 y lo que se obtendrá del levantamiento del 2016, cuyos resultados se darán a conocer a mediados del 2017.

Una implicancia de este ejercicio en términos de comparabilidad es que los datos de la ENIGH 2016 debieran quedar sujetos asimismo a un ejercicio *Delta*, alineando a los deciles del 2014, del mismo modo que se hizo con el 2015. Dicho ejercicio debiera aproximar aún más los datos entre el 2015 y 2016.

Sin duda, un supuesto mayor es no cuestionar nada del levantamiento del 2014, ya que éste se utilizó como referente en el procedimiento para detectar hibridaciones en los deciles (*peranzanas*), tal como fue descrito en el apartado metodológico.

co. La construcción del *pool*, como se describió en dicho apartado, solo puede involucrar dos años: el de referencia y el que se alinea. Un *pool* que involucre la muestra de más levantamientos solo estaría exagerando el efecto de las decisiones que se tomaron respecto al MCS 2015, pues al rellenar los microdatos con más registros del ICT menores tendría, por consecuencia, hacer mayores los tramos *manzana* en cada decil, lo que culminaría en un sobreajuste. El método conlleva, por su concepción misma, una fuerte dependencia respecto al levantamiento que se utilice para alinearlos.

El método, asimismo, entraña una pérdida de diversidad de información pues, básicamente, sobreexpande la diagonal de la matriz al final de las iteraciones; es decir, los registros del MCS 2015 más estables dentro de cada decil o cuya asignación no fue afectada por el colapso de frecuencias de los ICT más bajos. Por lo pronto, el ejercicio da una idea de la corrección en los niveles de ingreso y modalidades de pobreza que debieron manifestarse de no haber ocurrido una serie de factores —cualesquiera que hayan sido— que incidieron en el colapso de las frecuencias del ICT más bajas entre el 2014 y 2015, ello sin eliminar hallazgos válidos del MCS 2015 con respecto a una mayor captación de ingresos por receptor y perceptores por hogar en el ámbito nacional.

El que la sobreexpansión de la diagonal a partir de registros neutralizados (*manzana*) sea un procedimiento similar al de haber tratado a dichos registros como una no respuesta, puede dar lugar a la inquietud de qué tanto ello pudo afectar las precisiones estadísticas. En el *Anexo: precisiones estadísticas* puede verse el comparativo para las medias de 44 variables a nivel nacional y las diferencias habidas a niveles de errores estándar y coeficientes de variación. Asimismo, es de subrayar que de las 44 variables consideradas, en 36 se da un traslape de intervalos a 95% entre el MCS 2015 antes de los procedimientos y después del ejercicio *Delta* (con todo y ajuste demográfico). Las ocho variables en las que no ocurre esto (lo que sugiere diferencias estadísticamente significativas) son aquellas que corresponden a las ma-

yores incógnitas que planteaba el MCS 2015 y que no resultaban avalados por otras fuentes, como el incremento excesivo de ocupación y de las remuneraciones al trabajo. Es justo ahí donde termina actuando el ejercicio *Delta*, pero sin haber fijado la corrección de tales variables como punto de partida, todo lo cual lo hace un genuino resultado del procedimiento, no un tautológico ejercicio predeterminado.

6. Conclusión

El ejercicio presentado no sustituye ni altera, en ningún caso, el vector de ingresos obtenido de los informantes en el levantamiento MCS 2015; la alteración se da en el vector correspondiente a los factores de expansión.

El colapso de los fractiles de ingresos inferiores cambió la *decilización* del 2015 respecto al 2014, de modo que para completar el número de hogares en el decil quedaron asentados fractiles que, un año atrás, hubieran sido parte del decil II y así sucesivamente; lo que se ha denominado aquí un efecto *peranzana*. Buena parte de las perplejidades que suscitó el análisis comparativo entre el 2014 y 2015, en especial en lo que concierne a los primeros deciles, tienen en ello su origen.

Sin duda, el decil más difícil de comparar en el MCS 2015 respecto al levantamiento del 2014 es el primero, pues se vio afectado en sus dos tramos: por arriba (*manzana* en la *pera*) y por abajo, por el colapso de los fractiles. Ello, desde luego, se tradujo en variaciones muy importantes en sus promedios del ICT.

El método detecta los tramos *manzana* comparando la *decilización* del MCS 2015 con la del *pool 14-15*, donde los registros del 2014 aportan, sobre todo, las viviendas con fractiles colapsados.

Una vez que se detectan tramos *manzana* en la *decilización* del 2015 se neutralizan con factor de expansión = 1 y se redistribuye la expansión hacia el tramo *pera* en un sentido descendente y en una

secuencia en cascada, de modo que el primer decil sea un receptor neto de factores de expansión. El proceso mantiene cierta similitud con un tratamiento de no respuesta para los registros *manzana* y la consiguiente sobreexpansión de otros que guarden alguna conexión con aquéllos, como lo es el pertenecer a los mismos dominios de muestra, así como de estrato socioeconómico. Cuando no puede cumplirse con esa correspondencia en un sentido descendente, no hay lugar para el proceso de neutralización-reasignación.

El proceso se hace iterativamente y se interrumpe cuando un decil inferior ya no puede recibir factores de expansión del decil superior.

Los resultados del ejercicio *Delta* antes del ajuste demográfico tienen por efecto que los microdatos del MCS 2015 en vez de arrojar una pobreza multidimensional de 36.7% de la población, la sitúan en 40.4% (5.8 puntos porcentuales por debajo de la que se desprende del 2014).

Sin embargo, el ejercicio *Delta*, al redistribuir los factores de expansión, tuvo por consecuencia una disminución de la población una vez expandidos de nueva cuenta los microdatos del MCS 2015. Esto significa que en los hogares neutralizados con factor = 1 hay, en promedio, más población que en los hogares receptores y, por ello, la sobreexpansión de éstos no alcanzó a compensar en términos poblacionales.

Al incorporarse, entonces, el ajuste demográfico vía imputación, se recupera el monto poblacional con un efecto añadido en el ingreso per cápita al que había logrado el puro efecto *Delta*, teniendo por consecuencia que, de los microdatos, se desprende ahora un porcentaje de pobreza de 42.1%: 4.1 puntos (no 9.5) inferior al del 2014.

Comentarios finales

Es cierto que este ejercicio no parte de un planteamiento consagrado en alguna literatura. Es el resultado del desarrollo de una metodología para

resolver un problema específico y, en más de un sentido, inédito, al menos en la experiencia del INEGI. Es, pues, un planteamiento original con el riesgo que conlleva. El procedimiento puede ser muy dependiente del punto de referencia que se haya elegido para alinear los deciles, tal como se señaló en el apartado 5, sin embargo, ello no significa que sea un planteamiento *ad-hoc* que no pueda generalizarse. El problema de la pérdida o colapso de frecuencias del ICT más bajos en el MCS 2015 pudo ser la consecuencia de decisiones y criterios de validación adoptados, pero no hay que descartar tampoco que cualquier acción gubernamental (federal o local) que apunte en esa dirección (o, asimismo, dificultades de acceso a zonas rurales dispersas por problemas de seguridad y/o conflictividad) sean factores que tengan el mismo efecto a futuro. Una metodología de medición de la pobreza basada en el ingreso y, en general, toda comparación intertemporal de deciles son muy vulnerables a esta problemática. En la reflexión que aquí se ofrece, quienes en años venideros se topen con perplejidades como las que dieron lugar a la controversia con el MCS 2015, podrían encontrar elementos de diagnóstico y experimentación a tomar en cuenta.

Sin duda, una limitante del ejercicio es que incurre en pérdida de diversidad cuando se neutraliza la parte superior de los deciles para sobreexpandir la parte estable en ellos. Que *Delta* sin ajuste demográfico terminara con una merma de población es indicativo de que la transferencia de factores de expansión ocurre entre hogares que no dejan de mantener diferencias entre sí. Más allá de que este ejercicio pueda o no afinarse frente a esta válida observación, no está de sobra responder con una analogía: pensemos en una persona que sufrió un accidente con una quemadura grave en uno de sus pies, se decide obtener piel de su espalda para hacer un injerto, el punto no es si la espalda es como los pies, sino si el procedimiento funciona o no.

7. Anexo estadístico

Tabla 7.1

Hogares con ingresos iguales o menores a 3 mil pesos trimestrales

Hogares con ingresos <= 3 mil pesos			Variación porcentual anual		
Año	ICT	ICM	Periodo	ICT	ICM
2010	769	2 889			
2012	791	3 209	2010-2012	1.42	5.39
2014	520	2 394	2012-2014	-18.92	-13.63
2015	93	1 275	2014-2015	-82.12	-46.74

Tabla 7.2

Ingreso corriente total^a y hogares

Decil	ICT						Hogares					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
Total	1 266 333	1 284 468	1 318 163	1 524 263	1 426 945	1 412 646	29 638 940	31 359 562	32 150 400	33 218 037	33 150 484	32 859 786
I	17 616	18 055	19 662	27 137	24 008	23 805	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
II	33 747	34 493	36 680	45 939	39 671	39 306	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
III	46 942	48 037	50 224	61 079	52 601	52 091	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
IV	60 175	61 991	63 588	76 890	66 797	66 128	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
V	75 230	77 338	78 976	94 598	82 981	82 151	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
VI	93 394	96 077	97 024	115 818	101 831	100 816	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
VII	116 317	119 993	120 349	142 938	128 102	126 822	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
VIII	149 974	154 002	154 661	179 838	165 421	163 775	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
IX	209 216	211 422	213 831	243 329	229 949	227 655	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979
X	463 721	463 061	483 167	536 696	535 585	530 099	2 963 894	3 135 956	3 215 040	3 321 804	3 315 048	3 285 979

^a Millones de pesos a precios del 2015.

Tabla 7.3

Continúa

Perceptores totales y perceptores por hogar

Decil	Perceptores totales						Perceptores por hogar					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
Total	66 344 483	72 090 107	75 381 871	80 911 445	77 858 812	77 232 956	2.2	2.3	2.3	2.4	2.3	2.4
I	5 106 269	5 373 508	5 925 336	6 226 511	5 870 873	5 819 702	1.7	1.7	1.8	1.9	1.8	1.8
II	5 836 328	6 207 704	6 725 440	6 778 853	6 393 971	6 344 101	2.0	2.0	2.1	2.0	1.9	1.9
III	5 854 619	6 388 987	6 856 385	7 219 234	6 789 882	6 722 959	2.0	2.0	2.1	2.2	2.0	2.0
IV	6 161 013	6 743 166	7 166 736	7 563 063	7 148 397	7 102 603	2.1	2.2	2.2	2.3	2.2	2.2
V	6 437 621	6 966 398	7 320 377	7 934 187	7 594 139	7 507 110	2.2	2.2	2.3	2.4	2.3	2.3
VI	6 611 770	7 257 725	7 720 394	8 203 636	7 786 673	7 746 388	2.2	2.3	2.4	2.5	2.3	2.4

Tabla 7.3

Concluye

Perceptores totales y perceptores por hogar

Decil	Perceptores totales						Perceptores por hogar					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
VII	7 189 107	7 728 185	8 046 727	8 739 751	8 536 091	8 472 292	2.4	2.5	2.5	2.6	2.6	2.6
VIII	7 522 836	8 199 028	8 492 589	9 207 690	8 829 888	8 756 333	2.5	2.6	2.6	2.8	2.7	2.7
IX	7 817 455	8 663 564	8 650 216	9 558 109	9 468 729	9 395 595	2.6	2.8	2.7	2.9	2.9	2.9
X	7 807 465	8 561 843	8 477 671	9 480 410	9 440 168	9 365 873	2.6	2.7	2.6	2.9	2.8	2.9

Tabla 7.4

Promedio de ingreso corriente total^a por hogar y por perceptor

Decil	ICT por hogar						ICT por perceptor					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
Total	42 725	40 959	41 000	45 887	43 044	42 990	19 087	17 818	17 486	18 839	18 327	18 291
I	5 944	5 757	6 116	8 169	7 242	7 245	3 450	3 360	3 318	4 358	4 089	4 090
II	11 386	10 999	11 409	13 830	11 967	11 962	5 782	5 557	5 454	6 777	6 204	6 196
III	15 838	15 318	15 622	18 387	15 867	15 852	8 018	7 519	7 325	8 461	7 747	7 748
IV	20 303	19 768	19 778	23 147	20 150	20 124	9 767	9 193	8 873	10 167	9 344	9 310
V	25 382	24 662	24 564	28 478	25 032	25 000	11 686	11 102	10 788	11 923	10 927	10 943
VI	31 511	30 637	30 178	34 866	30 718	30 681	14 125	13 238	12 567	14 118	13 078	13 015
VII	39 245	38 264	37 433	43 030	38 643	38 595	16 180	15 527	14 956	16 355	15 007	14 969
VIII	50 600	49 108	48 105	54 139	49 900	49 840	19 936	18 783	18 211	19 531	18 734	18 704
IX	70 588	67 419	66 510	73 252	69 365	69 281	26 763	24 404	24 720	25 458	24 285	24 230
X	156 457	147 662	150 283	161 568	161 562	161 321	59 395	54 084	56 993	56 611	56 735	56 599

^a Millones de pesos a precios del 2015.

Tabla 7.5

Continúa

Porcentaje de pobreza multidimensional por entidad federativa: pobreza y no pobre y no vulnerable

	Pobreza						No pobre y no vulnerable					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
E.U.M.	46.1	45.5	46.2	36.7	40.4	42.1	19.9	19.8	20.5	23.9	22.3	21.6
Ag.	38.1	37.8	34.8	25.9	29.5	30.8	27.6	27.6	31.0	33.3	31.6	30.8

**Porcentaje de pobreza multidimensional por entidad federativa:
pobreza y no pobre y no vulnerable**

	Pobreza						No pobre y no vulnerable					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
BC	31.5	30.2	28.6	24.2	28.5	31.6	24.2	23.5	26.5	31.1	27.8	26.8
BCS	31.0	30.1	30.3	22.6	24.9	28.0	31.9	32.0	29.8	32.7	32.0	30.3
Camp.	50.5	44.7	43.6	34.2	38.2	38.8	20.3	21.2	20.4	23.1	21.5	21.2
Coah. de Z.	27.8	27.9	30.2	19.0	22.3	23.9	33.7	34.9	34.5	39.6	37.1	35.6
Col.	34.7	34.4	34.3	26.0	29.0	32.7	26.7	27.6	26.1	27.9	27.3	25.5
Chis.	78.5	74.7	76.2	65.6	69.2	69.7	6.1	6.4	6.0	8.4	7.9	7.8
Chih.	38.8	35.3	34.4	21.1	24.8	26.1	25.4	26.6	27.8	36.1	33.6	32.8
Cd. de Méx.	28.5	28.9	28.4	25.5	29.3	30.7	31.7	32.1	35.7	38.7	37.2	35.6
Dgo.	51.6	50.1	43.5	32.4	37.1	38.5	18.6	17.0	21.8	25.0	22.9	22.1
Gto.	48.5	44.5	46.6	30.0	34.5	35.5	16.9	18.1	17.2	21.6	20.1	19.7
Gro.	67.6	69.7	65.2	59.3	63.2	66.2	7.5	6.4	6.0	6.3	5.6	5.4
Hgo.	54.7	52.8	54.3	41.4	46.2	47.1	13.7	13.5	14.8	16.1	14.7	14.4
Jal.	37.0	39.8	35.4	24.9	28.6	29.1	23.0	23.8	27.1	31.0	29.1	29.0
Méx.	42.9	45.3	49.6	40.7	44.6	46.2	19.3	17.4	17.4	22.8	20.8	19.9
Mich. de O.	54.7	54.4	59.2	44.4	49.7	53.8	12.3	11.5	11.7	14.0	12.3	11.8
Mor.	43.2	45.5	52.3	36.8	39.7	42.2	17.3	17.9	15.8	20.8	19.5	19.0
Nay.	41.4	47.6	40.5	31.4	35.5	36.5	20.9	18.6	23.5	24.3	21.9	21.6
NL	21.0	23.2	20.4	17.6	19.9	21.6	39.1	39.2	42.5	44.2	41.9	40.2
Oax.	67.0	61.9	66.8	59.7	64.3	66.6	9.5	10.3	7.9	10.1	9.3	9.2
Pue.	61.5	64.5	64.5	50.1	52.4	54.9	11.3	9.4	10.5	12.8	12.3	12.0
Qro.	41.4	36.9	34.2	24.0	26.5	29.6	21.8	24.0	24.9	30.6	29.6	28.0
Q. Roo	34.6	38.8	35.9	30.0	34.0	36.8	24.6	24.6	23.7	25.4	24.4	23.3
SLP	52.4	50.5	49.1	36.5	40.3	42.6	19.5	18.3	19.0	24.2	22.6	21.1
Sin.	36.7	36.3	39.4	27.5	30.1	32.6	22.3	23.8	23.4	27.2	25.6	24.6
Son.	33.1	29.1	29.4	22.5	26.0	27.7	28.4	29.6	31.3	35.0	33.4	32.5
Tab.	57.1	49.7	49.6	36.7	41.0	42.9	11.5	13.4	10.6	11.2	10.7	10.7
Tamps.	39.0	38.4	37.9	32.0	36.2	37.0	24.8	26.0	26.5	29.3	25.4	24.9
Tlax.	60.3	57.9	58.9	43.3	47.2	49.2	13.0	12.2	10.8	16.5	15.2	14.5
Ver. de l. de la Ll.	57.6	52.6	58.0	50.7	54.5	54.9	14.3	12.8	12.2	13.1	11.7	11.7
Yuc.	48.3	48.9	45.9	34.4	37.2	39.9	19.2	17.9	19.5	23.9	22.4	21.4
Zac.	60.2	54.2	52.3	44.7	47.8	47.9	14.5	19.0	16.0	19.8	18.2	18.1

Tabla 7.6

**Porcentaje de pobreza multidimensional por entidad federativa:
pobreza moderada y pobreza extrema**

	Pobreza moderada						Pobreza extrema					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
E.U.M.	34.8	35.7	36.6	30.9	33.4	34.6	11.3	9.8	9.5	5.8	7.0	7.5
Ags.	34.4	34.4	32.7	24.4	27.6	28.9	3.8	3.4	2.1	1.5	1.9	1.9
BC	28.1	27.5	25.5	22.6	26.2	29.2	3.4	2.7	3.1	1.7	2.3	2.4
BCS	26.4	26.4	26.3	20.9	22.9	25.5	4.6	3.7	3.9	1.7	1.9	2.5
Camp.	36.7	34.2	32.5	29.7	32.3	32.5	13.8	10.4	11.1	4.6	5.9	6.3
Coah. de Z.	24.9	24.7	26.4	17.6	20.2	21.7	2.9	3.2	3.7	1.4	2.0	2.2
Col.	32.2	30.4	30.9	24.1	26.8	29.8	2.5	4.0	3.4	1.9	2.3	2.8
Chis.	40.2	42.5	44.4	40.3	40.8	41.3	38.3	32.2	31.8	25.3	28.4	28.5
Chih.	32.2	31.5	28.9	20.3	23.4	24.5	6.6	3.8	5.4	0.9	1.4	1.6
Cd. de Méx.	26.4	26.4	26.7	24.7	28.0	29.3	2.2	2.5	1.7	0.9	1.3	1.4
Dgo.	41.1	42.6	38.2	30.1	33.8	34.6	10.5	7.5	5.3	2.2	3.2	3.9
Gto.	40.1	37.6	41.0	28.3	32.1	33.1	8.4	6.9	5.5	1.7	2.3	2.3
Gro.	35.7	38.0	40.8	41.4	43.0	44.3	31.8	31.7	24.5	17.9	20.2	21.9
Hgo.	41.2	42.8	42.0	34.1	37.5	38.2	13.5	10.0	12.3	7.3	8.7	8.9
Jal.	31.8	34.0	32.1	22.5	25.7	26.1	5.3	5.8	3.2	2.4	2.9	3.0
Méx.	34.3	39.5	42.4	36.4	39.4	40.3	8.6	5.8	7.2	4.3	5.1	5.9
Mich. de O.	41.2	39.9	45.2	35.8	39.2	42.4	13.5	14.4	14.0	8.6	10.5	11.3
Mor.	36.3	39.1	44.4	31.4	32.4	34.1	6.9	6.3	7.9	5.4	7.3	8.0
Nay.	33.0	35.7	32.0	27.1	30.2	31.2	8.3	11.9	8.5	4.3	5.2	5.3
NL	19.2	20.8	19.0	16.6	18.7	19.9	1.8	2.4	1.3	1.0	1.3	1.6
Oax.	37.7	38.6	38.4	40.7	42.3	43.1	29.2	23.3	28.3	19.0	22.0	23.5
Pue.	44.5	46.9	48.4	42.9	43.9	45.6	17.0	17.6	16.2	7.2	8.5	9.2
Qro.	34.0	31.8	30.3	21.7	24.1	26.9	7.4	5.2	3.9	2.4	2.4	2.8
Q. Roo	28.2	30.4	28.9	25.0	27.4	29.6	6.4	8.4	7.0	5.0	6.6	7.2
SLP	37.1	37.7	39.6	30.5	33.0	34.9	15.3	12.8	9.5	6.0	7.3	7.7
Sin.	31.2	31.8	34.1	25.7	27.8	29.9	5.5	4.5	5.3	1.7	2.3	2.7
Son.	28.0	24.2	26.1	20.4	23.3	24.7	5.1	5.0	3.3	2.1	2.7	3.0
Tab.	43.5	35.4	38.5	31.0	33.3	34.4	13.6	14.3	11.0	5.7	7.7	8.5
Tamps.	33.5	33.7	33.6	28.7	31.7	32.7	5.5	4.7	4.3	3.3	4.4	4.3
Tlax.	50.4	48.8	52.4	40.5	43.5	44.8	9.9	9.1	6.5	2.8	3.7	4.4
Ver. de l. de la Ll.	38.8	38.4	40.9	40.3	41.6	41.7	18.8	14.3	17.2	10.4	12.9	13.2
Yuc.	36.6	39.0	35.2	28.6	30.5	32.7	11.7	9.8	10.7	5.8	6.7	7.2
Zac.	49.4	46.7	46.6	40.1	42.4	42.5	10.8	7.5	5.7	4.6	5.4	5.5

Tabla 7.7

Porcentaje de pobreza multidimensional por entidad federativa: vulnerable por carencias sociales y vulnerable por ingresos

	Vulnerable por carencias sociales						Vulnerable por ingresos					
	2010	2012	2014	2015			2010	2012	2014	2015		
				Original	Delta sin ajuste	Delta con ajuste				Original	Delta sin ajuste	Delta con ajuste
E.U.M.	28.1	28.6	26.3	33.9	31.3	29.8	5.9	6.2	7.1	5.4	6.0	6.6
Ags.	26.2	24.5	25.3	33.1	30.8	29.6	8.1	10.1	8.9	7.7	8.1	8.8
BC	37.9	37.6	38.3	38.8	36.7	33.5	6.3	8.6	6.6	5.9	7.0	8.1
BCS	32.5	30.0	34.4	40.1	37.8	35.1	4.5	7.9	5.6	4.6	5.4	6.7
Camp.	24.9	28.6	32.1	39.8	36.9	36.4	4.3	5.6	4.0	2.9	3.4	3.6
Coah. de Z.	25.6	24.4	24.2	30.9	28.8	27.7	12.9	12.7	11.1	10.5	11.8	12.8
Col.	33.8	31.7	33.2	41.0	38.6	35.0	4.9	6.3	6.4	5.0	5.1	6.8
Chis.	13.0	17.2	15.3	24.1	21.3	20.8	2.4	1.7	2.5	1.9	1.7	1.6
Chih.	22.8	27.4	25.8	34.7	33.5	32.4	13.0	10.7	12.0	8.1	8.2	8.7
Cd. de Méx.	34.4	32.4	28.0	28.9	26.2	25.1	5.4	6.6	8.0	6.9	7.3	8.6
Dgo.	21.1	21.8	24.3	35.0	32.7	31.4	8.8	11.1	10.5	7.6	7.3	7.9
Gto.	28.9	32.6	27.7	43.7	39.9	38.9	5.7	4.9	8.6	4.7	5.5	5.9
Gro.	23.0	21.7	26.2	32.5	29.6	26.7	2.0	2.3	2.6	1.8	1.6	1.7
Hgo.	27.5	30.7	25.8	39.3	35.7	34.7	4.2	3.0	5.1	3.2	3.4	3.8
Jal.	33.7	28.3	29.6	38.7	35.6	35.1	6.2	8.1	7.9	5.4	6.6	6.7
Méx.	32.2	29.5	23.7	27.9	25.1	23.8	5.6	7.8	9.3	8.6	9.5	10.1
Mich. de O.	28.6	30.7	25.0	38.6	34.4	30.1	4.3	3.5	4.1	3.0	3.7	4.3
Mor.	33.6	32.0	25.7	38.5	36.4	34.1	5.8	4.6	6.1	3.9	4.4	4.7
Nay.	33.4	28.2	29.4	38.8	35.6	34.5	4.3	5.6	6.5	5.5	7.1	7.3
NL	31.6	29.1	28.0	30.2	28.6	27.4	8.2	8.4	9.1	8.0	9.5	10.8
Oax.	22.2	26.1	23.3	28.5	24.8	22.4	1.3	1.7	2.1	1.6	1.6	1.8
Pue.	21.6	22.0	19.8	32.8	30.8	28.3	5.6	4.2	5.1	4.3	4.5	4.9
Qro.	31.7	32.6	33.1	40.4	38.2	35.6	5.0	6.4	7.8	4.9	5.7	6.8
Q. Roo	36.1	30.4	34.2	39.8	36.8	34.1	4.7	6.2	6.3	4.8	4.8	5.8
SLP	20.9	24.7	24.3	34.4	32.4	30.5	7.2	6.6	7.6	4.9	4.8	5.8
Sin.	33.3	33.5	30.4	40.4	38.5	36.1	7.7	6.4	6.9	4.9	5.8	6.7
Son.	31.6	36.6	32.1	36.8	34.4	32.9	6.8	4.7	7.3	5.7	6.1	6.9
Tab.	27.2	34.0	37.5	50.5	46.7	44.5	4.1	3.0	2.3	1.6	1.6	1.9
Tamps.	26.9	26.8	24.4	31.7	30.8	30.0	9.4	8.8	11.2	6.9	7.6	8.1
Tlax.	19.3	23.9	22.6	33.7	30.7	28.5	7.4	6.0	7.7	6.4	7.0	7.8
Ver. de l. de la Ll.	23.6	30.6	24.8	33.3	30.5	30.0	4.5	4.0	5.0	2.9	3.3	3.3
Yuc.	26.0	27.0	27.7	36.9	35.3	32.8	6.4	6.3	7.0	4.8	5.1	5.8
Zac.	18.4	20.4	24.4	30.5	28.8	28.7	6.9	6.4	7.3	5.0	5.1	5.3

8. Anexo: precisiones estadísticas

Tabla 8.1

Precisiones estadísticas de variables seleccionadas para información de los ejercicios original y Delta

Variables seleccionadas	Estimación			Error estándar			Coeficiente de variación			Intervalo 95%, ¿se traslapa?
	Original	Delta 6	Diferencia	Original	Delta 6	Diferencia	Original	Delta 6	Diferencia	
Integrantes del hogar										
Total de integrantes	3.65	3.69	0.04	0.011	0.013	0.002	0.003	0.004	0.001	SÍ
Hombres	1.78	1.79	0.01	0.007	0.008	0.001	0.007	0.008	0.001	SÍ
Mujeres	1.87	1.90	0.03	0.007	0.008	0.001	0.007	0.008	0.001	SÍ
12 y más años	2.88	2.90	0.03	0.009	0.010	0.001	0.009	0.010	0.001	SÍ
Menos de 12 años	0.78	0.79	0.01	0.006	0.007	0.001	0.006	0.007	0.001	SÍ
De 12 a 64 años	2.59	2.60	0.01	0.009	0.011	0.002	0.009	0.011	0.002	SÍ
65 y más años	0.28	0.30	0.02	0.004	0.004	0.001	0.004	0.004	0.001	SÍ
Ocupados	1.64	1.57	-0.07	0.007	0.007	0.001	0.007	0.007	0.001	NO
Perceptores de ingreso	2.44	2.35	-0.09	0.009	0.010	0.001	0.009	0.010	0.001	NO
Perceptores ocupados	1.59	1.52	-0.08	0.006	0.007	0.001	0.006	0.007	0.001	NO
Ingresos por hogar según la fuente										
Ingreso corriente total	45 887	42 990	-2 896	466.10	480.30	14.20	466.102	480.300	14.198	NO
Por trabajo	30 790	28 414	-2 376	236.25	249.85	13.60	236.250	249.855	13.605	NO
Asalariado	25 853	23 791	-2 062	224.41	233.92	9.50	224.414	233.918	9.504	NO
Sueldos	20 478	18 873	-1 605	186.82	193.51	6.68	186.824	193.508	6.684	NO
Horas extra	199.32	181.93	-17.38	7.07	7.01	-0.07	7.074	7.007	-0.067	SÍ
Comisiones	838.47	772.44	-66.03	27.61	28.11	0.50	27.614	28.113	0.499	SÍ
Aguinaldo	1 201.95	1 100.57	-101.38	24.18	24.78	0.60	24.181	24.776	0.596	NO
Indemnizaciones de trabajo	110.43	103.91	-6.52	10.09	10.18	0.09	10.085	10.177	0.091	SÍ
Otras remuneraciones	608.57	554.50	-54.07	16.06	15.99	-0.07	16.061	15.993	-0.068	SÍ
Remuneraciones especiales	2 416.31	2 204.89	-211.42	58.51	58.09	-0.42	58.510	58.091	-0.419	SÍ
Trabajo independiente	3 568.35	3 329.90	-238.45	87.59	86.75	-0.84	87.591	86.751	-0.840	SÍ
No agropecuario	2 942.30	2 733.13	-209.18	68.14	66.58	-1.57	68.143	66.575	-1.568	SÍ
Industria	614.08	569.51	-44.57	23.22	23.08	-0.13	23.219	23.085	-0.135	SÍ
Comercio	1 125.81	1 041.30	-84.51	44.26	43.82	-0.43	44.258	43.825	-0.433	SÍ
Servicios	1 202.42	1 122.32	-80.10	44.42	44.31	-0.11	44.419	44.311	-0.108	SÍ
Agropecuario	626.05	596.78	-29.27	54.00	54.00	-0.00	54.004	54.001	-0.004	SÍ
Agrícola	302.79	294.88	-7.91	33.12	33.67	0.55	33.116	33.669	0.552	SÍ
Pecuario	288.90	268.27	-20.63	36.46	35.60	-0.85	36.456	35.603	-0.853	SÍ
Reproducción de animales	7.68	7.66	-0.01	1.79	2.25	0.46	1.792	2.247	0.456	SÍ
Pesca	26.68	25.96	-0.71	5.63	5.40	-0.22	5.629	5.404	-0.224	SÍ
Otros trabajos	1 369.09	1 292.91	-76.17	25.05	26.29	1.24	25.053	26.291	1.237	SÍ
Rentas	3 276.77	3 177.25	-99.53	326.12	327.17	1.05	326.115	327.168	1.053	SÍ
Utilidad	2 710.32	2 647.57	-62.75	313.56	314.81	1.25	313.558	314.806	1.249	SÍ
Arrendamiento	566.46	529.68	-36.78	39.77	39.08	-0.69	39.774	39.084	-0.691	SÍ
Transferencias	6 779.64	6 555.60	-224.04	105.63	105.77	0.14	105.635	105.772	0.137	SÍ
Jubilación	2 892.43	2 750.45	-141.99	72.63	73.06	0.42	72.634	73.056	0.421	SÍ
Becas	172.44	163.19	-9.26	23.38	21.54	-1.84	23.379	21.541	-1.838	SÍ
Donativos	962.05	953.74	-8.31	23.74	25.68	1.94	23.741	25.682	1.942	SÍ
Remesas	326.68	309.31	-17.37	12.73	13.86	1.13	12.730	13.862	1.131	SÍ
Beneficios del gobierno	813.44	814.02	0.58	11.14	12.21	1.07	11.144	12.214	1.070	SÍ
Transferencias de otros hogares	1 021.89	991.19	-30.70	53.25	52.82	-0.43	53.253	52.818	-0.435	SÍ
Transferencias de instituciones	590.70	573.71	-16.99	34.00	34.39	0.40	33.996	34.393	0.397	SÍ
Alquiler de la vivienda	4 999.79	4 800.50	-199.29	65.26	65.98	0.73	65.255	65.982	0.727	SÍ
Otros ingresos	40.20	42.87	2.67	5.36	6.56	1.20	5.359	6.557	1.197	SÍ

Propuesta para recuperar la continuidad del MCS 2015 usando máquinas de soporte vectorial

A proposal for recovering MCS 2015's continuity using
Support Vector Machines

José Alejandro Ruiz Sánchez*

* Instituto Nacional de Estadística y Geografía (INEGI), jose.ruiz@inegi.org.mx

Nota: se agradece la colaboración e ideas proporcionadas por Miriam Romo y Benito Durán.



Robot sitting on chair, reading book/imagewerks/Getty Images

Debido a procesos de mejora llevados a cabo durante el levantamiento de la información del Módulo de Condiciones Socioeconómicas (MCS) 2015, la distribución del ingreso corriente total cambió de manera sustantiva con respecto a lo que se venía observando en levantamientos previos. Esta propuesta intenta recomponer la continuidad, modificando la representatividad que tenía un hogar en el 2015, de acuerdo con su asignación a un decil pronosticado o artificial creado a partir de un conjunto de variables monetarias y no-monetarias. Para asignar un hogar a un decil artificial, se usó una máquina de soporte vectorial. Como conjunto de entrenamiento se usaron los levantamientos ENIGH-MCS 2010, 2012 y 2014. Una vez realizada la imputación de deciles se ajustaron los factores de expansión de cada hogar, esto permite que hogares con ingresos bajos sean sobreexpandidos y que, para hogares con ingresos altos, su factor de expansión sea reducido.

Palabras clave: Módulo de Condiciones Socioeconómicas; MCS; continuidad; SVM; factores de expansión.

1. Introducción

La información relacionada con los ingresos de los hogares captada en el Módulo de Condiciones Socioeconómicas (MCS) 2015 perdió continuidad con levantamientos anteriores del mismo módulo; sin embargo, ello no invalidó ni volvió erróneos o sesgados los datos obtenidos. Es en este sentido que el problema de continuidad histórica puede ser abordado como uno de (re)clasificación de hogares, donde los conjuntos generados estén constituidos por observaciones con características similares.

Una manifestación de la pérdida de continuidad entre levantamientos con respecto a variables de ingresos se dio al realizar análisis por deciles: las tasas de crecimiento del 2014 al 2015 del ingreso corriente total (ICT) de cada decil fueron muy superiores a las que se venían observando históricamente.

El acercamiento metodológico en este documento consiste en pronosticar el decil al cual de-

Due to improvements in the Socio-economic Module's data collection 2015 (MCS 2015 by its Spanish acronym), the total income distribution shifted from what has been observed in recent years. Consequently, it is no longer possible to make income comparison with previous information. This methodology is intended to recompose continuity by modifying household weights in 2015, according to its classification into some artificial decile generated by monetary and non-monetary variables. To classify a household we use a technique called Support Vector Machine (SVM) and historic data (2010-2014) as a training data set. Once an artificial decile is imputed to a household in the MCS 2015, we modify its weight so that adding them up by decile represent 10 percent. This allows low-income households to be overexpanded and high-income households to be underexpanded.

Key words: Socio-economic Module; MCS; Continuity; SVM; weights.

bería pertenecer cada hogar en el 2015 a partir de información monetaria y no-monetaria, de forma tal que los hogares que pertenezcan a determinado decil sean consistentes con hogares en deciles iguales, pero de años anteriores. Con base en esta clasificación pronosticada o artificial, se hace un ajuste en los factores de expansión para otorgarle mayor o menor peso a cada hogar, de acuerdo con el decil al que haya sido asignado. De esta forma, hogares con ingresos bajos tendrán un mayor peso muestral que el que tenían originalmente, contrario a los de ingresos altos, quienes verán disminuido su peso muestral. Para pronosticar el decil de pertenencia de los hogares en el 2015, se utilizaron técnicas de *Machine Learning* para clasificación, específicamente máquinas de soporte vectorial (SVM, por sus siglas en inglés).

Así, la presente propuesta se desarrolló con el objetivo de generar una base de datos alternativa al MCS 2015 que recomponga la trayectoria histórica del ingreso y que tenga como prioridad el no modificar lo declarado por los hogares. Para lograr

cambios, entonces, se adecuarán los factores de expansión de tal forma que se modifique la representatividad original que tiene el ingreso declarado. El cambio en los factores de expansión se hará a partir de asignar un hogar a un decil pronosticado, al que teóricamente debería pertenecer, de acuerdo con valores que este hogar reportó sobre un conjunto de variables monetarias y no-monetarias. Por otra parte, y como consecuencia, cuando se usan los nuevos factores de expansión, es posible que la representatividad de subgrupos poblacionales en los cuales se basó el diseño muestral se vea alterada.

El documento se encuentra organizado en seis capítulos; en el segundo de ellos se presenta una descripción de los datos que dan sustento a la propuesta metodológica. En el tercer apartado se desarrolla la técnica de SVM, que es utilizada para la categorización y pronóstico; además, se muestra la forma para modificar los factores de expansión. En el cuarto capítulo se presentan resultados desglosados por entidad federativa y por decil. En el quinto apartado se realiza una validación de la técnica con años anteriores y en el último se muestran las conclusiones.

2. Marco teórico

Características propias al levantamiento del MCS 2015 generaron que la distribución del vector de ingresos se recorriera de manera importante con respecto a años anteriores (ver figura 2.1).

Ello implica que al hacer comparaciones que consideren agrupaciones a través de ingresos (como *decilizaciones*) se estén mezclando grupos con niveles de ingreso muy distintos. En el caso de las comparaciones por deciles del MCS 2014 y del MCS 2015 se están incluyendo hogares en un decil que, bajo otras circunstancias de recolección de información, pertenecerían a otro (ver figura 2.2).

El mecanismo que permita usar información del MCS para generar mediciones consistentes en el tiempo debe incluir el comportamiento histórico de los datos para realizar algún tipo de ajuste en

el MCS 2015. Si bien un camino es el ajuste directo al vector de ingresos, la propuesta del presente documento contempla respetar la información proporcionada por los hogares, actuando por medio de la modificación de los factores de expansión. La idea intuitiva del método es generar continuidad a través de la imputación (a cada hogar del 2015) del decil de pertenencia, usando el comportamiento histórico (2010, 2012 y 2014) de *decilización* y la relación que guarda éste con variables monetarias y/o no-monetarias.

Este método será efectivo en la medida en que las técnicas de clasificación sean certeras y en que los ajustes en los factores de expansión hagan que los hogares de determinado decil x sean representativos de los hogares que, bajo circunstancias similares a las de levantamientos anteriores al 2015, hubiesen pertenecido al mismo decil x .

Figura 2.1
Gráfica de densidad del ICT (precios del 2015)

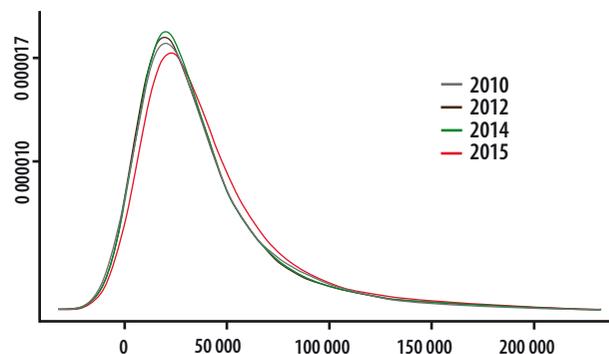
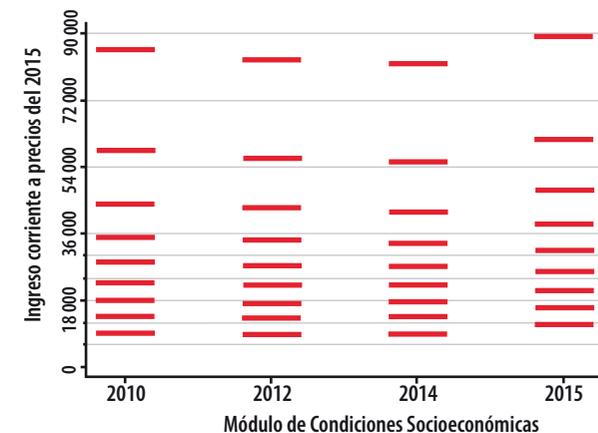


Figura 2.2
Puntos de corte para la *decilización*



3. Metodología

3.1 Su aplicación

Existen distintos métodos que permiten pronosticar la clasificación de nuevas observaciones con base en parámetros calibrados a partir de una base de datos alterna donde las observaciones ya se encuentran categorizadas. Un conjunto de estos métodos se encuentra en lo que conocemos como *Machine Learning*, así como en la estadística clásica; algunos de ellos son: análisis discriminante lineal (LDA, por sus siglas en inglés), redes neuronales artificiales (ANN, por sus siglas en inglés), árboles de decisión, SVM y regresión logística multinomial, entre otras.

Aunque se utilizaron distintas técnicas para comparar niveles de precisión, solo desarrollaremos la técnica SVM, ya que los resultados que se presentan son con base en ella y por haber sido la que generó resultados aceptables de manera consistente bajo diversas especificaciones. La idea intuitiva de SVM es generar hiperplanos separadores que permitan identificar aquellas observaciones pertenecientes a una misma categoría (ver figura 3.1a).

Para su construcción, los hiperplanos deben ser tales que se tenga el menor error de clasificación; es decir, las clasificaciones pueden no ser perfectas (ver figura 3.1b). Para pronosticar la clasificación de una observación nueva, primero se entrena a la máquina con una base de datos ya clasificada y, a

partir de los resultados, la máquina decide el grupo de pertenencia de la nueva observación.

De manera formal,¹ sea un conjunto de N parejas $(x_1, y_1), (x_2, y_2) \dots (x_N, y_N)$, donde $x_i \in R^p$ y $y_i \in \{-1, 1\}$. La variable y_i define la categoría de pertenencia y x el conjunto de covariables.

Sea el hiperplano definido por:

$$\{x: f(x) = x^T \beta + \beta_0 = 0\} \quad (1)$$

donde $\|\beta\| = 1$. Una regla de clasificación está dada por la función signo: $G(x) = \text{signo}[x^T \beta + \beta_0]$. En el caso de clases totalmente separables, existe una función:

$$f(x) = x^T \beta + \beta_0 = 0, \text{ tal que } y_i * f(x_i) > 0 \forall i \quad (2)$$

y, por lo tanto, podemos encontrar el hiperplano separador que tenga la mayor distancia con respecto a los puntos. A tal distancia se le conoce como margen (M).

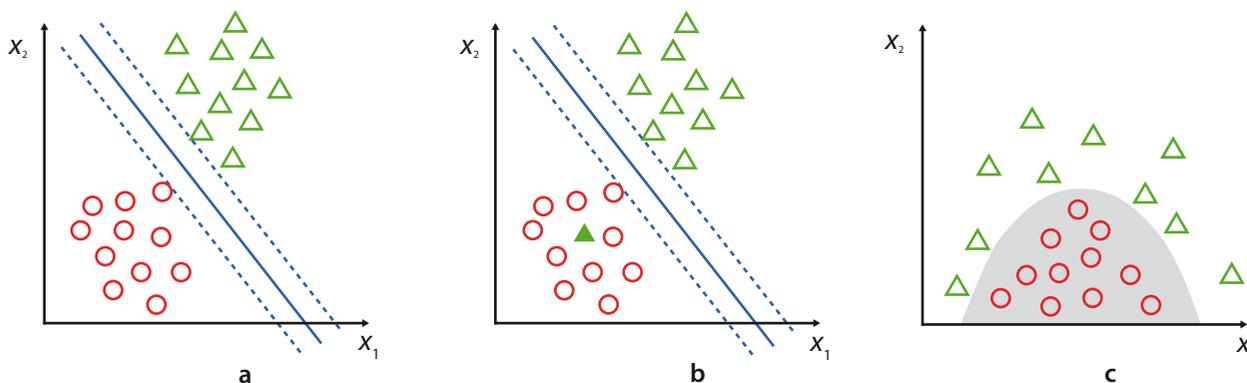
Entonces, el problema de optimización sería el siguiente:

$$\begin{aligned} & \max M \\ \text{s.a. } & y_i * (x_i^T \beta + \beta_0) \geq M, \quad i = 1, 2, \dots, N \end{aligned} \quad (3)$$

¹ El presente apartado es un extracto del capítulo 12 del libro *The Elements of Statistical Learning. Data Mining. Inference and Prediction*. Para mayor detalle, se recomienda consultarlo directamente.

Figura 3.1

Representación gráfica de un hiperplano separador



Cuando no puede darse una separación perfecta de las clases, se permite que algunos puntos se encuentren mal clasificados a través de la siguiente especificación:

$$\begin{aligned} & \max M \\ & \text{s.a.:} \\ & y_i^*(x_i^T\beta + \beta_0) \geq M(1 - \xi_i), \quad i = 1, 2, \dots, N \quad (4) \end{aligned}$$

$\xi_i \geq 0$ y $\sum \xi_i \leq C$, donde C es una constante

cuya reformulación en términos de β quedaría de la siguiente forma:

$$\begin{aligned} & \min \frac{1}{2} \|\beta\|^2 \\ & \text{s.a.:} \\ & y_i^*(x_i^T\beta + \beta_0) \geq (1 - \xi_i), \quad i = 1, 2, \dots, N \quad (5) \end{aligned}$$

$\xi_i \geq 0$ y $\sum \xi_i \leq C$, donde C es una constante.

Formamos la función de Lagrange con sus respectivos multiplicadores α_i, μ_i .

$$L = \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^N \xi_i - \sum_{i=1}^N \alpha_i [y_i(x_i^T\beta + \beta_0) - (1 - \xi_i)] - \sum_{i=1}^N \mu_i \xi_i \quad (6)$$

que en su forma más conocida es representada como:

$$L = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{i'=1}^N \alpha_i \alpha_{i'} y_i y_{i'} \langle h(x_i), h(x_{i'}) \rangle \quad (7)$$

donde $K(x, x') = \langle h(x), h(x') \rangle$ es la función kernel, la cual puede tomar diferentes especificaciones para ampliar el espacio de las covariables. Las especificaciones más comunes son las siguientes:

$$\begin{aligned} \text{Lineal} \quad & K(x, x') = \langle x, x' \rangle \\ \text{Polinomial de grado } d \quad & K(x, x') = (\langle x, x' \rangle)^d \\ \text{Radial} \quad & K(x, x') = \exp(-\gamma \|x - x'\|^2) \end{aligned}$$

Dos de estas funciones kernel permiten capturar no linealidades como las de la figura 3.1c.

Para extender la metodología al caso de Q clases, existen diversos enfoques, dos de los más usados se explican a continuación: enfoque de *clasificación uno vs. uno*, construye $\binom{Q}{2}$ SVM's, cada uno del cual compara un par de clases. Para clasificar una observación nueva se usan los $\binom{Q}{2}$ clasificadores y se cuenta el número de veces que dicha observación cae en cada una de las Q clases. La clasificación final se hace asignando la observación nueva a aquella clase a la cual fue asignada más veces. El segundo es el de *clasificación uno vs. todos*, en el cual se realizan Q SVM's, comparando una clase con las $Q-1$ restantes; una nueva observación x^* será asignada a aquella categoría para la cual $\beta_{0q} + \beta_{1q}x_1^* + \beta_{2q}x_2^* + \dots + \beta_{pq}x_p^*$ es mayor.

En nuestro caso utilizaremos como datos de entrenamiento los deciles (que tendrán el rol de clases o categorías) y las covariables a nivel hogar de los MCS 2010, 2012 y 2014;² con esta información se *entrena* a la máquina para que aprenda a realizar la *decilización* y ese aprendizaje será aplicado a la base de datos del 2015, de tal forma que la máquina de soporte vectorial asignará un decil pronosticado x a un hogar en el 2015, siempre y cuando los niveles de las covariables para ese hogar sean similares a aquellos hogares pertenecientes al mismo decil x en años anteriores.

La base de datos de la cual partimos para la *decilización* artificial es aquella proveniente del artículo "Ajuste demográfico por imputación" (publicado en este mismo número especial de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*, y cuyo objetivo es únicamente reconstituir la estructura demográfica del MCS 2015). Una vez obtenido el vector de deciles pronosticados, la siguiente etapa consiste en modificar los factores de expansión, de forma tal que hogares con menores ingresos estén sobrerrepresentados y hogares con ingresos mayores sean subrepresentados.

La forma de reasignar pesos muestrales está basada en los deciles pronosticados a los que perte-

² Las variables monetarias se encuentran en pesos del 2015.

nece cada hogar: cuando un hogar es asignado a un decil pronosticado o artificial x , esa asignación no considera el factor de expansión de dicho hogar y, por lo tanto, no considera el número de hogares a los que representa. Por otra parte, y dado que los montos del ingreso corriente que definen los cortes de la *decilización* son similares durante el periodo del 2010 al 2014 y que son visiblemente superiores en el 2015 (ver figura 2.2), la *decilización* artificial (la cual considera el comportamiento histórico) categorizará en el decil I a un número de hogares que representarán menos del 10% poblacional, de acuerdo con sus factores de expansión; es decir, el número de hogares muestrales que la máquina de soporte vectorial asignará al decil I será menor al número de hogares que caen en el mismo decil I cuando se consideran los datos del MCS 2015. Dado que ese menor número de hogares en el decil I representa menos del 10% poblacional, habrá que incrementar sus factores de expansión; esto permitirá que hogares con ingresos más bajos sean sobreexpandidos, mientras que los de ingresos mayores verán disminuidos sus factores de expansión; es decir, que representarán a una menor cantidad de hogares.

Cabe mencionar que esta *decilización* artificial solo se utiliza en esta etapa de modificación de factores de expansión; una vez creados los nuevos factores, nos olvidamos por completo de los deciles artificiales. Para la *decilización* final, usada en los análisis de resultados de la cuarta sección, se recurrió al vector de ingreso corriente total, pero incorporando los nuevos factores de expansión.

De manera formal: sea un hogar h y fh su factor de expansión, al formar deciles se debe cumplir que:

$$\frac{\sum_{h \in d} fh}{\sum_{d=1}^{10} \sum_{h \in d} fh} = 0.1 \quad (a)$$

para cada decil d . Sin embargo, dado que los factores de expansión no son considerados en la *decilización* artificial, nada garantiza que se cumpla

la igualdad (a) al utilizar la técnica de SVM para categorizar hogares ya que, en este caso, el número de hogares pertenecientes a cada decil se asigna a partir de un criterio de optimización, donde la condición (a) no tiene rol alguno; por lo tanto, para cumplir con la condición del 10% se generará un nuevo vector de factores de expansión $fh^{(n)}$, el cual resultará del ajuste a los factores de expansión originales del MCS 2015.

Sea h un hogar perteneciente al MCS 2015 y sea fh el factor de expansión original asociado a ese hogar. Denotemos como $h_{d,svm}$ a un hogar que fue clasificado en el decil d utilizando la técnica SVM (decil pronosticado).

Entonces, el factor de ajuste para los hogares del decil pronosticado d se define como:

$$A^d = \frac{\sum_h fh/10}{\sum_{h_{d,svm}} fh}$$

y el nuevo factor de expansión para un hogar que pertenece al decil pronosticado d , es:

$$fh_d^{(n)} = (A^d) * fh$$

Este proceso se realiza para cada una de las entidades federativas por separado, con lo cual se procura cuidar la representatividad por entidad en el agregado nacional.

3.2 Nota técnica

- El *software* estadístico que se utilizó para la metodología descrita en el apartado anterior fue *R*, a través de su IDE *RStudio* versión 1.0.44. Las librerías usadas fueron MASS y NNET. Para el caso particular de la función SVM, el kernel usado fue el lineal, con los parámetros por *default*.

- Se utilizaron los MCS 2010, 2012 y 2014 como base de datos de entrenamiento para el algoritmo de SVM.

Se usaron tres bases de datos dentro de cada MCS: población, vivienda y hogares. De la segunda se utilizaron variables asociadas a las características físicas de la vivienda, como: tipo de piso, paredes, dotación de agua, etcétera. De la tercera se emplearon tanto el ICT como variables sobre la pertenencia de bienes muebles (autos y televisores, por ejemplo). Por último, de la base de población se usaron índices por hogar sobre características sociodemográficas de sus integrantes: proporción de miembros del hogar que se encuentran estudiando primaria, promedio de edad de los integrantes, proporción de integrantes que cuentan con estudios universitarios, etcétera. La tabla del *Anexo* contiene una descripción de las variables usadas en el 2015, así como el tratamiento que se les dio para su incorporación como covariables dentro de las diversas técnicas³ de *Machine Learning* usadas. Para los MCS 2010, 2012 y 2014 se utilizaron las mismas variables.

4. Resultados

4.1 Resultados a nivel nacional, por entidad federativa y por decil con la nueva base del MCS 2015

Una deducción de la lectura del apartado 3 es que a nivel microdatos no se afectó el vector de ingresos reportado por los hogares; son los cambios en los factores de expansión por hogar los que permiten tener resultados distintos a los del MCS 2015.

En esta sección se muestran los principales cambios respecto al ingreso y la pobreza bajo distintos

³ Entre las técnicas probadas se encuentran redes neuronales, árboles de decisión y regresión logística multinomial.

niveles de agregación. La gráfica 4.1 muestra los efectos en el ICT por decil. Con respecto al 2014, el MCS 2015 presentó un incremento de 15.6% en el ingreso corriente total, mientras que con la presente propuesta de ajuste, el incremento para el mismo periodo fue de 4.5 por ciento.

Para los resultados a nivel entidad federativa, se presentan las tablas 4.1 y 4.2. La primera tiene información sobre los ingresos en valores absolutos y a precios del 2015; la segunda, información en términos de variaciones porcentuales con respecto al levantamiento del 2014.

En la tabla 4.2 se puede observar que la tasa de variación tanto del ingreso corriente como del promedio por hogar para todas las entidades federativas disminuyó como consecuencia del ajuste a los factores de expansión.

En términos de fuentes de ingreso, la tabla 4.3 muestra sus variaciones porcentuales entre el 2014 y 2015; este último año considerando el ajuste propuesto.

Si bien los ajustes en las tasas de crecimiento de las variables monetarias son a la baja, faltaría verificar si esta propuesta tiene el efecto esperado en pobreza; es decir, un aumento con respecto al MCS 2015 y un alineamiento con los años anteriores; para ello se presenta una serie de tablas a nivel entidad federativa con los distintos tipos de pobreza calculados por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), los cuales ya incorporan el ajuste correspondiente en el 2015.

La tabla 4.4 muestra los resultados de pobreza para cada uno de los estados y para el país. Debido a que esta metodología considera datos no solo del 2014 sino que también incorpora información anterior de entidades federativas, el comportamiento de los nuevos resultados de pobreza (aquellos resultantes de ajustar factores de expansión, con la técnica descrita en apartados anteriores) es más parecido al 2010 y el 2012 que al 2014.

Gráfica 4.1

Variación porcentual del ICT para el periodo 2014-2015, por deciles

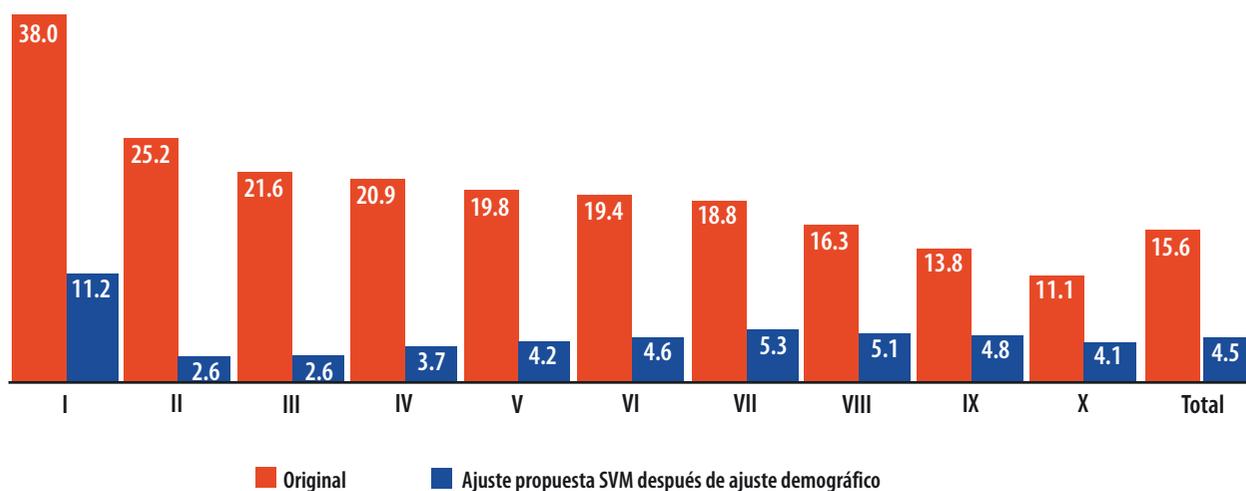


Tabla 4.1

Ingreso corriente promedio por hogar

Continúa

Entidad federativa	2012	2014	2015	2015 SVM
1 Aguascalientes	44 149	47 529	59 434	54 942
2 Baja California	49 258	49 017	54 616	50 705
3 Baja California Sur	52 094	49 327	61 361	58 056
4 Campeche	42 933	44 388	50 782	47 178
5 Coahuila de Z.	46 528	51 407	53 301	51 419
6 Colima	43 599	45 443	48 293	44 391
7 Chiapas	24 118	22 417	29 648	25 599
8 Chihuahua	42 886	38 778	56 015	51 596
9 Ciudad de México	56 884	64 907	61 622	58 650
10 Durango	33 552	33 364	41 709	37 473
11 Guanajuato	41 158	35 739	46 457	43 615
12 Guerrero	23 848	26 782	27 584	23 897
13 Hidalgo	32 263	32 792	37 707	33 228
14 Jalisco	46 034	47 369	54 917	52 351
15 México	43 505	39 867	42 114	39 042
16 Michoacán de O.	32 859	30 611	38 112	34 601
17 Morelos	37 761	35 894	43 314	40 062
18 Nayarit	34 639	37 753	44 661	40 077
19 Nuevo León	62 218	61 376	66 836	63 672
20 Oaxaca	27 158	24 629	28 715	24 815
21 Puebla	30 067	36 723	37 176	35 075
22 Querétaro	48 030	48 394	56 048	53 431

Tabla 4.1

Concluye

Ingreso corriente promedio por hogar

Entidad federativa	2012	2014	2015	2015 SVM
23 Quintana Roo	43 276	48 032	45 547	42 309
24 San Luis Potosí	34 364	33 991	40 789	36 695
25 Sinaloa	42 583	43 848	58 056	53 279
26 Sonora	53 234	51 726	58 017	53 874
27 Tabasco	37 267	35 834	42 209	38 127
28 Tamaulipas	41 851	41 170	44 147	40 391
29 Tlaxcala	31 047	30 055	36 888	34 237
30 Veracruz de I. de la Ll.	31 376	28 977	31 328	28 268
31 Yucatán	38 303	41 510	48 698	45 069
32 Zacatecas	31 578	31 061	35 749	32 175
Nacional	40 959	41 000	45 887	41 994

Cifras a precios del 2015.

Tabla 4.2

Continúa

Variaciones porcentuales con respecto al MCS inmediato anterior

Entidad federativa	2012-2014		2014-2015		2014-2015 SVM	
	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar
1 Aguascalientes	12.1%	7.7%	28.3%	25.0%	14.8%	15.6%
2 Baja California	1.4%	-0.5%	20.0%	11.4%	2.3%	3.4%
3 Baja California Sur	-2.8%	-5.3%	33.1%	24.4%	16.3%	17.7%
4 Campeche	4.3%	3.4%	16.0%	14.4%	6.7%	6.3%
5 Coahuila de Z.	11.3%	10.5%	7.0%	3.7%	-0.9%	0.0%
6 Colima	7.3%	4.2%	12.4%	6.3%	-2.1%	-2.3%
7 Chiapas	-4.0%	-7.1%	32.7%	32.3%	28.1%	14.2%
8 Chihuahua	-10.8%	-9.6%	48.6%	44.5%	30.6%	33.1%
9 Ciudad de México	12.0%	14.1%	-1.7%	-5.1%	-13.8%	-9.6%
10 Durango	3.2%	-0.6%	29.6%	25.0%	15.6%	12.3%
11 Guanajuato	-8.4%	-13.2%	31.7%	30.0%	20.9%	22.0%
12 Guerrero	10.6%	12.3%	9.4%	3.0%	1.1%	-10.8%
13 Hidalgo	7.6%	1.6%	17.1%	15.0%	8.5%	1.3%
14 Jalisco	11.8%	2.9%	15.0%	15.9%	5.2%	10.5%
15 México	-3.4%	-8.4%	10.3%	5.6%	-1.0%	-2.1%
16 Michoacán de O.	-5.2%	-6.8%	31.6%	24.5%	18.5%	13.0%
17 Morelos	-2.8%	-4.9%	26.6%	20.7%	13.8%	11.6%
18 Nayarit	15.2%	9.0%	22.2%	18.3%	9.6%	6.2%
19 Nuevo León	0.0%	-1.4%	14.2%	8.9%	-0.1%	3.7%
20 Oaxaca	-10.8%	-9.3%	24.6%	16.6%	15.3%	0.8%

Variaciones porcentuales con respecto al MCS inmediato anterior

Entidad federativa	2012-2014		2014-2015		2014-2015 SVM	
	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar	Ingreso corriente	ICT promedio por hogar
21 Puebla	22.2%	22.1%	4.9%	1.2%	-0.7%	-4.5%
22 Querétaro	2.7%	0.8%	21.9%	15.8%	7.4%	10.4%
23 Quintana Roo	18.8%	11.0%	1.6%	-5.2%	-10.5%	-11.9%
24 San Luis Potosí	0.8%	-1.1%	26.0%	20.0%	13.1%	8.0%
25 Sinaloa	5.6%	3.0%	35.9%	32.4%	21.3%	21.5%
26 Sonora	-1.7%	-2.8%	14.6%	12.2%	2.8%	4.2%
27 Tabasco	-2.0%	-3.8%	20.3%	17.8%	9.9%	6.4%
28 Tamaulipas	2.0%	-1.6%	11.0%	7.2%	-0.2%	-1.9%
29 Tlaxcala	0.0%	-3.2%	26.6%	22.7%	16.5%	13.9%
30 Veracruz de I. de la Ll.	-6.5%	-7.6%	8.6%	8.1%	6.5%	-2.4%
31 Yucatán	11.9%	8.4%	23.1%	17.3%	9.8%	8.6%
32 Zacatecas	1.8%	-1.6%	13.6%	15.1%	6.9%	3.6%
Nacional	2.6%	0.1%	15.6%	11.9%	4.5%	2.4%

Tabla 4.3

Variación porcentual 2014-2015 por fuentes de ingreso

Decil	Ingreso corriente total		Ingreso por trabajo asalariado		Ingreso por trabajo independiente		Renta de la propiedad	
	2015	2015 SVM	2015	2015 SVM	2015	2015 SVM	2015	2015 SVM
I	38.0%	11.2%	61.5%	-4.4%	54.9%	23.2%	132.1%	70.2%
II	25.2%	2.6%	31.3%	-12.9%	26.9%	8.1%	78.7%	47.5%
III	21.6%	2.6%	17.6%	-10.7%	23.0%	17.1%	79.8%	14.7%
IV	20.9%	3.7%	13.1%	-8.3%	38.5%	18.4%	103.4%	45.6%
V	19.8%	4.2%	12.8%	-6.5%	53.6%	18.0%	76.2%	57.0%
VI	19.4%	4.6%	13.1%	-3.9%	52.3%	12.5%	101.7%	57.0%
VII	18.8%	5.3%	11.8%	-1.6%	36.1%	23.2%	144.1%	112.9%
VIII	16.3%	5.1%	9.7%	-2.5%	20.5%	18.4%	111.4%	69.9%
IX	13.8%	4.8%	6.2%	-2.0%	50.3%	11.1%	102.8%	83.4%
X	11.1%	4.1%	-1.8%	-7.1%	-15.1%	17.0%	63.7%	44.8%
Nacional	15.6%	4.5%	6.4%	-5.0%	22.9%	16.4%	72.3%	50.7%

Tabla 4.4

Porcentaje de población en pobreza

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	38.1	37.8	34.8	25.9	30.4
Baja California	31.5	30.2	28.6	24.2	30.1
Baja California Sur	31.0	30.1	30.3	22.6	27.7
Campeche	50.5	44.7	43.6	34.2	38.9
Coahuila de Z.	27.8	27.9	30.2	19	22.1
Colima	34.7	34.4	34.3	26	32.4
Chiapas	78.5	74.7	76.2	65.6	70.9
Chihuahua	38.8	35.3	34.4	21.1	25.0
Ciudad de México	28.5	28.9	28.4	25.5	29.1
Durango	51.6	50.1	43.5	32.4	38.3
Guanajuato	48.5	44.5	46.6	30	34.4
Guerrero	67.6	69.7	65.2	59.3	68.1
Hidalgo	54.7	52.8	54.3	41.4	47.5
Jalisco	37.0	39.8	35.4	24.9	27.1
México	42.9	45.3	49.6	40.7	45.8
Michoacán de O.	54.7	54.4	59.2	44.4	52.2
Morelos	43.2	45.5	52.3	36.8	43.5
Nayarit	41.4	47.6	40.5	31.4	37.6
Nuevo León	21.0	23.2	20.4	17.6	21.6
Oaxaca	67.0	61.9	66.8	59.7	68.2
Puebla	61.5	64.5	64.5	50.1	54.1
Querétaro	41.4	36.9	34.2	24	28.5
Quintana Roo	34.6	38.8	35.9	30	35.8
San Luis Potosí	52.4	50.5	49.1	36.5	43.6
Sinaloa	36.7	36.3	39.4	27.5	32.7
Sonora	33.1	29.1	29.4	22.5	26.5
Tabasco	57.1	49.7	49.6	36.7	42.8
Tamaulipas	39.0	38.4	37.9	32	37.5
Tlaxcala	60.3	57.9	58.9	43.3	47.8
Veracruz de I. de la Ll.	57.6	52.6	58.0	50.7	56.8
Yucatán	48.3	48.9	45.9	34.4	40.3
Zacatecas	60.2	54.2	52.3	44.7	49.7
Nacional	46.1	45.5	46.2	36.7	42.5

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.5

Porcentaje de población en pobreza moderada

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	34.4	34.4	32.7	24.4	28.4
Baja California	28.1	27.5	25.5	22.6	28.0
Baja California Sur	26.4	26.4	26.3	20.9	25.0
Campeche	36.7	34.2	32.5	29.7	32.4
Coahuila de Z.	24.9	24.7	26.4	17.6	20.1
Colima	32.2	30.4	30.9	24.1	29.5
Chiapas	40.2	42.5	44.4	40.3	40.6
Chihuahua	32.2	31.5	28.9	20.3	23.6
Ciudad de México	26.4	26.4	26.7	24.7	27.7
Durango	41.1	42.6	38.2	30.1	34.6
Guanajuato	40.1	37.6	41.0	28.3	32.0
Guerrero	35.7	38.0	40.8	41.4	42.8
Hidalgo	41.2	42.8	42.0	34.1	37.5
Jalisco	31.8	34.0	32.1	22.5	23.9
México	34.3	39.5	42.4	36.4	39.8
Michoacán de O.	41.2	39.9	45.2	35.8	40.2
Morelos	36.3	39.1	44.4	31.4	35.7
Nayarit	33.0	35.7	32.0	27.1	30.8
Nuevo León	19.2	20.8	19.0	16.6	19.3
Oaxaca	37.7	38.6	38.4	40.7	42.7
Puebla	44.5	46.9	48.4	42.9	45.5
Querétaro	34.0	31.8	30.3	21.7	25.2
Quintana Roo	28.2	30.4	28.9	25.0	28.4
San Luis Potosí	37.1	37.7	39.6	30.5	35.0
Sinaloa	31.2	31.8	34.1	25.7	29.9
Sonora	28.0	24.2	26.1	20.4	23.4
Tabasco	43.5	35.4	38.5	31.0	34.5
Tamaulipas	33.5	33.7	33.6	28.7	32.9
Tlaxcala	50.4	48.8	52.4	40.5	43.9
Veracruz de I. de la Ll.	38.8	38.4	40.9	40.3	43.0
Yucatán	36.6	39.0	35.2	28.6	32.3
Zacatecas	49.4	46.7	46.6	40.1	44.0
Nacional	34.8	35.7	36.6	30.9	34.3

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.6

Porcentaje de población en pobreza extrema

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	3.8	3.4	2.1	1.5	2.1
Baja California	3.4	2.7	3.1	1.7	2.2
Baja California Sur	4.6	3.7	3.9	1.7	2.7
Campeche	13.8	10.4	11.1	4.6	6.5
Coahuila de Z.	2.9	3.2	3.7	1.4	2.0
Colima	2.5	4.0	3.4	1.9	2.9
Chiapas	38.3	32.2	31.8	25.3	30.3
Chihuahua	6.6	3.8	5.4	0.9	1.4
Ciudad de México	2.2	2.5	1.7	0.9	1.4
Durango	10.5	7.5	5.3	2.2	3.7
Guanajuato	8.4	6.9	5.5	1.7	2.3
Guerrero	31.8	31.7	24.5	17.9	25.3
Hidalgo	13.5	10.0	12.3	7.3	10.0
Jalisco	5.3	5.8	3.2	2.4	3.3
México	8.6	5.8	7.2	4.3	6.0
Michoacán de O.	13.5	14.4	14.0	8.6	12.0
Morelos	6.9	6.3	7.9	5.4	7.8
Nayarit	8.3	11.9	8.5	4.3	6.8
Nuevo León	1.8	2.4	1.3	1	2.3
Oaxaca	29.2	23.3	28.3	19	25.5
Puebla	17.0	17.6	16.2	7.2	8.5
Querétaro	7.4	5.2	3.9	2.4	3.3
Quintana Roo	6.4	8.4	7.0	5	7.4
San Luis Potosí	15.3	12.8	9.5	6	8.6
Sinaloa	5.5	4.5	5.3	1.7	2.8
Sonora	5.1	5.0	3.3	2.1	3.1
Tabasco	13.6	14.3	11.0	5.7	8.2
Tamaulipas	5.5	4.7	4.3	3.3	4.6
Tlaxcala	9.9	9.1	6.5	2.8	4.0
Veracruz de I. de la Ll.	18.8	14.3	17.2	10.4	13.8
Yucatán	11.7	9.8	10.7	5.8	8.0
Zacatecas	10.8	7.5	5.7	4.6	5.7
Nacional	11.3	9.8	9.5	5.8	8.2

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.7

Porcentaje de población vulnerable por carencias sociales

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	26.2	24.5	25.3	33.1	30.1
Baja California	37.9	37.6	38.3	38.8	34.2
Baja California Sur	32.5	30.0	34.4	40.1	36.6
Campeche	24.9	28.6	32.1	39.8	37.1
Coahuila de Z.	25.6	24.4	24.2	30.9	28.9
Colima	33.8	31.7	33.2	41.0	36.4
Chiapas	13.0	17.2	15.3	24.1	20.6
Chihuahua	22.8	27.4	25.8	34.7	32.7
Ciudad de México	34.4	32.4	28.0	28.9	26.8
Durango	21.1	21.8	24.3	35.0	30.9
Guanajuato	28.9	32.6	27.7	43.7	40.8
Guerrero	23.0	21.7	26.2	32.5	25.1
Hidalgo	27.5	30.7	25.8	39.3	34.4
Jalisco	33.7	28.3	29.6	38.7	37.9
México	32.2	29.5	23.7	27.9	25.0
Michoacán de O.	28.6	30.7	25.0	38.6	32.2
Morelos	33.6	32.0	25.7	38.5	33.1
Nayarit	33.4	28.2	29.4	38.8	35.0
Nuevo León	31.6	29.1	28.0	30.2	28.3
Oaxaca	22.2	26.1	23.3	28.5	22.0
Puebla	21.6	22.0	19.8	32.8	29.6
Querétaro	31.7	32.6	33.1	40.4	36.9
Quintana Roo	36.1	30.4	34.2	39.8	35.6
San Luis Potosí	20.9	24.7	24.3	34.4	30.0
Sinaloa	33.3	33.5	30.4	40.4	36.9
Sonora	31.6	36.6	32.1	36.8	34.7
Tabasco	27.2	34.0	37.5	50.5	45.3
Tamaulipas	26.9	26.8	24.4	31.7	28.4
Tlaxcala	19.3	23.9	22.6	33.7	30.0
Veracruz de I. de la Ll.	23.6	30.6	24.8	33.3	28.0
Yucatán	26.0	27.0	27.7	36.9	32.6
Zacatecas	18.4	20.4	24.4	30.5	28.0
Nacional	28.1	28.6	26.3	33.9	30.3

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.8

Porcentaje de población vulnerable por ingresos

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	8.1	10.1	8.9	7.7	8.9
Baja California	6.3	8.6	6.6	5.9	7.6
Baja California Sur	4.5	7.9	5.6	4.6	6.0
Campeche	4.3	5.6	4.0	2.9	3.0
Coahuila de Z.	12.9	12.7	11.1	10.5	11.7
Colima	4.9	6.3	6.4	5	6.7
Chiapas	2.4	1.7	2.5	1.9	1.7
Chihuahua	13.0	10.7	12.0	8.1	8.7
Ciudad de México	5.4	6.6	8.0	6.9	8.9
Durango	8.8	11.1	10.5	7.6	9.0
Guanajuato	5.7	4.9	8.6	4.7	5.1
Guerrero	2.0	2.3	2.6	1.8	2.0
Hidalgo	4.2	3.0	5.1	3.2	4.2
Jalisco	6.2	8.1	7.9	5.4	5.4
México	5.6	7.8	9.3	8.6	9.2
Michoacán de O.	4.3	3.5	4.1	3	3.9
Morelos	5.8	4.6	6.1	3.9	4.5
Nayarit	4.3	5.6	6.5	5.5	5.8
Nuevo León	8.2	8.4	9.1	8	9.2
Oaxaca	1.3	1.7	2.1	1.6	1.9
Puebla	5.6	4.2	5.1	4.3	4.5
Querétaro	5.0	6.4	7.8	4.9	6.0
Quintana Roo	4.7	6.2	6.3	4.8	5.8
San Luis Potosí	7.2	6.6	7.6	4.9	6.0
Sinaloa	7.7	6.4	6.9	4.9	5.9
Sonora	6.8	4.7	7.3	5.7	6.4
Tabasco	4.1	3.0	2.3	1.6	1.9
Tamaulipas	9.4	8.8	11.2	6.9	7.8
Tlaxcala	7.4	6.0	7.7	6.4	7.7
Veracruz de I. de la Ll.	4.5	4.0	5.0	2.9	3.1
Yucatán	6.4	6.3	7.0	4.8	5.7
Zacatecas	6.9	6.4	7.3	5	5.0
Nacional	5.9	6.2	7.1	5.4	6.1

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

Tabla 4.9

Porcentaje de población no pobre y no vulnerable

Entidad federativa	2010	2012	2014	2015	2015 SVM
Aguascalientes	27.6	27.6	31.0	33.3	30.6
Baja California	24.2	23.5	26.5	31.1	28.0
Baja California Sur	31.9	32.0	29.8	32.7	29.8
Campeche	20.3	21.2	20.4	23.1	20.9
Coahuila de Z.	33.7	34.9	34.5	39.6	37.3
Colima	26.7	27.6	26.1	27.9	24.5
Chiapas	6.1	6.4	6.0	8.4	6.8
Chihuahua	25.4	26.6	27.8	36.1	33.6
Ciudad de México	31.7	32.1	35.7	38.7	35.2
Durango	18.6	17.0	21.8	25	21.9
Guanajuato	16.9	18.1	17.2	21.6	19.7
Guerrero	7.5	6.4	6.0	6.3	4.8
Hidalgo	13.7	13.5	14.8	16.1	13.9
Jalisco	23.0	23.8	27.1	31	29.6
México	19.3	17.4	17.4	22.8	20.0
Michoacán de O.	12.3	11.5	11.7	14	11.8
Morelos	17.3	17.9	15.8	20.8	18.9
Nayarit	20.9	18.6	23.5	24.3	21.6
Nuevo León	39.1	39.2	42.5	44.2	40.9
Oaxaca	9.5	10.3	7.9	10.1	7.9
Puebla	11.3	9.4	10.5	12.8	11.8
Querétaro	21.8	24.0	24.9	30.6	28.6
Quintana Roo	24.6	24.6	23.7	25.4	22.8
San Luis Potosí	19.5	18.3	19.0	24.2	20.4
Sinaloa	22.3	23.8	23.4	27.2	24.6
Sonora	28.4	29.6	31.3	35	32.4
Tabasco	11.5	13.4	10.6	11.2	10.0
Tamaulipas	24.8	26.0	26.5	29.3	26.4
Tlaxcala	13.0	12.2	10.8	16.5	14.4
Veracruz de I. de la Ll.	14.3	12.8	12.2	13.1	12.2
Yucatán	19.2	17.9	19.5	23.9	21.5
Zacatecas	14.5	19.0	16.0	19.8	17.3
Nacional	19.9	19.8	20.5	23.9	21.2

Fuente: estimaciones del CONEVAL con base en la ENIGH-MCS 2010, 2012, 2014 y 2015.

5. Validación y evaluación de la metodología

Una de las ventajas que tiene la metodología desarrollada en el apartado 3 consiste en que puede ser fácilmente evaluada con otras bases de datos. El mecanismo de validación comprende la información de los MCS 2010 y 2012 como datos de entrenamiento; como datos de prueba se usaron los del MCS 2014.

A continuación, se presentan los resultados de validación para cuatro entidades seleccionadas de forma aleatoria. La diagonal principal indica el porcentaje de hogares para quienes la *decilización* artificial y la original coincidieron; es decir, que la clasificación artificial no cambió el decil de pertenencia del hogar. Como se observa, el porcentaje de hogares para los cuales la clasificación resultó acertada es elevado.

		Durango									
		Decil pronosticado									
		I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	86.8%	13.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	8.1%	72.5%	19.4%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	5.7%	84.4%	9.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	19.8%	76.3%	4.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.6%	22.1%	71.2%	6.1%	0.0%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	23.0%	71.7%	5.3%	0.0%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	22.4%	75.1%	2.5%	0.0%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	23.7%	74.2%	2.2%	0.0%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.7%	20.1%	77.0%	2.2%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17.1%	82.9%

		Zacatecas									
		Decil pronosticado									
		I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	91.9%	8.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	6.5%	86.2%	7.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	11.1%	76.8%	12.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	17.0%	76.0%	7.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.0%	15.5%	72.3%	12.3%	0.0%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	14.9%	81.8%	3.4%	0.0%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	32.3%	60.2%	7.5%	0.0%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	1.4%	25.7%	65.0%	7.9%	0.0%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	17.5%	80.0%	2.5%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	5.4%	94.6%

Veracruz de Ignacio de la Llave

Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	96.1%	3.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	10.9%	77.1%	11.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	14.2%	70.9%	14.9%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	17.0%	63.5%	19.5%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.0%	10.1%	70.4%	19.6%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	8.5%	69.5%	22.0%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.0%	9.3%	66.7%	24.0%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.8%	80.0%	16.2%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.1%	75.5%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	3.5%

Oaxaca

Decil pronosticado

	I	II	III	IV	V	VI	VII	VIII	IX	X
Decil original	I	95.4%	4.6%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	II	7.8%	83.3%	8.9%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	III	0.0%	8.5%	82.2%	9.4%	0.0%	0.0%	0.0%	0.0%	0.0%
	IV	0.0%	0.0%	11.7%	82.2%	6.1%	0.0%	0.0%	0.0%	0.0%
	V	0.0%	0.0%	0.8%	17.1%	75.2%	7.0%	0.0%	0.0%	0.0%
	VI	0.0%	0.0%	0.0%	0.0%	27.5%	69.6%	2.9%	0.0%	0.0%
	VII	0.0%	0.0%	0.0%	0.0%	0.9%	30.2%	66.4%	2.6%	0.0%
	VIII	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	22.4%	74.8%	2.8%
	IX	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	18.6%	73.2%
	X	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	15.9%

6. Conclusión y comentarios finales

Como consecuencia de las mejoras realizadas en el levantamiento del MCS 2015 se observa un corrimiento del ingreso reportado por los hogares, rompiendo la continuidad con levantamientos del

MCS-ENIGH anteriores; sin embargo, ello no implica que la información recabada por el INEGI sea errónea o sesgada.

Una manifestación del corrimiento del vector de ingreso se da al comparar resultados por deciles del MCS 2015 con deciles del MCS 2014: las

tasas de crecimiento del 2014 al 2015 del ingreso corriente total de cada decil fueron muy superiores a las que se venían observado.

La metodología que aquí se propone para recomponer la continuidad del MCS 2015 con los años anteriores consiste en la modificación de los factores de expansión, dejando inalterada la declaración de ingresos por parte de los hogares.

La forma de lograr continuidad consiste en modificar la representatividad que tiene un hogar de acuerdo con su asignación a un decil pronosticado o artificial, creado para ese propósito a partir de un conjunto de variables monetarias y no-monetarias. La asignación de un hogar a un decil pronosticado se realiza a través de una técnica de *Machine Learning*, conocida como máquina de soporte vectorial; esta técnica consiste en entrenar una máquina con datos de MCS anteriores al 2015 con el fin de pronosticar el decil de pertenencia de un hogar en el 2015. Para hacer comparaciones válidas con los resultados de los MCS anteriores, el supuesto en el que se basa la propuesta es que los hogares que pertenecen a determinado decil, considerando sus nuevos factores de expansión, representan a aquellos hogares que bajo circunstancias similares a las históricas hubiesen pertenecido a ese decil.

Una ventaja de este método es que utiliza información histórica para hacer pronósticos en el 2015 sin modificar la información reportada por los hogares, además de que trata de minimizar las clasificaciones erróneas y sus resultados son fácilmente validables y comparables. Sin embargo, debe con-

siderarse que al ser una técnica de optimización, los resultados dependerán de las covariables incluidas y de los parámetros de ajuste especificados dentro del paquete computacional, aunados a las posibles consecuencias en el diseño muestral a causa de la modificación de los factores de expansión.

En términos de resultados, la variación porcentual entre el 2014 y 2015 del ICT a nivel nacional pasó de 15.6 a 4.5% con la presente propuesta de ajuste. Respecto a la variación porcentual en el decil I, ésta pasó de 38 a 11.2%; mientras que en el decil X fue de 11.1 a 4.1 por ciento. En términos de medición de pobreza, el porcentaje de población en pobreza según el MCS 2015 fue de 36.7, en tanto que con el ajuste aumentó a 42.5 por ciento. La pobreza extrema creció de 5.8 a 8.2%, mientras que la moderada fue de 30.9 a 34.3 por ciento.

Fuentes

- Cortes, Corina y Vladimir Vapnik. "Support vector networks", en: *Machine Learning*. Boston, United States of America, Kluwer Academic Publishers, 1995.
- Hastie, Trevor; Robert Tibshirani y Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York, United States of America, Springer, 2009.
- Schölkopf, Bernhard y Alexander J. Smola. *Learning with kernels: Support vector machines, regularization, optimization, and beyond*. Cambridge, England, MA, MIT Press, 2002.
- Witten, Ian y Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco, CA, United States of America, Morgan Kaufmann, 2011.

Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
Población	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	numren	Número consecutivo en el registro de personas integrantes del hogar.	Ninguno.
	edad	Años transcurridos entre la fecha de nacimiento de la persona y la fecha de la entrevista.	Promedio de edad de los integrantes del hogar.
	hablaind	Personas de 3 años o más que hablan alguna lengua indígena o dialecto.	Proporción de integrantes del hogar que habla alguna lengua indígena.
	alfabetism	Situación que distingue a las personas de 3 años o más según sepan leer y escribir un recado.	Proporción de integrantes del hogar mayores de 12 años que no sabe escribir/ leer un recado.
	nivel	Nivel al que asiste a la escuela el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional.	Proporción de integrantes del hogar que asiste a cada uno de los niveles escolares.
	tipoes	Tipo de escuela a la que asiste el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional, según sea pública, privada o de otro tipo.	Proporción de integrantes del hogar que asiste a cada uno de los tipos de escuela.
	tiene_b	Apoyo en dinero o especie que se proporciona a los integrantes del hogar que asisten a la escuela.	Proporción de integrantes del hogar que tiene beca.
	tiene_c	Aportación que recibe el integrante del hogar que asiste a la escuela para el pago de colegiatura, hospedaje, alimentación, materiales, etcétera. Quien lo recibe está comprometido a pagarlo en parte o en su totalidad al concluir sus estudios.	Proporción de integrantes del hogar que tiene crédito educativo.
nivelaprob	Año máximo aprobado en la escuela por el integrante del hogar de 3 años o más dentro del Sistema Educativo Nacional.	Proporción de integrantes del hogar para cada nivel de educación.	
trabajo_mp	Es la situación que distingue a la población de 12 años o más de acuerdo con el desempeño o no de una actividad económica en el periodo de referencia, que permite clasificarla como parte de la Población Económicamente Activa (PEA) o como parte de la Población No Económicamente Activa (PNEA).	Proporción de integrantes del hogar que trabaja.	
act_pnea1	Actividad o situación de las personas no económicamente activas.	a) Proporción de integrantes que se dedicó a buscar trabajo.	
act_pnea1	Actividad o situación de las personas no económicamente activas.	b) Proporción de integrantes que se dedicó a quehaceres del hogar.	
Vivienda	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.

Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	tipo_viv	Clase de vivienda particular.	Generación de una variable dicotómica por cada una de las categorías.
	mat_pared	Material predominante en las paredes de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	mat_techos	Material predominante en el techo de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	cocina	Cuarto de la vivienda dispuesto para preparar o calentar alimentos, aunque también tenga otros usos.	Generación de una variable dicotómica por cada una de las categorías.
	cocina_dor	Cocina utilizada de manera habitual también para dormir.	Generación de una variable dicotómica por cada una de las categorías.
	cuart_dorm	Número de cuartos de la vivienda que son usados habitualmente para dormir, aunque también tengan otros usos.	Ninguno.
	num_cuarto	Número total de cuartos que tiene la vivienda, independientemente de su uso.	Ninguno.
	excusado	Instalación sanitaria o excusado en la vivienda para el desalojo de los desechos humanos.	Generación de una variable dicotómica por cada una de las categorías.
	uso_compar	Uso compartido de la instalación sanitaria con los habitantes de otra vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	sanit_agua	Funcionamiento de la instalación sanitaria con o sin conexión de agua.	Generación de una variable dicotómica por cada una de las categorías.
	drenaje	Disponibilidad de un sistema de drenaje para desalojar de la vivienda los desechos humanos y las aguas utilizadas.	Generación de una variable dicotómica por cada una de las categorías.
	disp_elect	Fuente de donde se obtiene la energía eléctrica en la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	combustible	Combustible más utilizado para preparar o calentar alimentos en la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	eli_basura	Formas de eliminar la basura de la vivienda.	Generación de una variable dicotómica por cada una de las categorías.
	tenencia	Dominio o posesión legal de la vivienda por sus ocupantes.	Generación de una variable dicotómica por cada una de las categorías.
	lavadero	Lavadero.	Generación de una variable dicotómica por cada una de las categorías.
	fregadero	Fregadero o tarja.	Generación de una variable dicotómica por cada una de las categorías.
	regadera	Regadera.	Generación de una variable dicotómica por cada una de las categorías.

Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	tinaco_azo	Tinaco en la azotea.	Generación de una variable dicotómica por cada una de las categorías.
	cisterna	Cisterna o aljibe.	Generación de una variable dicotómica por cada una de las categorías.
	pileta	Pileta, tanque o depósito de agua.	Generación de una variable dicotómica por cada una de las categorías.
	calent_gas	Calentador de gas u otros.	Generación de una variable dicotómica por cada una de las categorías. Se renombró como calentador.
	medidor_luz	Medidor de luz.	Generación de una variable dicotómica por cada una de las categorías.
	bomba_agua	Bomba de agua.	Generación de una variable dicotómica por cada una de las categorías.
	tanque_gas	Tanque de gas estacionario.	Generación de una variable dicotómica por cada una de las categorías.
	aire_acond	Aire acondicionado.	Generación de una variable dicotómica por cada una de las categorías.
	calefacc	Calefacción.	Generación de una variable dicotómica por cada una de las categorías.
	ubica_geo	Contiene la ubicación geográfica de la vivienda. Los dos primeros dígitos representan la clave de la entidad y los tres últimos dígitos la clave del municipio. Éstos corresponden al <i>Catálogo único de claves de áreas geoestadísticas estatales, municipales y localidades</i> , que está disponible en el sitio del INEGI.	Se extrajo la entidad y se formó una nueva variable llamada edo.
Hogar	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	telefono	Servicio de línea telefónica en el hogar ya sea con instalación alámbrica o satelital (antena).	Generación de una variable dicotómica por cada una de las categorías.
	celular	Hogares que cuentan con el servicio de teléfono móvil aunque sea un solo integrante del hogar.	Generación de una variable dicotómica por cada una de las categorías.
	tv_paga	Hogares que cuentan con los servicios de televisión por pago, considerando que el servicio está en funcionamiento.	Generación de una variable dicotómica por cada una de las categorías.
	conex_inte	Hogares que cuentan con el servicio de línea telefónica, tarjeta prepagada o telecable.	Generación de una variable dicotómica por cada una de las categorías.
	num_auto	Número de automóviles con los que cuenta el hogar.	Ninguno.

Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	anio_auto	Año de adquisición del último automóvil.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_van	Número de camionetas cerradas o con cabina (<i>van</i> o <i>minivan</i>) con las que cuenta el hogar.	Ninguno.
	anio_van	Año de adquisición de la última camioneta cerrada o con cabina.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_pickup	Número de camionetas con caja con las que cuenta el hogar.	Ninguno.
	anio_pickup	Año de adquisición de la última camioneta de caja.	Se generó una nueva variable con categorías de acuerdo con el rango de antigüedad del auto: 1. El vehículo tiene a lo más un año de antigüedad. 2. El vehículo tiene entre 2 y 5 años de antigüedad. 3. El vehículo tiene entre 6 y 10 años de antigüedad. 4. El vehículo tiene más de 10 años de antigüedad. 5. Tienen vehículo pero no respondieron sobre la antigüedad. 6. No tienen vehículo.
	num_ester	Número de estéreos, modulares o minicomponentes con los que cuenta el hogar.	Ninguno
	num_grab	Número de grabadoras con las que cuenta el hogar.	Ninguno.

Variables utilizadas, 2015

Base de datos	Nombre de la variable	Definición	Tratamiento
	num_radio	Número de radios con las que cuenta el hogar.	Ninguno.
	num_tva	Número de televisores analógicos con los que cuenta el hogar.	Su sumó con la variable num_tvd y se formó una nueva variable llamada num_tv.
	num_tvd	Número de televisores digitales con los que cuenta el hogar.	Su sumó con la variable num_tva y se formó una nueva variable llamada num_tv.
	num_dvd	Número de DVD, <i>Blue-ray</i> (reproductores de discos de video) con los que cuenta el hogar.	Ninguno.
	num_video	Número de videocaseteras con las que cuenta el hogar.	Ninguno.
	num_licua	Número de licuadoras con las que cuenta el hogar.	Ninguno.
	num_tosta	Número de tostadores con los que cuenta el hogar.	Ninguno.
	num_micro	Número de hornos de microondas con los que cuenta el hogar.	Ninguno.
	num_refri	Número de refrigeradores con los que cuenta el hogar.	Ninguno.
	num_estuf	Número de estufas de gas o eléctricas con las que cuenta el hogar.	Ninguno.
	num_lavad	Número de lavadoras con las que cuenta el hogar.	Ninguno.
	num_planc	Número de planchas eléctricas con las que cuenta el hogar.	Ninguno.
	num_maqui	Número de máquinas de coser con las que cuenta el hogar.	Ninguno.
	num_venti	Número de ventiladores con los que cuenta el hogar.	Ninguno.
	num_aspir	Número de aspiradoras con las que cuenta el hogar.	Ninguno.
	num_compu	Número de computadoras con las que cuenta el hogar.	Ninguno.
	num_impre	Número de impresoras con las que cuenta el hogar.	Ninguno.
	num_juego	Número de videojuegos (<i>Wii, Playstation, Xbox</i> , otros) con los que cuenta el hogar.	Ninguno.
Concentrado	folioviv	Identificador de la vivienda compuesto por dos dígitos con la clave de la entidad federativa, uno con el ámbito (urbano, código diferente a 6; rural, código 6), cuatro dígitos del número consecutivo de la upm, un dígito de la decena de levantamiento y dos dígitos con un número consecutivo para la vivienda seleccionada.	Ninguno.
	foliohog	El código 1 identifica al hogar principal y del 2 al 5, los hogares adicionales.	Ninguno.
	ing_cor	Suma de los ingresos por trabajo, los provenientes de rentas, de transferencias, de estimación del alquiler y de otros ingresos.	Ninguno.

Fuente: INEGI. *Módulo de Condiciones Socioeconómicas 2015*. Descripción de la base de datos.

Ajuste de una base de datos vía *modelos lineales*

Setting up a database via *linear models*

Víctor Alfredo Bustos y de la Tijera*

En vista de que, a diferencia del ingreso de los hogares y de algunos elementos de carácter demográfico, las variables no monetarias recabadas por el Módulo de Condiciones Socioeconómicas (MCS) 2015 no muestran cambios importantes, y de su asociación con el ingreso permanente, se decidió explorar sus relaciones para aproximar el comportamiento de los ingresos en el 2015 desde la perspectiva de los levantamientos anteriores, una vez corregida la demografía de los hogares mediante posestratificación de acuerdo con estructuras de la Encuesta Intercensal 2015. En consecuencia, el propósito del presente trabajo es el de explorar y, en su caso, explorar la relación histórica entre variables no monetarias y el ingreso corriente total (ICT), recogidos en uno o más de los levantamientos anteriores del MCS, para aproximar el comportamiento del ICT para cada hogar en el 2015. Para permitir mayor flexibilidad en los modelos que aproximan tal relación, se modeló por separado el comportamiento del ICT al interior de cada uno de los cuatro estratos socioeconómicos en los que han sido clasificadas las unidades primarias de muestreo del Marco Nacional de Viviendas. En cada caso, se ignoraron observaciones con valores atípicos (residuos más allá de tres desvia-

Given that, unlike household income and some demographic elements, the non-monetary variables collected by the MCS 2015 do not show significant changes, and their association with permanent income, it was decided to explore their relationships to approximate the income behavior in 2015 from the perspective of the previous surveys, and once the demography of the households has been corrected through post-stratification in accordance with the structures of the 2015 Intercensus Survey. Consequently, the purpose of this paper is to explore and, if appropriate, exploit the historical relationship between non-monetary variables and the current quarterly income (ICT), collected in one or more of the previous surveys of the Socioeconomic Conditions Module, to approximate the ICT behavior for each household, in the year 2015. To allow for greater flexibility when modeling their relationship, the behavior of ICT within each of the four socioeconomic strata in which the primary sampling units (PSUs) of the National Housing Framework have been classified, were separately modeled. In each case, observations with atypical values (residues beyond 3 s.d. of their mean) were ignored. The database was also post-stratified to

* Instituto Nacional de Estadística y Geografía, alfredo.bustos@inegi.org.mx



Crowd forming a positive graph/Henrik Sorensens/Getty Images

ciones estándar de su promedio). La base de datos fue, asimismo, posestratificada para corregir la demografía de los hogares. A pesar de lo anterior, los ingresos estimados por esta vía aún no son comparables con los ingresos del 2014, pues el modelo ignora las especificidades del momento del levantamiento (p. ej., la desocupación del único receptor de ingresos en el hogar da lugar a un desequilibrio entre el ingreso declarado y el equipamiento de la vivienda). En otras palabras, los ingresos estimados a partir del modelo no toman en cuenta la discrepancia entre la distribución de los errores supuesta en el momento del ajuste y lo que la realidad refleja. Es por ello que en este trabajo se exploran dos formas de, por así decirlo, *emparejar el piso* (además de la posestratificación de las bases de datos de diversos años) con el fin de llevar a cabo las comparaciones requeridas. En primera instancia, los ingresos del 2014 son reemplazados por estimaciones obtenidas a partir del mismo modelo, con lo que los efectos coyunturales son removidos. En el segundo caso, se generan valores aleatorios de dichos efectos coyunturales a partir de la distribución empírica de los residuos del ajuste en el 2014, mismos que son añadidos a las estimaciones obtenidas para el 2015 a partir del modelo.

Palabras clave: modelos lineales; distribución empírica; simulación.

correct household demographics. In spite of the above, the income estimated in this way is not yet comparable with the income of 2014 because the model ignores the specificities of the time of the survey (eg. the unemployment of the only recipient of income in the household leads to an imbalance between declared income and housing equipment). In other words, the estimated income from the model does not take into account the discrepancy between the assumed distribution of errors at the time of adjustment and what the reality reflects. For this reason, this paper explores two ways of “evening the floor” (as well as the post-stratification of databases of different years) in order to carry out the required comparisons. The revenues of 2014 are replaced by estimates obtained from the same model, with the result that the cyclical effects are removed. In the second case, random values of these conjunctural effects are simulated from the empirical distribution of the residuals of the adjustment in 2014, which are added to the estimates obtained for 2015 from the model.

Key words: linear models; empirical distribution; simulation.

1. Introducción

La petición por parte de los responsables de dar seguimiento al levantamiento de información a través del Módulo de Condiciones Socioeconómicas (MCS) 2015, en el sentido de que tanto los entrevistadores como sus supervisores insistieran con los informantes acerca de la necesidad de declarar de manera fidedigna sus ingresos, parece haber dado lugar a diversos cuestionamientos planteados por los usuarios de la información sobre los posibles cambios metodológicos introducidos en el ejercicio más reciente. Parece claro que una de las consecuencias más importantes de este proceder fue la reducción del número de ingresos subreportados por los hogares. Este solo hecho resultó en un incremento del ingreso promedio del primer decil, donde se hubiera ubicado a los hogares a partir de su declaración original de ingresos. Más aún, es posible que el nuevo valor declarado llevara a algunos de los hogares a ser ubicados en deciles superiores. Con ello, se haría necesario reconstituir al primero trayendo hogares cuyos ingresos los hubieran ubicado en el segundo decil. Nuevamente, la inclusión de ingresos superiores a los del primer decil llevaría a su ingreso promedio a mostrar un valor mayor y, en consecuencia, a mostrar una tasa de crecimiento *sorprendente* cuando se le comparase con el mismo decil del 2014. A su vez, fue necesario reconstituir el segundo decil con hogares del tercero y así sucesivamente, según fuese necesario, por lo que sus crecimientos con respecto al 2014 ya no fueron tan altos. Se desprende de lo anterior que la insistencia en precisar la declaración de ingreso no se concentró en los hogares *pobres* sino en aquellos cuyas condiciones y equipamiento de sus viviendas llevaban a pensar en una subdeclaración del ingreso.

Es en este contexto que surge el presente ejercicio estadístico y econométrico. Bajo el supuesto de que los ejercicios anteriores se levantaron bajo circunstancias similares, surge la pregunta acerca de cuáles hubieran sido los resultados en el 2015 de no haberse dado ninguna modificación en las instrucciones al personal de campo. Es decir, si tales instrucciones modificaron o no en alguna medida la relación estructural entre las variables no mone-

tarias y las correspondientes variables monetarias referidas al ingreso, recogidas por anteriores levantamientos del MCS o si ellas afectaron en alguna medida a la distribución de los errores. Parece pensarse que de no haber sido éste el caso, no solamente los agregados monetarios derivados de la información recopilada sino la relación entre variables monetarias y no monetarias habrían operado también a lo largo del nuevo levantamiento.

Por todo lo anterior, se decidió la conveniencia de explorar, en una primera instancia, la posible estabilidad de la relación entre las mencionadas variables y, en su caso, aprovecharla para aproximar el comportamiento de la variable ingreso corriente total (ICT) de los hogares en el 2015, lo que permitiría contar con una base de datos comparable con los ejercicios anteriores. Al resultado de éste, al igual que al de otros ejercicios analíticos similares, se le ha denominado *base de datos ajustada vía modelos lineales* del MCS 2015.

Queda claro que la sola determinación y uso de la estructura común, en su caso, deja fuera la posible existencia de otros factores exógenos al propio levantamiento, y aún al de algunos otros endógenos al mismo; ejemplo de esto último lo da la estructura de valores de ingresos pequeños, la cual, como ya se ha indicado, marca una de las mayores diferencias entre el levantamiento del 2015 y los anteriores. Más aún, como se verá, esa misma estructura sufrirá una distorsión mayor al imputar los ingresos del 2015 mediante el uso de modelos que aproximen la relación estructural ya citada.

En las siguientes secciones de este documento se presentan los fundamentos teóricos detrás de los modelos desarrollados. En el tercer apartado se discute la metodología utilizada para lograr la estimación de los ingresos corrientes totales de los hogares con base en las relaciones estructurales analizadas. Por su parte, la cuarta sección muestra los resultados obtenidos, así como una discusión de sus bondades y deficiencias. En la quinta se discute la forma en que los resultados obtenidos pueden ser evaluados y, finalmente, la última pretende alcanzar algunas conclusiones.

2. Marco teórico

2.1 Relación ingreso corriente y acervos acumulados por los hogares

La metodología a seguir tiene como fundamento la relación observada entre las características de la vivienda que ocupan los hogares bajo estudio (entre las que se cuentan los materiales con que han sido construidas y los servicios de los que éstas disponen) con el ingreso que los mismos hogares perciben. Con el propósito de mejorar la aproximación al ingreso observado, es posible incluir, además, la relación entre el propio ingreso y los bienes y acervos, tanto físicos como humanos, de los que disponen en el hogar. Para el ejemplo que nos ocupa, dicha relación es aproximada mediante modelos lineales. Por su propia naturaleza, algunas de las mencionadas características de la vivienda y del hogar tienen un carácter cualitativo por lo que el modelo habrá de incluir un coeficiente para cada uno de los niveles que dichas variables presentan en la población. Por su parte, cada una de las variables cuantitativas incluidas en el modelo habrá de tener asociado un coeficiente.

Para el ajuste óptimo de tales modelos mediante el método de mínimos cuadrados, es usual solicitar que los errores estadísticos sean homoscedásticos y no correlacionados. Sin embargo, en el contexto actual, los datos son obtenidos mediante muestreo estadístico de poblaciones finitas con un diseño complejo. Para el caso que nos ocupa, la selección de las unidades a las que se aplicará el cuestionario del MCS se lleva a cabo en dos etapas. En la primera de ellas, las unidades primarias de muestreo (UPM) son seleccionadas con probabilidades proporcionales al tamaño, de entre las que se encuentran en el mismo estrato socioeconómico de los cuatro en que dichas unidades primarias han sido previamente clasificadas. En la segunda etapa se elige un número determinado de las viviendas contenidas en las unidades primarias con igual probabilidad. Lo anterior dificulta que los requerimientos del método de mínimos cuadrados sean satisfechos. Por ello, se hace uso de rutinas que reconocen el diseño muestral y lo incorporan

al ajuste de modelos. En nuestro caso, la paquetería estadística utilizada para llevar a cabo dichos ajustes es la conocida con el nombre de *Stata*, que cuenta con algunas rutinas a las que es posible anteponer el prefijo SVY con el fin de lograr una estimación más precisa de los segundos momentos obtenidos durante el proceso de estimación.

Los apéndices de resultados¹ muestran paso a paso el procedimiento seguido para adecuar cada una de las bases de datos, para declarar el diseño muestral de cada una de ellas y su uso en la obtención de modelos ajustados a los datos correspondientes a los levantamientos 2008, 2010, 2012, 2014 y 2015. En particular, los resúmenes aportados por el paquete, y cuyo propósito es el de evaluar la adecuación del modelo ajustado para explicar el comportamiento del logaritmo natural del ingreso corriente total de los hogares, parecen aportar evidencia que soporta la supuesta estabilidad estructural en la relación entre el ingreso corriente total de los hogares y las variables no monetarias seleccionadas para ser incluidas en los modelos. Como se observará a partir de la tabla 1, los modelos ajustados aportan una explicación significativa del comportamiento de los ingresos, así como de una calidad similar de acuerdo con uno de los criterios más utilizados para establecer la bondad del ajuste alcanzado conocido como la R^2 que para los modelos ajustados en cada uno de los años toma un valor superior a 0.6. Por su parte, la tabla 2 proporciona un listado de las variables consideradas en cada uno de los ajustes, así como sus mnemónicos; adicionalmente, variables *dummy* asociadas con cada una de las entidades federativas fueron incluidas.

Es preciso destacar que la aparente estabilidad se da a pesar del cambiante entorno económico en que tuvo lugar cada uno de los levantamientos. Como se recordará, el 2008 se tiene como el inicio de una de las peores crisis económicas del planeta. México, por supuesto, no fue inmune a los efectos

¹ Revisar en la versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>.

Tabla 1

Estadísticas resumen del ajuste de modelos lineales a cada uno de los levantamientos del MCS para los años que se indica

2008:			
Number of strata	= 4	Number of obs.	= 53 203
Number of PSUs	= 8 954	Population size	= 24 092 985
		Design df	= 8 950
		F(142, 8 809)	= 303.43
		Prob > F	= 0.0000
		R-squared	= 0.6449
2010:			
Number of strata	= 4	Number of obs.	= 46 600
Number of PSUs	= 8 420	Population size	= 21 974 408
		Design df	= 8 416
		F(142, 8 275)	= 236.93
		Prob > F	= 0.0000
		R-squared	= 0.6611
2012:			
Number of strata	= 4	Number of obs.	= 43 053
Number of PSUs	= 7 079	Population size	= 23 231 938
		Design df	= 7 075
		F(142, 6 934)	= 196.03
		Prob > F	= 0.0000
		R-squared	= 0.6417
2014:			
Number of strata	= 4	Number of obs.	= 57 624
Number of PSUs	= 8 110	Population size	= 31 895 373
		Design df	= 8 106
		F(142, 7 966)	= 296.84
		Prob > F	= 0.0000
		R-squared	= 0.6260
2015:			
Number of strata	= 4	Number of obs.	= 43 492
Number of PSUs	= 7 882	Population size	= 25 105 148
		Design df	= 7 878
		F(142, 7 737)	= 253.49
		Prob > F	= 0.0000
		R-squared	= 0.6297

Fuente: cálculos propios a partir de las bases de datos de los levantamientos del MCS para los años 2008, 2010, 2012, 2014 y 2015.

Lista de las variables explicativas consideradas en los modelos ajustados para aproximar el comportamiento del logaritmo del ICT de los hogares, y sus mnemónicos

Descripción	mnemónico
Disposición de aire acondicionado	aire_acond
Frecuencia con la que llega el agua a la vivienda	dotac_agua
Material predominante en sus paredes	mat_pared
Número de <i>dvd</i> del hogar	num_dvd
Número de microondas del hogar	num_micro
Número de ventiladores en el hogar	num_venti
Tiempo que tiene la vivienda de haber sido construida	antiguedad
Disponibilidad de un sistema de drenaje en el hogar	drenaje
Material predominante en sus pisos	mat_pisos
Número de <i>estéreos</i> en el hogar	num_ester
Número de motocicletas en el hogar	num_moto
Número de videocaseteras en el hogar	num_video
Disposición de bomba de agua	bomba_agua
Educación formal del jefe del hogar	educa_jefe
Material predominante en el techo de la vivienda	mat_techos
Número de estufas en el hogar	num_estuf
Número de vehículos en el hogar	num_otro
Número de personas que perciben ingreso corriente monetario	perc_ocupa
Disposición de calefacción	calefacc
Formas de eliminar la basura de la vivienda	eli_basura
Disposición de medidor de luz	medidor_luz
Número de focos en el hogar	num_focos
Número de camionetas con caja con las que se cuenta en el hogar	num_pickup
Número de personas que perciben ingreso corriente monetario	percep_ing
Disposición de calentador solar o de gas	calentador
Disposición de excusado, sanitario, letrina u hoyo	excusado
Número de aspiradoras del hogar	num_aspir
Número de radiograbadoras del hogar	num_grab
Número de planchas en el hogar	num_planc
Disposición de pileta o tanque	pileta
Disposición de cisterna o aljibe en el hogar	cisterna
Número de automóviles en el hogar	num_auto
Número de impresoras en el hogar	num_impre
Número de radios en el hogar	num_radio
Disposición de regaderas en el hogar	regadera
Disposición de cocina en el hogar	cocina
Disposición de fregadero o tarja	fregadero
Número de bicicletas en el hogar	num_bici

Tabla 2

Concluye

Lista de las variables explicativas consideradas en los modelos ajustados para aproximar el comportamiento del logaritmo del ICT de los hogares, y sus mnemónicos

Descripción	mnemónico
Número de videojuegos en el hogar	num_juego
Número de refrigeradores en el hogar	num_refri
Sexo del jefe del hogar	sexo_jefe
Combustible más usado para preparar o calentar alimentos	combustible
Número de hombres en el hogar	hombres
Número de canoas, lanchas u otros en el hogar	num_canoa
Número de lavadoras en el hogar	num_lavad
Número de tostadores en el hogar	num_tosta
Disposición de tanque de gas estacionario	tanque_gas
Número de cuartos en la vivienda usados para dormir	cuart_dorm
Número de carretas en el hogar	num_carret
Número de licuadoras en el hogar	num_licua
Número de triciclos en el hogar	num_trici
Dominio o posesión legal de la vivienda por sus ocupantes	tenencia
Fuente de donde se obtiene la energía eléctrica en la vivienda	disp_elect
Disposición de lavadero estacionario	lavadero
Número de computadoras en el hogar	num_compu
Número de máquinas de coser en el hogar	num_maqui
Número de televisores en el hogar	num_tv
Disposición de tinaco en la azotea	tinaco_azo
Número de camionetas en el hogar	num_van
Número de integrantes en el hogar	tot_integ

de la crisis, lo que se reflejó en el comportamiento de los ingresos percibidos por los hogares. Para el 2010 se tiene, tal vez, la evidencia más palpable de los efectos de la mencionada crisis. A partir de ese momento es perceptible una tendencia ligera hacia la recuperación del entorno económico y del ingreso.

Vale la pena observar que la población de hogares a la que cada modelo expande no coincide con la reportada en cada ejercicio del MCS. Lo anterior se debe a que el ajuste no considera las observaciones para las cuales los valores de una o más de las variables contempladas en el mode-

lo no fueron recogidas durante el levantamiento. Consecuentemente, todos los hogares cuyo ingreso fue reportado con un valor 0 y, en consecuencia, cuyo logaritmo natural no puede ser calculado son excluidos del ajuste. Del mismo modo, cuando el valor de una variable cuantitativa, como el número de focos, no fue registrado en el cuestionario el hogar es excluido del ajuste. Algo similar ocurre con las variables cualitativas.

En adición a la estimación del ingreso corriente de los hogares para el 2015 a partir de los modelos ajustados a los datos de los levantamientos anteriores, la base de datos correspondiente a este

año fue posestratificada con el propósito de corregir también la demografía de los hogares. Para este fin, se utilizó información recabada durante la Encuesta Intercensal (EIC) 2015, en particular de la tabla *Hogares* 3² de la sección de hogares que resume la estructura de los hogares de México según la edad y el sexo del jefe del hogar, así como del número de integrantes de éstos.

Las figuras 1, 2 y 3 muestran el efecto de esta corrección en la estructura relativa de los hogares por edad y sexo del jefe, en la que se refiere al tamaño del hogar, así como las de otras variables demográficas, respectivamente. La figura 1 presenta tres barras para cada grupo de edad y cada sexo del jefe del hogar. La primera de ellas muestra la proporción que ese grupo de hogares representa del total según el MCS 2015; la segunda, la misma cantidad según la EIC 2015; y la tercera, la que corresponde al MCS 2015 una vez posestratificado. Es claro que en la nueva base de datos la sub y sobrerepresentación de cada grupo en el MCS 2015 ha sido corregida y llevada a los niveles correspondientes a la EIC 2015.

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

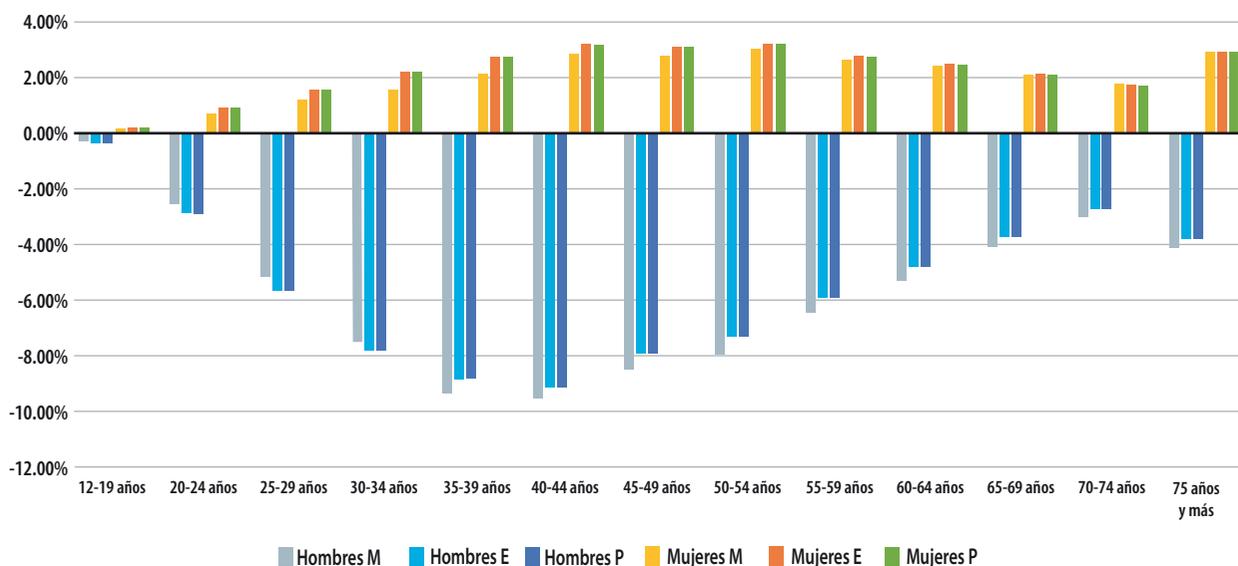
La figura 2 hace evidente la sobrerepresentación en la muestra de los hogares pequeños a costa de los grandes. Nuevamente, la posestratificación corrige las desviaciones en uno u otro sentido apeándose a lo estipulado para la Encuesta Intercensal 2015.

La figura 3 exhibe la manera en que la posestratificación da lugar a un mayor promedio del número de integrantes del hogar que el que dio como resultado la base de datos original del MCS 2015. Asimismo, muestra las correcciones a las que da lugar la posestratificación.

2.2 Introducción o no de variables a nivel estatal

Como se indicó con anterioridad, además de las anteriores variables, todos los modelos ajustados incluyeron una variable *dummy* para cada una de las entidades federativas y, en su caso, otra de estas variables para cada uno de los estratos sociodemográficos en los que el Marco Nacional de Viviendas (MNV) clasificó a cada una de las unidades primarias de muestreo. La presencia del primer conjunto

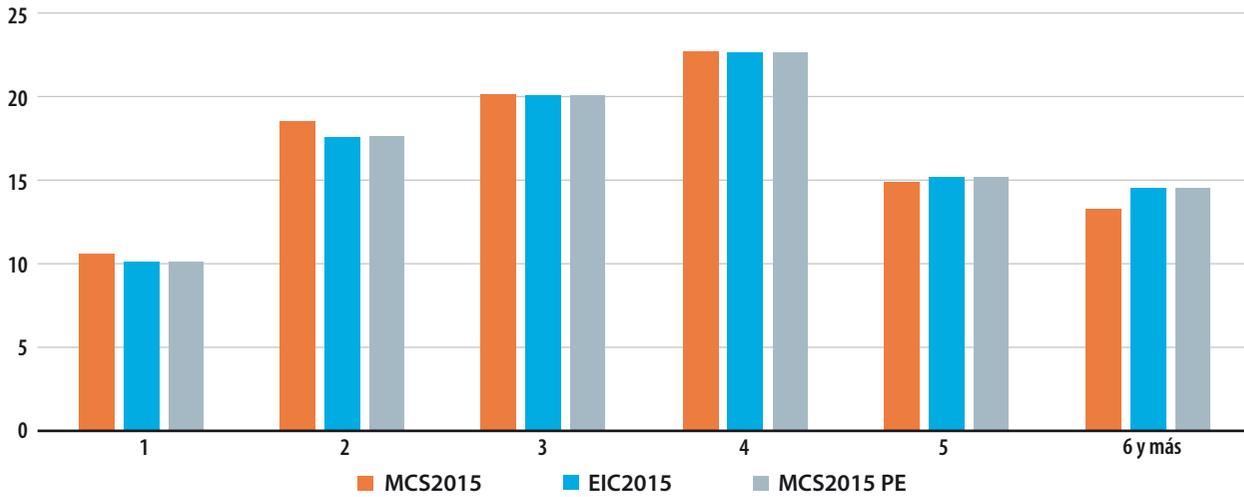
Figura 1
Efecto de la posestratificación en las estructuras relativas por edad y sexo del jefe del hogar, 2015



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

Figura 2

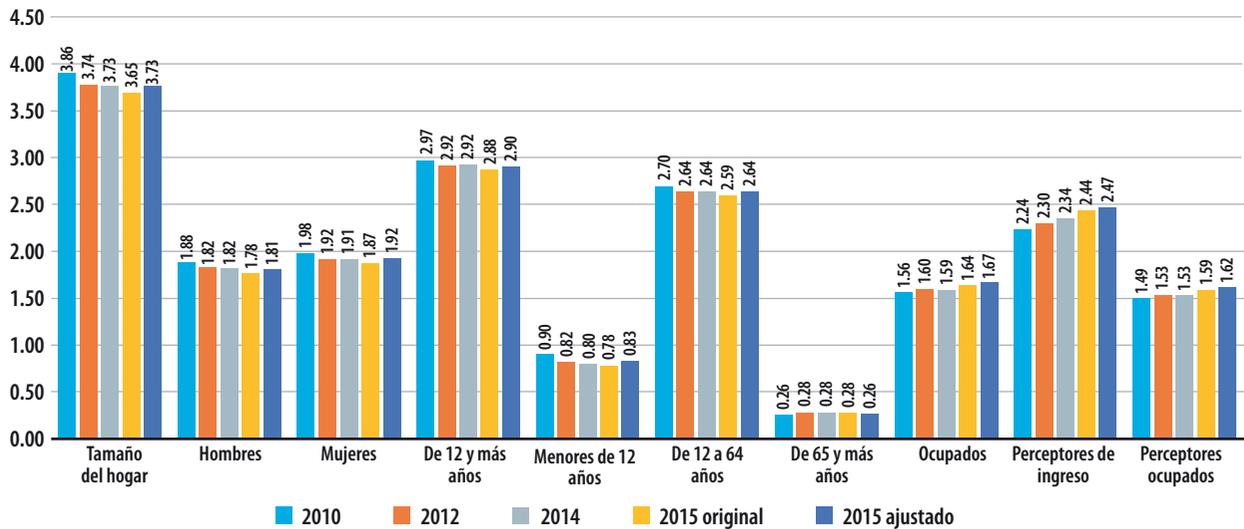
Efecto de la posestratificación en los tamaños del hogar, 2015



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

Figura 3

Efecto de la posestratificación en otras variables



Fuente: cálculos propios a partir de la base de datos del MCS 2015.

recoge todas las diferencias presentadas entre las entidades. Es decir, se hace innecesario incluir nuevas variables numéricas, como el producto interno bruto de la entidad. Es fácil demostrar la veracidad de la anterior aseveración. Para ello, se requiere establecer alguna anotación.

Parámetros del modelo: β_0

Parámetros del modelo: $\beta_0 \beta_1 \dots \beta_{32} \alpha_1 \dots \alpha_q$

Columnas por variable: $X = [1 \ e_1 \dots e_{32} \ x_1 \dots x_q]$

La matriz diseño X tiene tantos renglones como observaciones hay en la muestra y tantas columnas como el número de variables incluidas en el modelo + 33. La columna 1 tiene todas sus entradas iguales a 1 pues está asociada con el término constante. Las columnas $e_j, j=1, \dots, 32$ son tales que sus componentes son iguales a 1 si el renglón corresponde a una observación de la entidad a la que están asociadas y 0, en otro caso. Ya que ninguna observación pertenece a más de una entidad, debe tenerse que $\mathbf{1} = e_1 + \dots + e_{32}$; o que, despejando, $e_1 = \mathbf{1} - (e_2 + \dots + e_{32})$. En otras palabras, las primeras 33 columnas de la matriz diseño exhiben una redundancia tal que una de ellas puede ser expresada en función de las restantes 32. Lo anterior introduce el problema aritmético conocido como colinealidad. En consecuencia, con el propósito de obtener una única solución de las ecuaciones normales se acostumbra eliminar una columna de este conjunto, lo que equivale a reparametrizar el modelo en términos de funciones lineales de los parámetros originales, para las cuales sí es posible obtener un único estimador. De este modo, cuando se elige a la primera entidad como entidad de referencia, eliminando la columna que se le asociaba se tiene que:

$$X = \begin{bmatrix} \beta_0 + \beta_1 & \beta_2 - \beta_1 & \beta_{32} - \beta_1 & \alpha_1 \dots \alpha_q \\ \mathbf{1} & e_2 & \dots & e_{32} & x_1 & \dots & x_q \end{bmatrix}$$

En cambio, cuando el vector eliminado es el asociado al término constante se obtiene la siguiente reparametrización:

$$X = \begin{bmatrix} \beta_0 + \beta_1 & \beta_0 + \beta_2 & \beta_0 + \beta_{32} & \alpha_1 \dots \alpha_q \\ e_1 & e_2 & \dots & e_{32} & x_1 & \dots & x_q \end{bmatrix}$$

La adición de variables a nivel estatal no mejorará el ajuste del modelo. En efecto, la columna y asociada a una de esas nuevas variables tendría el mismo valor para cada observación perteneciente a una entidad. Denótese dicho valor por $y_i, i = 1, \dots, 32$. Es fácil ver que $y = y_1 * e_1 + y_2 * e_2 + \dots + y_{32} * e_{32}$. En otras palabras, la *nueva variable* puede ser expresada como combinación lineal de las *dummies*

estatales ya en el modelo. En consecuencia, también está contenida en el subespacio lineal generado por éstas por lo que no aporta información adicional y, en cambio, resulta en el problema aritmético de colinealidad. Asimismo, basta con usar $\mathbf{1} - (e_2 + \dots + e_{32})$ en lugar de e_1 en la anterior expresión para ver que y también puede ser expresado como combinación lineal de los vectores que caracterizan a la primera reparametrización; i. e., $y = y_1 * \mathbf{1} + (y_2 - y_1) * e_2 + \dots + (y_{32} - y_1) * e_{32}$. Por lo tanto, se repite en este caso que y no aporta información adicional y sí resulta en el problema aritmético de colinealidad.

3. Metodología

3.1 Su aplicación³

Modelo lineal ajustado a datos del 2014

Con el fin de aproximar lo que habrían sido los valores del ICT de los hogares en el 2015 de no haber mediado cambio alguno en las circunstancias, tanto exógenas como endógenas, en que la información fue recopilada, se recurre a la experiencia de levantamientos anteriores. Aunque, como ya se mencionó, se experimentó tanto con modelos para cada uno de los años entre el 2008 y 2014 como para la base de datos conjunta para el mismo periodo obteniendo ajustes de calidades similares, se ejemplificará la aplicación de la metodología con los resultados obtenidos con el modelo ajustado a los datos del 2014.

En primera instancia, es necesario construir el archivo a partir del cual se construirá el modelo que servirá de base para imputar el valor de los ingresos corrientes totales de los hogares en el 2015. Las variables requeridas para este fin se encuentran dispersas en tres archivos por lo menos. Por esta razón, se requiere hacer uso de la instrucción *merge*

³ Para mayor detalle, favor de referirse al apéndice 3 en la versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>.

un par de veces. En el apéndice dos⁴ se ejemplifica con el caso del levantamiento del MCS 2014.

En vista de que la forma en que las variables han sido codificadas a lo largo del tiempo puede haber cambiado, se aplican diversas transformaciones con el propósito de alcanzar homogeneidad. A manera de ejemplo se tiene a la variable *antigüedad*, cuya presentación fue modificada ya que en el ejercicio del 2008 se las reportaba a través de ocho categorías, en tanto que los ejercicios posteriores dicha variable adquirió un carácter numérico representando los años desde que fue construida la vivienda.

Acto seguido, y con el propósito de realizar el ajuste en condiciones tan similares como sea posible a las que tendrá la base ajustada del MCS 2015, la base de datos es posestratificada del mismo modo que se llevó a cabo esta operación para el MCS 2015, según se describió líneas arriba.

Con el fin de permitir una mayor flexibilidad de los coeficientes de los modelos, se procede a realizar el ajuste de cuatro modelos semejantes, uno para cada uno de los estratos socioeconómicos en que se clasifican las unidades primarias de muestreo.

3.2 Imputación de los valores del ICT por hogar en el 2015⁵

Los valores de los coeficientes resultantes en la estimación anterior son utilizados para identificar observaciones correspondientes al estrato de que se trate, cuyo valor de la variable ingresos resulta estar alejado del promedio para el estrato. Estas observaciones son marcadas como aberrantes y no serán consideradas en ajustes posteriores del mismo modelo ya que pueden sesgar sus resultados. Se procede iterativamente de la misma manera hasta el momento en que ninguna otra observa-

⁴ Ver la versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>.

⁵ Ver apéndice 2 en versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>.

Instrucción

Continúa

...

```
replace lict_15_14_2= cond(ent==2,0.2084838,0,0)+cond(ent==3,0.2902873,0,0)+
cond(ent==4,0.2692921,0,0)+cond(ent==5,0.054058,0,0)+cond(ent==6,0.1329593,0,0)+
cond(ent==7,0.0138338,0,0)+cond(ent==8,-0.039243,0,0)+cond(ent==9,0.207832,0,0)
+cond(ent==10,-0.0118056,0,0)+cond(ent==11,0.0263964,0,0)+cond(ent==
12,0.0540763,0,0)+cond(ent==13,0.0520245,0,0)+cond(ent==14,0.080892,0,0)+
cond(ent==15,0.146192,0,0)+cond(ent==16,-0.0407519,0,0)+cond(ent==17,0.0226898,0,0)+
cond(ent==18,0.0431858,0,0)+cond(ent==19,0.1715621,0,0)+cond(ent==20,0.112118,0,0)+
cond(ent==21,-0.0475403,0,0)+cond(ent==22,0.2300557,0,0)+cond(ent==23,0.2887406,0,0)
+cond(ent==24,0.0406937,0,0)+cond(ent==25,0.0119451,0,0)+cond(ent==26,0.102442,0,0)+
cond(ent==27,0.1931561,0,0)+cond(ent==28,0.0250334,0,0)+cond(ent==29,-0.0780366,0,0)+
cond(ent==30,0.0302744,0,0)+cond(ent==31,0.051414,0,0)+cond(ent==32,-0.1043155,0,0)
+cond(est_socio==2,0,0,0)+cond(est_socio==3,0,0,0)+cond(est_socio==4,0,0,0)+cond
(sexo_jefe==2,-0.0239431,0,0)+cond(educ_a_jefe==2,-0.0727829,0,0)+cond(educ_a_jefe==
3,-0.0126718,0,0)+cond(educ_a_jefe==4,0.0522974,0,0)+cond(educ_a_jefe==5,0.0541993,0,0)
+cond(educ_a_jefe==6,0.0984958,0,0)+cond(educ_a_jefe==7,0.1205507,0,0)+
```

```

cond(educ_a_jefe==8,0.1860793,0,0)+cond(educ_a_jefe==9,0.2323969,0,0)+cond
(educ_a_jefe==10,0.4338892,0,0)+cond(educ_a_jefe==11,0.6434069,0,0)+cond(mat_pared==
2,-0.1118209,0,0)+cond(mat_pared==3,-0.0526059,0,0)+cond(mat_pared==4,-0.1952233,0,0)+
cond(mat_pared==5,-0.1640854,0,0)+cond(mat_pared==6,-0.1006038,0,0)+cond
(mat_pared==7,-0.1498196,0,0)+cond(mat_pared==8,-0.109697,0,0)+cond(mat_techos==
2,-0.0061939,0,0)+cond(mat_techos==3,-0.0498149,0,0)+cond(mat_techos==4,-0.0255521,0,0)+
cond(mat_techos==5,-0.1267531,0,0)+cond(mat_techos==6,-0.0521155,0,0)+cond(mat_techos==
7,-0.0312835,0,0)+cond(mat_techos==8,-0.100727,0,0)+cond(mat_techos==9,-0.0088913,0,0)
+cond(mat_pisos==2,0.0456613,0,0)+cond(mat_pisos==3,0.097119,0,0)+cond(cocina==2,
-0.0484888,0,0)+cond(cocina==3,-0.0050013,0,0)+cond(dotac_agua==2,0.008699,0,0)+cond
(dotac_agua==3,0.0134096,0,0)+cond(dotacagua==4,0.008181,0,0)+cond(dotac_agua==
5,0.0066653,0,0)+cond(excusado==2,-0.0565791,0,0)+cond(excusado==3,-0.0655634,0,0)
+cond(drenaje==2,-0.0239128,0,0)+cond(drenaje==3,-0.0437316,0,0)+cond(drenaje==4,
-0.0631023,0,0)+cond(drenaje==5,-0.0501556,0,0)+cond(disp_elect==2,0.1026179,0,0)+
cond(disp_elect==3,0.1407885,0,0)+cond(disp_elect==4,0.1148549,0,0)+cond(disp_elect==
4,0,0,0)+cond(combustible==2,0.0699868,0,0)+cond(combustible==3,0.2135501,0,0)
+cond(combustible==4,0.2715266,0,0)+cond(combustible==5,0.1605489,0,0)+cond(combustible==
6,0.3482146,0,0)+cond(eli_basura==2,-0.0050362,0,0)+cond(eli_basura==3,-0.061659,0,0)+
cond(eli_basura==4,-0.0525909,0,0)+cond(eli_basura==5,0.0874761,0,0)+cond(eli_basura==
6,-0.244286,0,0)+cond(tenencia==2,0.0544615,0,0)+cond(tenencia==3,0.1852454,0,0)+
cond(tenencia==4,0.0963684,0,0)+cond(tenencia==5,0.0213532,0,0)+cond(tenencia==
6,0.0524394,0,0)+cond(lavadero==2,-0.0236408,0,0)+cond(fregadero==2,-0.0458118,0,0)
+cond(regadera==2,-0.0226608,0,0)+cond(tinaco_azo==2,0.0143681,0,0)+cond
(cisterna==2,-0.0195102,0,0)+cond(pileta==2,-0.0034837,0,0)+cond(calentador==
2,-0.051597,0,0)+cond(medidor_luz==2,-0.0200378,0,0)+cond(bomba_agua==2,-0.0129787,0,0)
+cond(tanque_gas==2,-0.0449267,0,0)+cond(aire_acond==2,-0.0628015,0,0)+cond
(calefact==2,-0.1035879,0,0)+cond(tam_loc==2,-0.0190362,0,0)+cond(tam_loc==3,-0.0361273,0,0)
+cond(tam_loc==4,-0.0344768,0,0)+cond(antiguedad==2,-0.0086855,0,0)
+cond(antiguedad==3,-0.0323411,0,0)+cond(antiguedad==4,-0.0442789,0,0)+cond
(antiguedad==5,-0.0236441,0,0)+cond(antiguedad==6,0.0120303,0,0)+cond(antiguedad==
7,-0.0025602,0,0)+cond(antiguedad==8,0,0,0)+0.0005391*edad_jefe+0.021793*tot_
integ+0.0633045*percep_ing+0.2257462*perc_ocupa+0.0168226*num_focos+
0.0162275*cuart_dorm+0.1299516*num_auto+0.1059354*num_van+0.0632263*
num_pickup-0.0012772*num_moto-0.0316606*num_bici-0.1118853*num_trici-0.0262355*
num_carret-0.1601213*num_canoa+0.079093*num_otro+0.0290637*num_ester-0.0105932*
num_grab-0.0109887*num_radio+0.033758*num_tv+0.0430486*num_dvd-0.0185274*
num_video+0.0033888*num_licua+0.048953*num_tosta+0.0411183*num_micro+0.0782377*
num_refri+0.0014575*num_estuf+0.0416103*num_lavad+0.0308168*num_planc-
0.0060453*num_maqui+0.0365429*num_venti+0.0457647*num_aspir+0.0638511*
num_compu+0.0395288*num_impre+0.0178456*num_juego+9.060962 if est_socio==2
...

generate ict_15_14_2 = exp(lict_15_14_2)*1.023848383

```

ción sea identificada como aberrante. A partir de este momento, el siguiente modelo ajustado permitirá definir un valor nuevo para el logaritmo de los ingresos corrientes totales de los hogares en la muestra del MCS 2015. La instrucción correspondiente se ejemplifica enseguida en el recuadro.⁶

En efecto, el nombre de la variable que contiene los valores imputados del ICT para el 2015 es `ict_15_14_2`, después de aplicar un factor inflacionario de casi 2.4 por ciento. Más aún, a partir de estos valores es posible definir nuevos deciles de ingreso para el mismo año 2015.

Comparación con el 2014

Como se verá en la siguiente sección, los resultados obtenidos por esta vía fueron también insatisfactorios. Mientras que, por un lado, el ingreso corriente total promedio por hogar resultó ser menor aún que la misma cantidad que se obtendría de levantamientos anteriores, expresadas éstas en pesos del 2015, por el otro, los mismos promedios, pero ahora por deciles, mostraban, para los deciles menores, incrementos aún mayores que los que resultaron de los datos levantados en campo y que ya eran considerados excesivos; para los deciles mayores, en cambio, se tenían decrementos importantes.

De este modo se hizo evidente que los resultados obtenidos del modelo para el 2015 eran aún menos comparables. Entonces, si el propósito es el de producir bases de datos comparables en el tiempo, se decidió explorar algunas alternativas que nos acercaran al cumplimiento de este objetivo. La búsqueda de opciones se basó en la consideración de que el modelo no consideraba la coyuntura de cada periodo. Quedaba claro ahora que dicha coyuntura limitaba la calidad del ajuste pues, aunque la R^2 resultaba ser significativamente diferente de 0, también era significativamente diferente de 1 para todos los modelos ajustados. Entonces, o bien se ig-

nora tal coyuntura para los otros levantamientos, o se incluye una similar a la del 2014 en las estimaciones del ingreso para el MCS 2015.

Estimación de ingresos para el 2014 usando el modelo

Cabe señalar que el mismo modelo puede ser usado para imputar el valor del ICT de los hogares en el propio 2014. Más aún, a partir de estos valores es posible definir nuevos deciles de ingreso para el mismo año 2014. Cuando los ingresos corrientes totales de los hogares para el 2014 fueron, de este modo, estimados con el modelo ajustado a datos para ese mismo año las comparaciones entre el 2014 y 2015 tomaron un giro muy favorable. Sin embargo, la comparación entre los levantamientos del 2012 y 2014 exhibió un importante deterioro.

Interpretamos estos hechos como que el concepto de ingreso corriente usado para el cálculo de valores en la encuesta y el que resulta de la aplicación del modelo son diferentes y, en consecuencia, incomparables. Entonces, aun cuando podíamos aplicar el modelo a las bases de datos de los levantamientos para el 2008, 2010 y 2012, con fines de comparación, con lo que obtendríamos valores diferentes a los utilizados por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) en sus mediciones de pobreza, decidimos intentar el enfoque alternativo de acercar al ingreso calculado a partir del modelo al concepto de ingreso corriente que se utiliza en dichas mediciones.

Complementación del modelo para el 2015 con errores de ajuste para el 2014

El inusual comportamiento de los residuos del ajuste para el 2014, en particular lo pesado de las colas de la distribución de dichos residuos, nos ha llevado a pensar que la variabilidad no explicada por el modelo tiene que ser tomada en cuenta de algún modo. Para conseguir este propósito, decidimos generar aleatoriamente tantos valores de di-

⁶ Ver apéndice 2 en versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol.9, número especial, 2018: <http://rde.inegi.org.mx>.

chos residuos como el tamaño muestral dentro de cada uno de los cuatro estratos sociodemográficos. Para ello, echamos mano de la distribución empírica de los residuos del 2014 utilizando para ello el factor de expansión asociado a cada observación y, en consecuencia, a cada residuo.

En vista de que los paquetes estadísticos a nuestra disposición no contemplan una simulación tal, fue necesario desarrollar una rutina en FORTRAN.⁷ Dicho programa toma en cuenta por separado cada uno de los cuatro estratos sociodemográficos. En uno de los archivos de entrada se ordenaron de menor a mayor los residuos del ajuste para el 2014. Se obtuvieron las sumas parciales de sus correspondientes factores de expansión, estandarizándolas de modo que el valor final fuese igual a 1. Es necesario destacar que se dejó abierta la posibilidad de que los residuos generados tomaran el valor 0, lo que más adelante se usó como indicativo de que el ingreso correspondiente también sería igual a 0. Acto seguido se generaron tantos valores de una variable aleatoria uniforme en el intervalo [0,1] como observaciones en la muestra del 2015 para cada estrato sociodemográfico. El valor simulado correspondió al que tenía asociado el intervalo dentro del cual se contenía el valor uniforme generado aleatoriamente. En vista de que el tamaño de la muestra en cada caso importaba varios miles, aún los valores más extremos de los residuos aparecieron entre los simulados. El nuevo vector de errores fue sumado, sin modificar su orden, al del logaritmo de los ingresos generado por el modelo. Esta suma fue convertida a pesos del 2014 y, aplicando un factor de inflación, éstos fueron expresados en pesos del 2015.

Comparación con el 2014 y otros periodos comparables

El nuevo conjunto de valores del ingreso para el 2015 incluyó una proporción con valores pequeños o cero semejante a la obtenida para el 2014. Como se verá en la siguiente sección, la compara-

ción de los resultados obtenidos con los datos así generados con los datos originales del 2014, y del 2012, y del 2010, etc., no da lugar a grandes comentarios. Y eso es, precisamente, lo destacable de este enfoque.

4. Resultados

4.1. Resultados del ajuste del modelo lineal a nivel nacional, por entidad federativa y por decil con la nueva base del MCS 2015

De manera similar a lo descrito en la sección anterior, diversos modelos fueron ajustados a datos para el periodo 2008-2014 tanto para cada año individual como para el conjunto de los cuatro levantamientos incluidos en el periodo. En cada caso, después de aplicar un factor inflacionario adecuado, se obtuvieron los valores para el ingreso corriente total de los hogares en el 2015. Los valores que se muestran a continuación son el resultado de ajustar por separado un modelo para cada uno de los cuatro estratos sociodemográficos en que se clasifican las unidades primarias de muestreo del Marco Nacional de Viviendas del Instituto Nacional de Estadística y Geografía (INEGI). Enseguida, la tabla 1.a muestra un resumen de los resultados de dichos ajustes.

La figura 4 ejemplifica lo anterior mostrando el promedio del ICT obtenido mediante la aplicación de tres modelos. El primero de ellos es el obtenido de la fusión de las cuatro bases de datos del periodo en una sola. El segundo es el obtenido antes de proceder a la eliminación de las observaciones aberrantes en el 2014. El tercero corresponde a lo descrito en la sección anterior. La cuarta barra de esta figura muestra el valor obtenido mediante estimación directa a partir de la base de datos del MCS 2015. Es claro que, en todos los casos anteriores, el promedio de los ingresos imputados asume un valor significativamente menor al publicado en julio del 2016. Aunque no se exhibe en la mencionada figura, los promedios de los valores imputados son también menores a los correspondientes publicados para los ejercicios del periodo, expresados todos ellos en pesos del 2015. Del mismo

⁷ Ver apéndice 2 en versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>.

Tabla 1.a

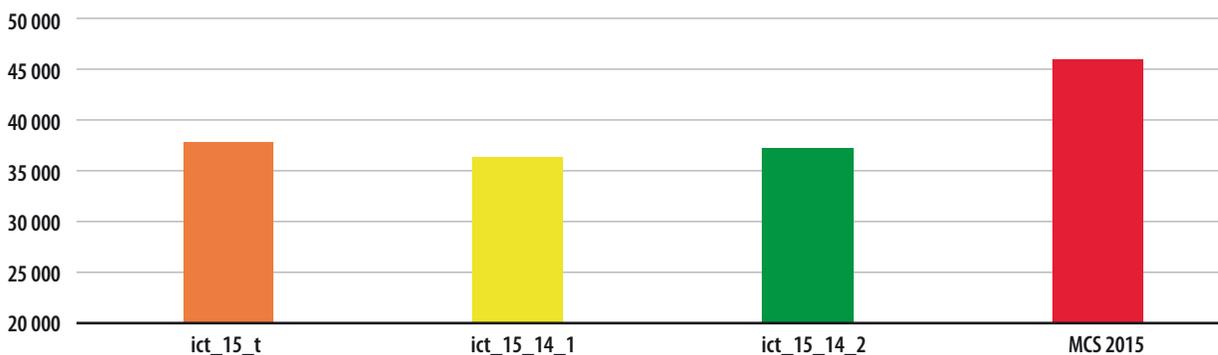
Estadísticas resumen del ajuste de modelos lineales para el levantamiento del MCS para el 2014

Estrato sociodemográfico bajo					
Number of strata	=	1	Number of obs.	=	10 921
Number of PSUs	=	653	Population size	=	5 837 412
N. of poststrata	=	156	Subpop. no. of obs.	=	10 492
			Subpop. size	=	5 575 071.3
			Design df	=	652
			F(143, 510)	=	259.45
			Prob > F	=	0.0000
			R-squared	=	0.5063
Estrato sociodemográfico medio bajo					
Number of strata	=	1	Number of obs.	=	30 240
Number of PSUs	=	3 786	Population size	=	15 924 126
N. of poststrata	=	156	Subpop. no. of obs.	=	29 740
			Subpop. size	=	15 762 860
			Design df	=	3 785
			F(144, 3 642)	=	143.52
			Prob > F	=	0.0000
			R-squared	=	0.5 592
Estrato sociodemográfico medio alto					
Number of strata	=	1	Number of obs.	=	12 150
Number of PSUs	=	2 486	Population size	=	7 133 851
N. of poststrata	=	156	Subpop. no. of obs.	=	11 969
			Subpop. size	=	7 047 327.5
			Design df	=	2 485
			F(142, 2 344)	=	74.97
			Prob > F	=	0.0000
			R-squared	=	0.5717
Estrato sociodemográfico alto					
Number of strata	=	1	Number of obs.	=	4 814
Number of PSUs	=	1 185	Population size	=	3 255 003
N. of poststrata	=	156	Subpop. no. of obs.	=	4 753
			Subpop. size	=	3 199 125.7
			Design df	=	1 184
			F(132, 1 053)	=	.
			Prob > F	=	.
			R-squared	=	0.6217

Fuente: cálculos propios a partir de las bases de datos del levantamiento del MCS para el 2014.

Figura 4

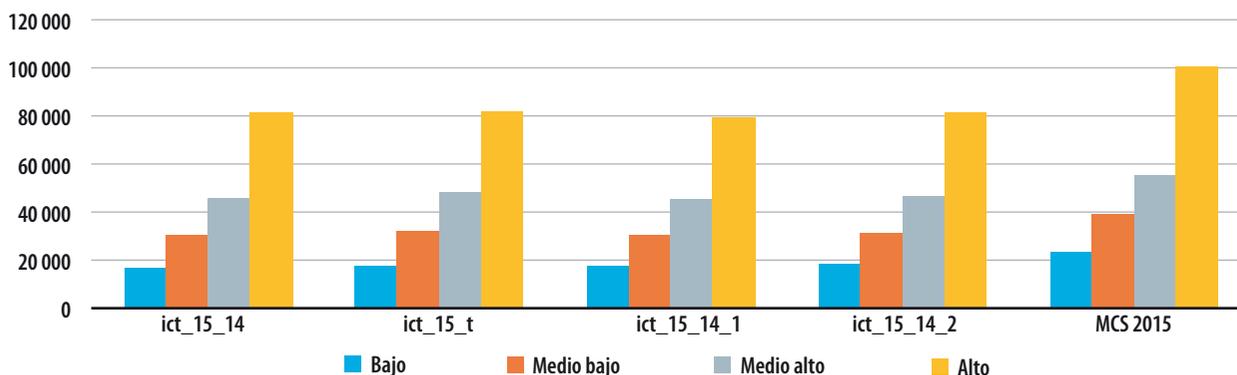
ICT promedio por hogar estimado por diferentes modelos, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 5

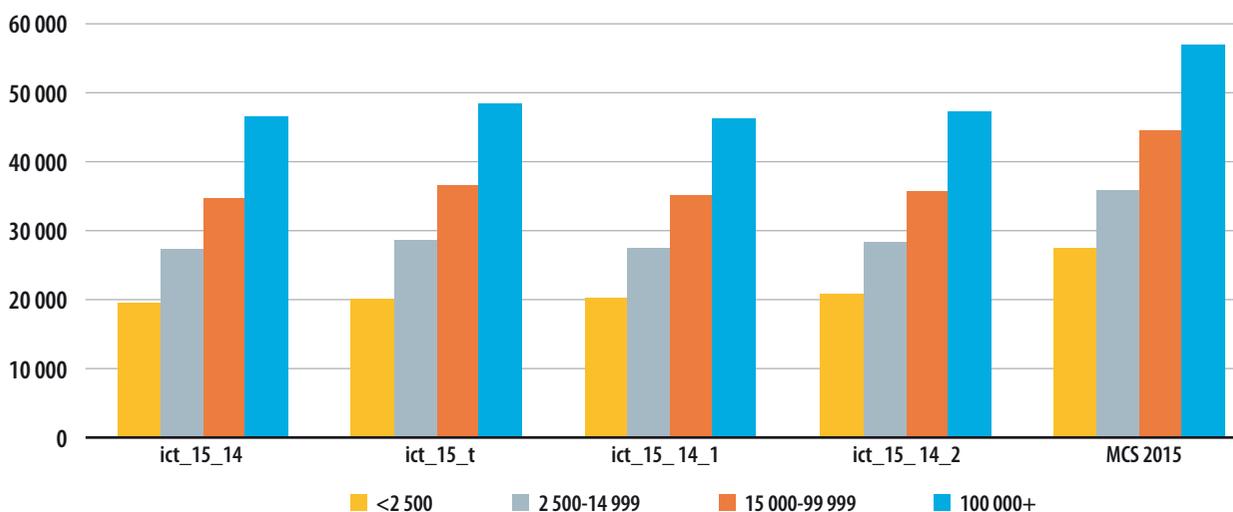
ICT promedio por hogar según estrato sociodemográfico y modelos ajustados, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 6

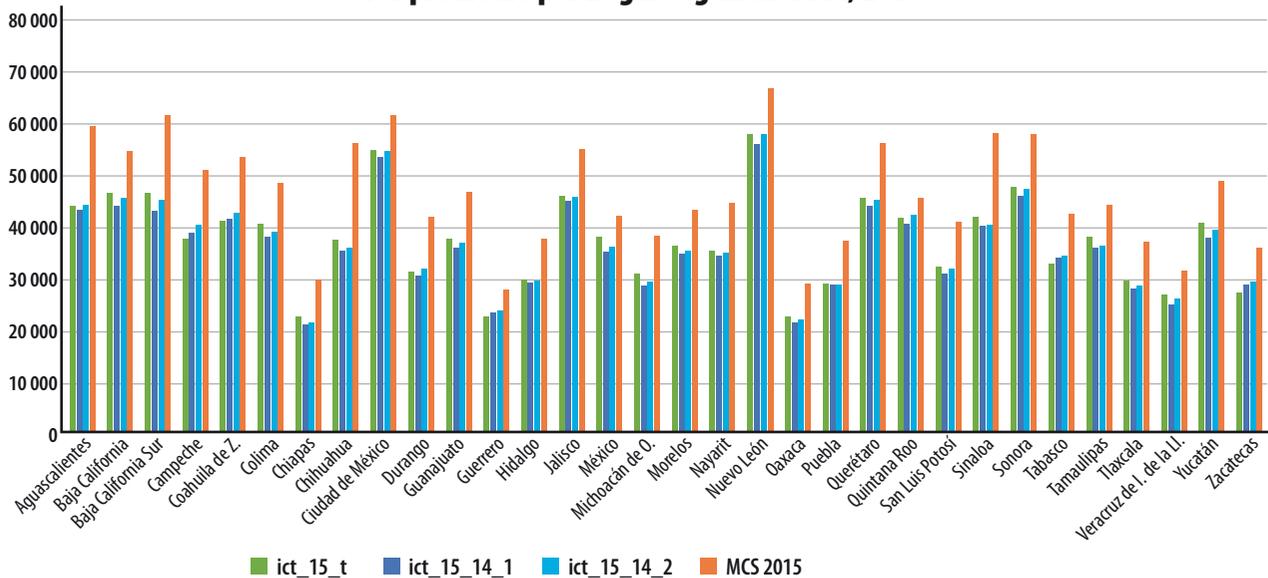
ICT promedio por hogar según tamaño de localidad, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 7

ICT promedio por hogar según entidad, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

modo, los correspondientes promedios por estrato socioeconómico, por tamaño de localidad y por entidad federativa se muestran en las figuras 5, 6 y 7, respectivamente. En todos los casos, los valores obtenidos a partir de los modelos ajustados se encuentran por debajo de los que fueron publicados en su momento por los ejercicios del periodo.

4.2 Comparación de los resultados con los del MCS 2015 y la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)-MCS 2012 y 2014, por decil

Cuando las comparaciones tienen lugar al nivel de los deciles, la imagen que obtenemos cambia de manera importante. En general, la discusión se basará en la tabla 3 en la que se presentan los resultados para los cálculos para el promedio del ICT de los hogares para los años 2012, 2014 y 2015, según fueron publicados por INEGI; se presentan, asimismo, los cálculos correspondientes a los ingresos según fueron imputados por el modelo ajustado a datos del 2014, para ese mismo año y para el 2015. Queda claro que los cálculos a nivel nacional resultan en valores menores cuando se obtienen a partir del modelo que los que resultan de la estimación directa.

Las figuras 8, 9 y 10 exhiben las tasas de crecimiento de los valores en la tabla 1, tanto entre el 2012 y 2014 como entre el 2014 y 2015. En particular, la figura 8 resume las causas de la sorpresa de los usuarios de los datos del MCS 2015, mostrando tasas de crecimiento entre el 2014 y 2015 exageradas, particularmente para los deciles de ingresos más bajos. El crecimiento exorbitante se hace más evidente cuando es comparado con el crecimiento observado a lo largo de dos años, entre el 2012 y 2014. No bastó con la advertencia de que los resultados del MCS 2015 no eran comparables con los de ejercicios anteriores para evitar que se llevaran a cabo precisamente dichas comparaciones. Es nuestra contención que los resultados del presente ejercicio permiten observar de manera más clara las razones por las cuales, en efecto, el levantamiento del 2015 no puede ni debe ser comparado con levantamientos anteriores sin antes llevar a cabo una serie de ajustes que eliminen las más importantes discrepancias entre ellos.

La figura 9 muestra nuevamente las tasas del crecimiento observado en el periodo 2012 al 2014, pero reemplaza las tasas observadas entre el 2014 y 2015 por las que se obtienen al estimar los promedios del ingreso usando los datos imputados para el 2015 a partir del modelo ajustado a datos del

Tabla 3

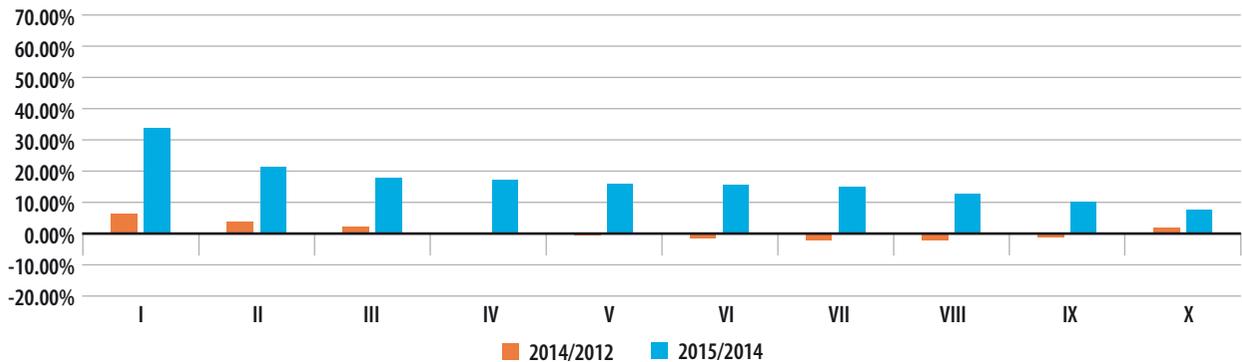
Promedio del ICT por deciles de hogares (precios constantes del 2015)

	2012	2014	2014 (imputado)	2015	2015 (Imputado)
Ingreso corriente	40 959	41 000	35 685	45 887	37 221
I	5 757	6 116	9 314	8 169	9 866
II	10 999	11 409	13 765	13 830	14 396
III	15 318	15 622	17 014	18 387	17 795
IV	19 768	19 778	20 419	23 147	21 184
V	24 662	24 564	24 190	28 478	24 928
VI	30 637	30 178	28 744	34 866	29 513
VII	38 264	37 433	34 376	43 030	35 391
VIII	49 108	48 105	42 561	54 139	43 652
IX	67 419	66 510	57 570	73 252	58 390
X	147 662	150 283	115 462	161 568	117 086

Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

Figura 8

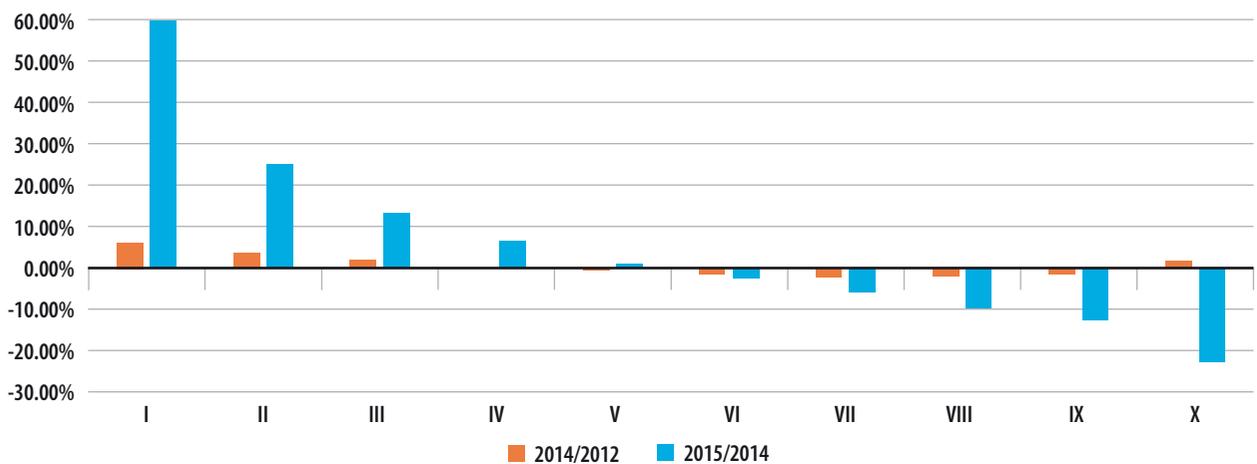
Crecimiento del ingreso corriente promedio, 2012-2014 y 2014-2015, según el INEGI



Fuente: INEGI, archivos del MCS 2015.

Figura 9

Crecimiento del ingreso corriente promedio, 2012-2014, según el INEGI, y 2014-2015, según modelo



Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

2014. El resultante crecimiento para los deciles de ingresos más bajos es ahora aún mayor, en tanto que para los deciles de más altos ingresos se observan decrementos que llegan a ser sustanciales. Por supuesto, cabe preguntarse sobre las razones por las que se observa un comportamiento aún más desfavorable para los ingresos menores. La respuesta parece estar en la modificación de la estructura de ingresos bajos, incluidos los que toman el valor 0, misma que es exacerbada por el modelo. En efecto, el número de ingresos con valor 0 en el levantamiento del 2015 se redujo sustancialmente en relación con lo ocurrido en años anteriores; algo similar ocurre con los ingresos positivos, pero pequeños. Por su parte, el modelo hace aún menos comparables los resultados ya que no solo reduce el número de ingresos con valor 0 ó pequeño, sino que los elimina por completo.

Cuando, haciendo uso del modelo, se modifica también la estructura del ingreso en su parte inferior también para el 2014, haciéndola comparable con la imputación del ingreso para 2015, se obtiene la figura 10. Bajo estas circunstancias, se observan crecimientos moderados entre el 2014 y 2015 para

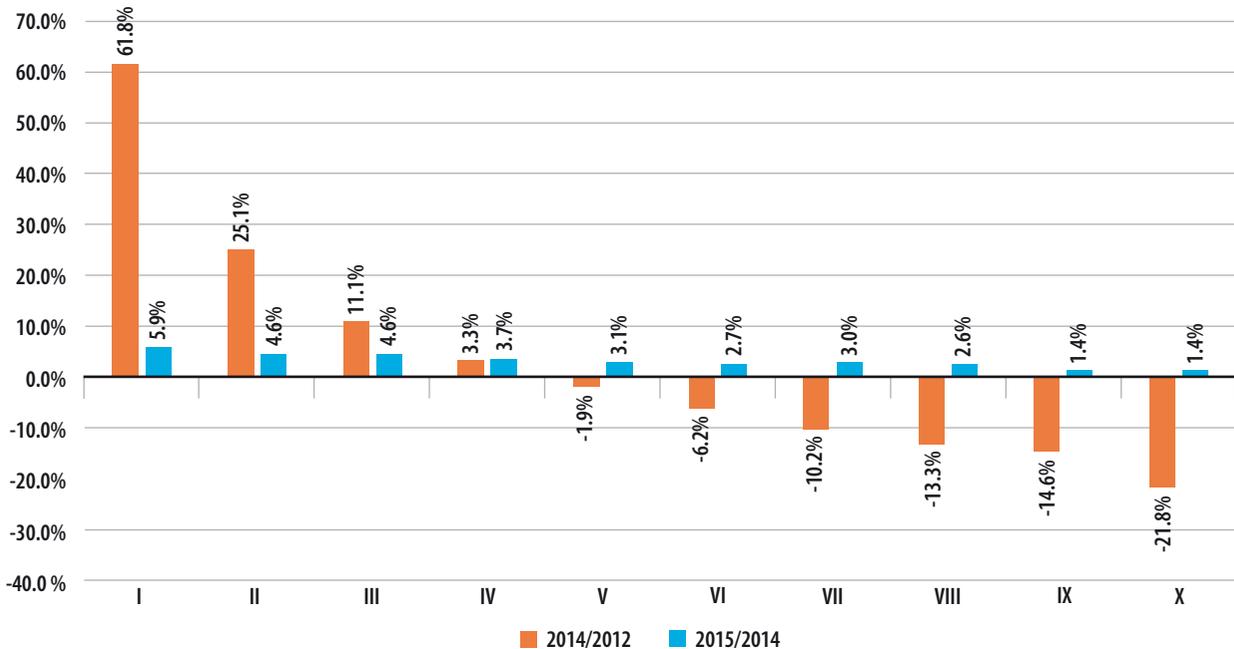
todos los deciles de ingreso. De hecho, el mayor crecimiento se da en el primer decil, pero éste alcanza *apenas* 5.9 por ciento. Por su parte, los deciles de ingresos mayores exhiben un crecimiento inferior a 1.5 por ciento. Al parecer, en este caso, las comparaciones se están llevando a cabo entre estructuras comparables por construcción.

Se observa, además, que es ahora la comparación entre los años 2012 y 2014 la que da lugar a crecimientos elevadísimos. Por supuesto, podemos decir que también en este caso se estaría realizando una comparación a todas luces indebida, pues están involucradas dos estructuras de ingreso incomparables, en particular, en la parte inferior de la distribución.

Llama la atención, además, el hecho de que el modelo, por así decirlo, *achate* la distribución recorriendo de manera notable las colas, tanto inferior como superior. Ello no parece deberse solamente a la eliminación de observaciones con valores aberrantes en el momento de llevar a cabo el ajuste de los modelos.

Figura 10

Crecimiento del ingreso corriente promedio, 2012-2014 y 2014-2015, según modelo



Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

4.3 Resultados del ajuste complementario al modelo lineal a nivel nacional por estado y decil con la nueva base del MCS 2015

En esta sección se presenta una forma alternativa de *emparejar el piso*, es decir, de construir una nueva base de datos que sea en efecto comparable con las anteriores. En el apartado anterior, dicha comparación se logró cuando tanto los ingresos del módulo 2014 como los del 2015 fueron reemplazados por los valores arrojados por el modelo lineal ajustado usando la base del 2014. En vista de la naturaleza de las variables explicativas en ese modelo, en alguna medida asociada a los *stocks* acumulados por los hogares a lo largo del tiempo, los resultados arrojados por el modelo podrían asociarse a un concepto más cercano a lo que se ha denominado *ingreso permanente* en la literatura económica. A nuestro juicio, al hacer caso omiso de las eventualidades que enfrentan los hogares y que afectan su ingreso corriente, da lugar a una simplificación conceptual y, con ello, permite la realización de comparaciones posiblemente sesgadas a causa de diversas influencias: muestrales, económicas, sociales, etcétera. Lo anterior aunado a la convicción de que los cambios operativos introducidos durante el levantamiento del 2015 tuvieron un fuerte impacto en la distribución de los ingresos recogidos en la muestra y, en particular, en los ingresos más bajos. La distorsión introducida de un año al siguiente es de tal manera sustancial que el ajuste obtenido a partir del modelo lineal solamente se acercará a ser comparable con los resultados obtenidos en el 2014 cuando los resultados para el 2015 sean afectados tomando en cuenta la distribución de las discrepancias entre las observaciones y el modelo según la experiencia del 2014.

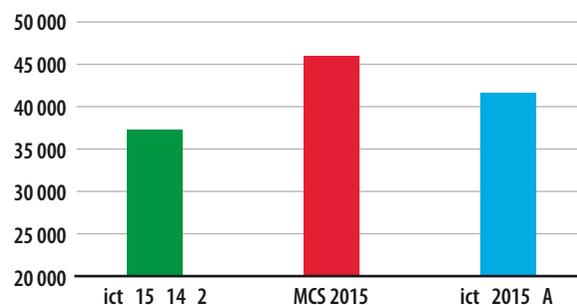
Con el propósito de evitar la introducción de decisiones arbitrarias al hacer la imputación de los errores que afectarán las estimaciones del 2015, se determinó que lo más conveniente era llevar a cabo una simulación de un número suficiente de discrepancias como para cubrir la totalidad de las unidades muestrales.⁸ De hecho, se realizó

⁸ Ver apéndice 3 en versión electrónica de *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. Vol. 9, número especial, 2018: <http://rde.inegi.org.mx>

una simulación para cada uno de los estratos sociodemográficos en que se clasifican las unidades primarias de muestreo. Asimismo, a lo largo de la simulación se concedió una probabilidad positiva, en cada uno de los estratos, de tener una respuesta igual a cero. Ya que la distribución específica obtenida a partir de la agregación de los factores de expansión y estandarizada para cada estrato no estaría disponible en los paquetes estadísticos tradicionales, se recurrió al uso de macros que pueden ser reproducidas al interior de hojas de *Excel*. Para este fin, fue suficiente el desarrollo de cinco de estas macros, la primera de las cuales toman el valor simulado y lo copia al final de la columna de valores simulados; la segunda repite 10 veces esta acción; la tercera, 10 veces a la segunda; la cuarta, 10 veces a la quinta; y, finalmente, la quinta, 10 veces a la cuarta. De este modo, con un solo llamado a la quinta macro se generan 10 mil valores de la distribución correspondiente. Los valores simulados son añadidos a las estimaciones de los ingresos producidas por el modelo. La única salvedad consiste en reemplazar al ingreso todo por el valor 0 cuando el valor simulado era también igual a cero. Con el fin de tener más de una simulación, se desarrolló la mencionada rutina en FORTRAN. La simulación seleccionada es la que resulta en la mediana del ICT promedio.

La figura 4.a permite una rápida y primera comparación a partir del primero del ingreso corriente total de los hogares mexicanos para el 2015. En

Figura 4.a
ICT promedio por hogar estimado por diferentes modelos, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

ella se muestra dicho promedio, en primer lugar, cuando los valores son obtenidos directamente a partir del modelo lineal ajustado; en segundo, cuando los valores proceden de la base de datos original del módulo; y, finalmente, cuando los valores procedentes del modelo se le han agregado los de los residuos simulados, a lo que en adelante se denotará por *ict_2015_A*. Es necesario destacar que ahora esta última cantidad se encuentra ligeramente por arriba de la que correspondería al 2014 cuando sus valores son expresados en pesos del 2015, pero sustancialmente menor al que se obtiene de los valores originales.

Cuando los promedios nacionales son calculados para cada estrato sociodemográfico, así como para cada uno de los tamaños de localidad, se tiene otra vez que la nueva base de datos da lugar a promedios mayores que los del modelo, pero inferiores a los de la base de datos original. Dicha circunstancia se ilustra en las figuras 5.a y 6.a, respectivamente.

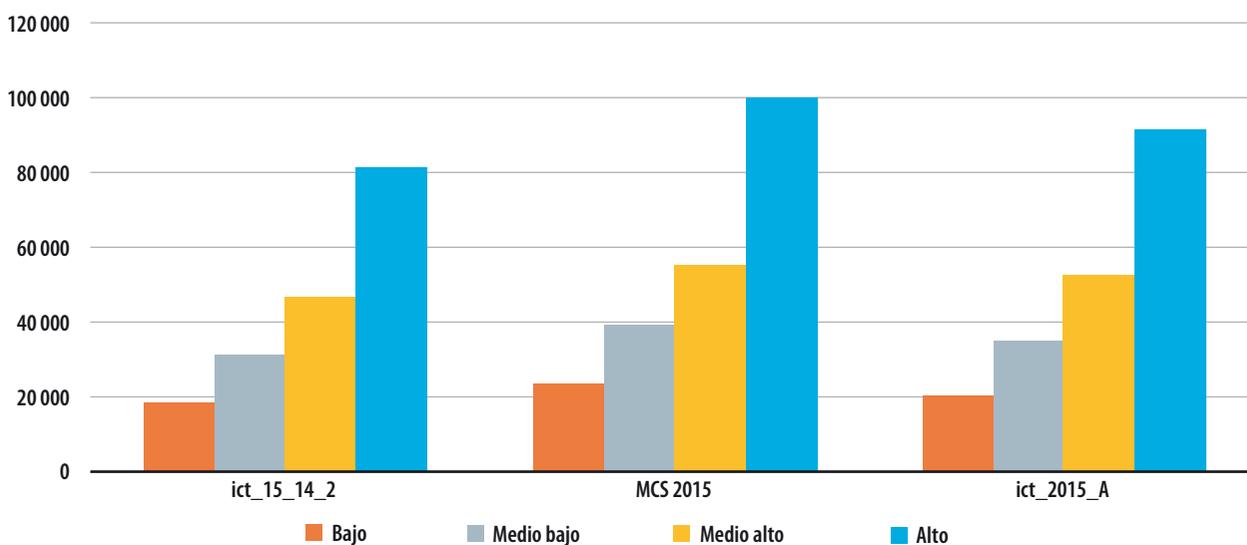
La figura 7.a, salvo por alguna excepción, exhibe un comportamiento similar para los promedios del ingreso corriente total cuando se le desagrega a nivel de entidad federativa.

La tabla 3.a muestra la desagregación por deciles de ingreso para los mismos promedios para diversos años. A nivel de los deciles de hogares se tendrían tanto algunos incrementos en los promedios como también algunos decrementos. Esta circunstancia no parece ser inusual ya que algo semejante se observa, por ejemplo, entre los años 2012 y 2014. Por lo que toca a los coeficientes de Gini, queda claro que el nuevo resultado está con una magnitud muy similar a la de años anteriores aunque con un ligero incremento.

Bajo estas circunstancias, se obtiene la figura 10.a. El incremento en el valor del promedio nacional de los ingresos corrientes totales entre el 2014 y 2015 asciende, ahora, a solamente 1.68 por ciento. Se observan, además, crecimientos moderados entre el 2014 y 2015 para algunos deciles de ingreso, pero en este caso también se exhiben decrementos para los deciles VI a IX. El mayor crecimiento se da en el primer decil, pero éste alcanza apenas 10.85 por ciento. Por su parte, el decil de ingresos mayores exhibe un crecimiento de 0.79 por ciento. Al parecer, nuevamente, las comparaciones se están llevando a cabo entre estructuras comparables por construcción.

Figura 5.a

ICT promedio por hogar según estrato sociodemográfico y modelos ajustados, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

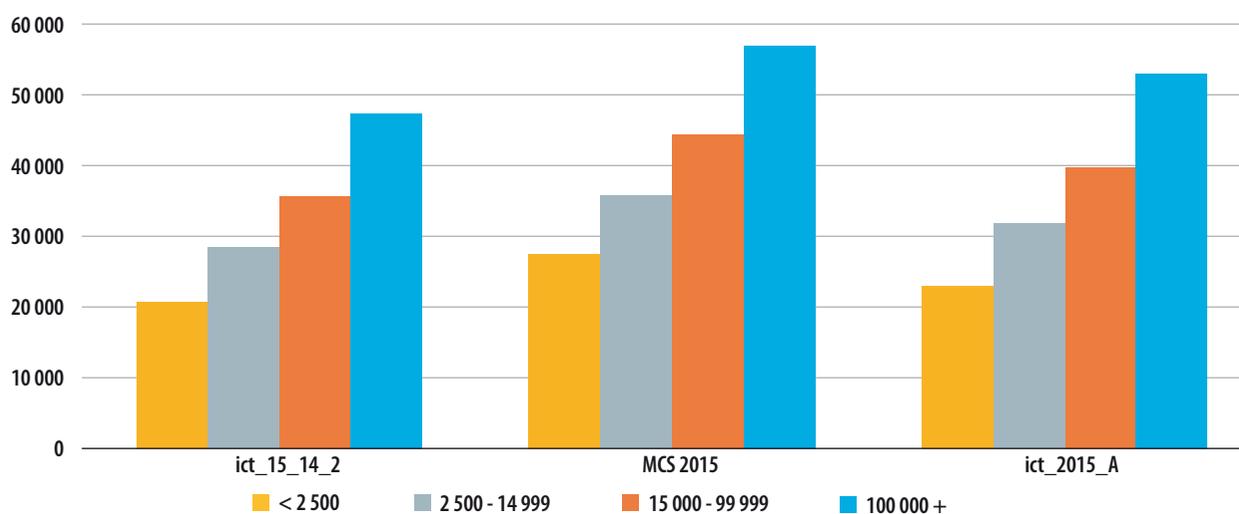
Tabla 3.a

Promedio del ICT por deciles de hogares y su coeficiente de Gini (precios constantes del 2015)

Deciles de hogares	2010	2012	2014	ict_2015_A
Ingreso corriente	42 725	40 959	41 000	41 549
I	5 944	5 757	6 116	6 746
II	11 386	10 999	11 409	11 638
III	15 838	15 318	15 622	15 665
IV	20 303	19 768	19 778	19 758
V	25 382	24 662	24 564	24 302
VI	31 511	30 637	30 178	29 901
VII	39 245	38 264	37 433	37 304
VIII	50 600	49 108	48 105	47 820
IX	70 588	67 419	66 510	66 825
X	156 457	147 662	150 283	155 527
Coefficiente de Gini	0.469	0.464	0.464	0.468

Figura 6.a

ICT promedio por hogar según tamaño de localidad, 2015



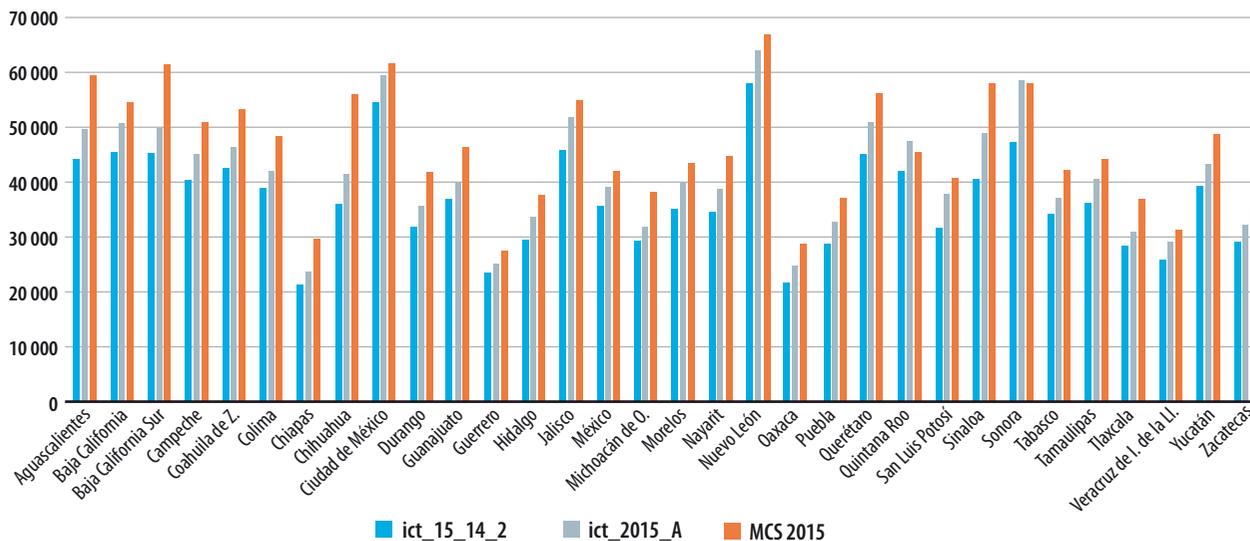
Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

En contraposición a lo observado en el caso anterior, el comportamiento de estos crecimientos porcentuales no se desvía significativamente del comportamiento exhibido por las mismas medidas en periodos anteriores. Ahora podemos decir que en este caso no se estaría realizando una comparación indebida, pues están involucradas dos estructuras

de ingreso comparables, en particular, en la parte inferior de la distribución; de hecho, una base de datos ajustada como la que se obtiene por esta segunda vía daría lugar a una mayor comparabilidad con todos los levantamientos previos y sin tener que recurrir a modificaciones a la información recopilada en éstos.

Figura 7.a

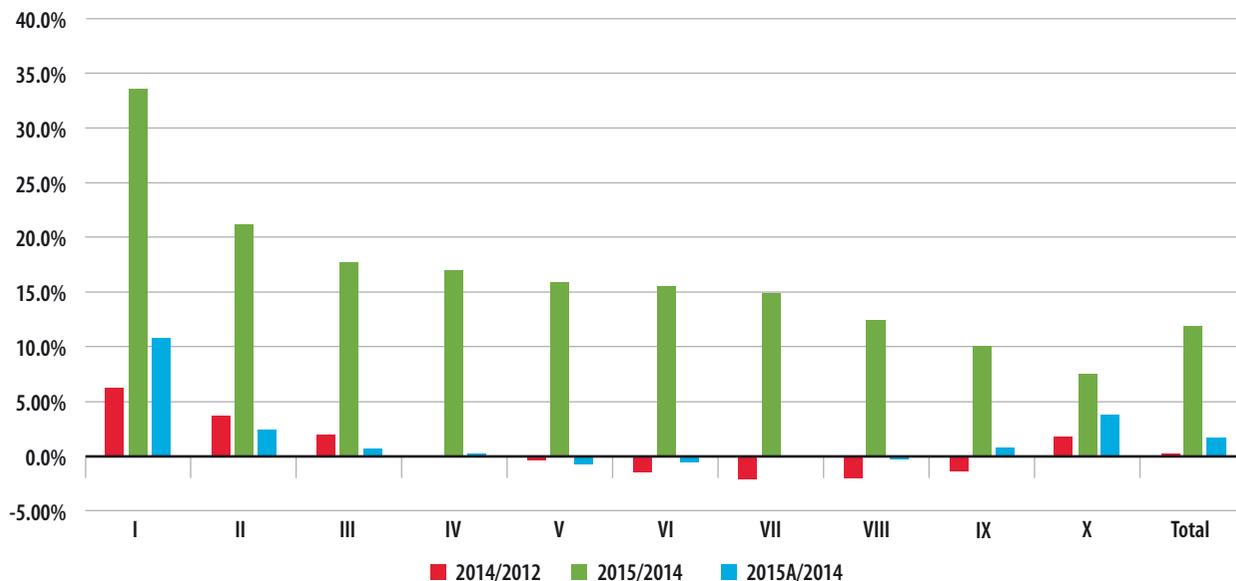
ICT promedio por hogar según entidad, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 10.a

Crecimiento del ingreso corriente promedio, 2012-2014 y 2014-2015, según modelo



Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

5. Validación y evaluación de la metodología

La tabla 3.a hace explícita la evolución de distintos rubros relacionados con el crecimiento del ingreso corriente total a lo largo del periodo 2010 a 2015. Para fines de comparación, solamente se hará refe-

rencia a la modificación designada ict_2015_A. La primera columna de cada bloque de cuatro contiene los crecimientos del ingreso corriente total para el periodo; la segunda, al del ingreso por perceptor; la tercera, al del número de perceptores por hogar; y, por último, la cuarta, al incremento en el número de hogares durante cada periodo.

Tabla 4

Descomposición del cambio del ingreso corriente total por deciles, 2012-2015

Decil	ICT ₁₂ /ICT ₁₀	IPP	PH	Hog	ICT ₁₄ /ICT ₁₂	IPP	PH	Hog
I	2.49%	-2.61%	-0.54%	5.81%	8.90%	-1.24%	7.56%	2.52%
II	2.21%	-3.90%	0.53%	5.81%	6.34%	-1.85%	5.68%	2.52%
III	2.33%	-6.23%	3.14%	5.81%	4.55%	-2.57%	4.68%	2.52%
IV	3.02%	-5.88%	3.44%	5.81%	2.58%	-3.49%	3.67%	2.52%
V	2.80%	-5.00%	2.28%	5.81%	2.12%	-2.82%	2.50%	2.52%
VI	2.87%	-6.28%	3.75%	5.81%	0.99%	-5.07%	3.76%	2.52%
VII	3.16%	-4.04%	1.60%	5.81%	0.30%	-3.67%	1.56%	2.52%
VIII	2.69%	-5.78%	3.01%	5.81%	0.43%	-3.04%	1.03%	2.52%
IX	1.05%	-8.81%	4.74%	5.81%	1.14%	1.30%	-2.61%	2.52%
X	-0.14%	-8.94%	3.65%	5.81%	4.34%	5.38%	-3.42%	2.52%
Total	1.43%	-6.65%	2.70%	5.81%	2.62%	-1.86%	1.99%	2.52%
Decil	ICT ₁₅ /ICT ₁₄	IPP	PH	Hog	ICT _{15A} /ICT ₁₄	IPP	PH	Hog
I	38.02%	31.34%	1.71%	3.32%	11.74%	-0.35%	10.70%	1.30%
II	25.24%	24.25%	-2.45%	3.32%	3.33%	0.45%	1.55%	1.30%
III	21.61%	15.50%	1.91%	3.32%	1.58%	-3.65%	4.08%	1.30%
IV	20.92%	14.58%	2.14%	3.32%	1.19%	-4.28%	4.36%	1.30%
V	19.78%	10.52%	4.90%	3.32%	0.21%	-4.64%	3.75%	1.30%
VI	19.37%	12.34%	2.84%	3.32%	0.37%	-4.58%	3.84%	1.30%
VII	18.77%	9.35%	5.12%	3.32%	0.95%	-2.37%	2.07%	1.30%
VIII	16.28%	7.25%	4.94%	3.32%	0.70%	-1.72%	1.15%	1.30%
IX	13.79%	2.99%	6.94%	3.32%	1.78%	-4.22%	4.90%	1.30%
X	11.08%	-0.67%	8.23%	3.32%	4.83%	-11.75%	17.27%	1.30%
Total	15.64%	7.73%	3.89%	3.32%	2.65%	-3.83%	5.38%	1.30%

De entrada, llaman poderosamente la atención los incrementos *aceptables* en el periodo 2010-2012 (ver tabla 4). Cabe recordar que un incremento de 3.32% entre el 2014 y 2015 estuvo entre los primeros indicios de un comportamiento inusual en ese levantamiento. Es claro que la demografía de un país sufre modificaciones al paso del tiempo, pero que dichas modificaciones ocurren de una manera más bien suave. Por ello, se procedió a corregir la demografía de los hogares en la muestra por la vía de la posestratificación y tomando en cuenta resultados de la Encuesta Intercensal 2015.

En estas condiciones, el crecimiento en el número de hogares entre el 2014 y 2015 resultó ser de tan solo 1.3%, lo que exhibe una mayor consistencia con el crecimiento poblacional aceptado como razonable.

Llama también la atención el hecho de que a lo largo del periodo el ingreso promedio por perceptor, global y por deciles de ingreso, exhiba un comportamiento decreciente salvo por excepciones que merecerían ser explicadas, pero que no es el propósito del presente documento. El levantamiento

del 2015 rompió esta tendencia exhibiendo en casi todos los casos crecimientos con respecto al 2014. La base de datos corregida para ese mismo año recupera la mencionada tendencia. Por su parte, el incremento en el número de perceptores por hogar había venido creciendo a lo largo de los levantamientos del periodo, y el levantamiento del 2015 no fue la excepción. Sin embargo, cabe aclarar que ambas bases de datos reportan incrementos por encima de la tendencia.

Por lo que toca a los indicadores de pobreza, la imputación del ingreso *ict_2015_A* atenúa de manera importante los cambios que implicaría el MCS 2015 original, como se muestra en la tabla 5. Si bien la pobreza disminuye, ahora alcanza un porcentaje de 42.4 de la población. Tanto la pobreza moderada como la pobreza extrema muestran un comportamiento similar. En contraste, la población vulnerable por carencias crece menos que lo que

el MCS diría en referencia a la población vulnerable por ingresos: se tienen que este porcentaje, en vez de ser menor que la cifra del 2014 ahora la rebasa por casi uno y medio puntos porcentuales. De esta manera, como consecuencia, la población no pobre y no vulnerable regresa para alcanzar casi el nivel que tenía en el 2014. Para la interpretación de las anteriores aseveraciones, se hace necesario tomar en cuenta que la suma de los anteriores indicadores es igual a 100% y debe seguir siéndolo. Por ello, cada vez que el valor de un indicador crece el de otro debe decrecer.

Las diferencias que muestran los indicadores de carencias sociales para la base de datos original y para el ajuste considerado son, en general, pequeñas. De hecho, la más grande de ellas apenas alcanza ocho décimas de punto porcentual para el indicador de acceso a la seguridad social, cuyos valores se encuentran por encima de 55 por ciento.

Tabla 5

Porcentaje de población por indicadores de pobreza, privación y carencias sociales

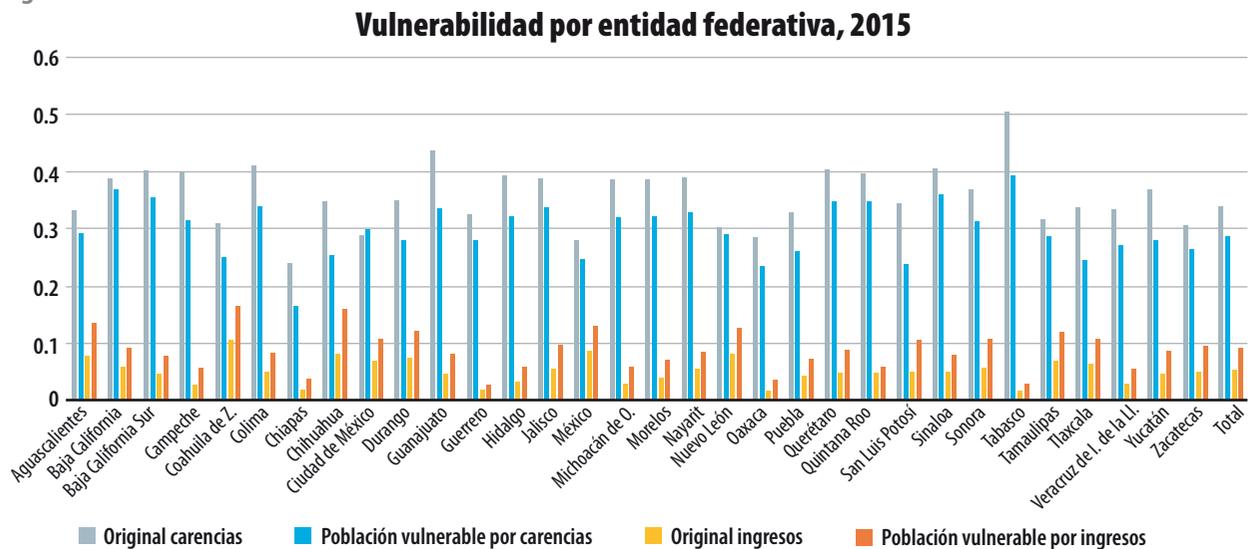
Indicador	2014	2015		Diferencia (ict_2015_A – MCS)
		MCS	ict_2015_A	
Pobreza	46.17	36.69	42.4	5.7
Pobreza moderada	36.62	30.85	34.4	3.6
Pobreza extrema	9.55	5.85	8.0	2.2
Población vulnerable por carencias	26.26	33.94	28.7	-5.2
Población vulnerable por ingresos	7.07	5.45	9.2	3.7
Población no pobre y no vulnerable	20.5	23.91	19.7	-4.3
Privación social				
Población con al menos una carencia	72.43	70.64	71.14	0.5
Población con al menos tres carencias	22.12	20.53	20.95	0.4
Indicadores de carencia social				
Rezago educativo	18.66	17.95	17.68	-0.3
Acceso a servicios de salud	18.16	16.87	17.02	0.1
Acceso a la seguridad social	58.47	56.59	57.37	0.8
Calidad y espacios de la vivienda	12.32	12.01	12.58	0.6
Acceso a servicios básicos de la vivienda	21.22	20.48	20.67	0.2
Acceso a la alimentación	23.35	21.75	22.13	0.4

El desglose a nivel estatal de los indicadores de vulnerabilidad se muestra en la figura 11. Para la casi totalidad de los casos, se tiene que el resultado obtenido para la vulnerabilidad por carencias a partir del Módulo de Condiciones Socioeconómicas original es mayor que el obtenido a partir de la base ajustada. Lo opuesto ocurre para el caso de la vulnerabilidad por ingresos, tal vez como parte del efecto de compensación mencionado antes.

el primer caso, se presentan los residuos para la totalidad de las observaciones en la muestra. Para los residuos del ajuste, se tiene un comportamiento simétrico alrededor del promedio, en este caso igual a 0, pero exhiben colas pesadas particularmente asociadas con valores pequeños del ingreso según fueron declarados durante el levantamiento de la información. Para el caso de los errores de pronóstico, es evidente un sesgo negativo en los valores imputados del ingreso. Para este caso, sin embargo, no son aparentes desviaciones de magnitudes similares a las observadas para el 2014.

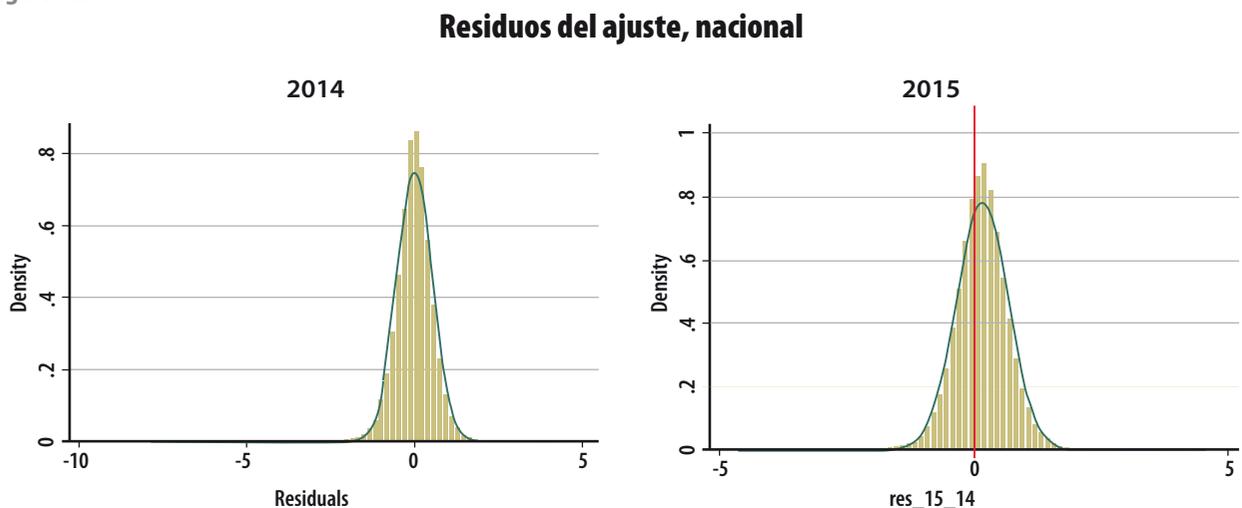
La figura 12 muestra los residuos del ajuste para el 2014 y los errores de pronóstico para el 2015. En

Figura 11



Fuente: cálculos propios a partir del MCS 2015 y de la base de datos ajustada ict_2015_A.

Figura 12



Fuente: cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

5.1 Modelo alternativo

Con el fin de añadir evidencia en favor o en contra del modelo elegido, y siguiendo sugerencias recibidas de quienes escucharon diversas presentaciones del modelo, se llevó a cabo el ajuste de un segundo modelo cuyas características destacables son las siguientes:

1. La variable dependiente para este caso fue el logaritmo del ingreso corriente per cápita de los hogares.
2. Se excluyeron las variables relacionadas con tamaños de diversas subpoblaciones:
 - a. Hombres.
 - b. Perceptores de ingresos.
 - c. Perceptores ocupados.
 - d. Tamaño del hogar.
3. Se incluyeron algunas de ellas en su versión relativa, es decir, como proporciones del tamaño del hogar:
 - a. Proporción de hombres (p_hombres).
 - b. Proporción de integrantes menores de 12 años (p_menores).
 - c. Proporción de integrantes mayores de 11 años y menores de 65 (p_12_65).
 - d. Proporción de integrantes que hablan alguna lengua indígena (p_hab_ind).

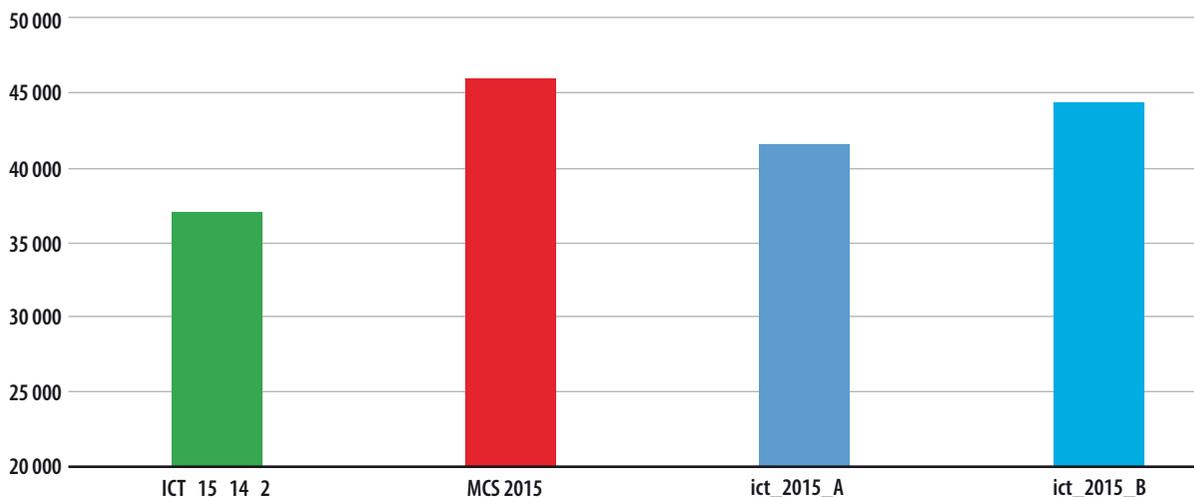
- e. Proporción de integrantes que no concluyeron la educación secundaria (p_rezago).
4. Se usaron nuevamente las dos etapas ya comentadas:
 - a. Modelo lineal.
 - b. Modelo estocástico de simulación.

Además de una reducción de los valores de R^2 para cada uno de los cuatro modelos lineales ajustados para cada estrato sociodemográfico, se obtuvieron los siguientes resultados, que serán a su vez denotados por ict_2015_B. Se incluyen ambos en las figuras 4.b a 6.b con fines de comparación.

La figura 4.b permite una rápida y primera comparación a partir del promedio del ICT de los hogares mexicanos para el 2015. En ella se muestra dicho promedio, en primer lugar, cuando los valores son obtenidos directamente a partir del MCS lineal ajustado; en segundo, cuando los valores proceden de la base de datos original del MCS; en seguida, cuando el modelo estima directamente el ICT del hogar, complementado con una simulación de los residuos; y, finalmente, cuando a los valores procedentes del modelo se le han agregado los de los residuos simulados. Es preciso destacar que

Figura 4.b

ICT promedio por hogar estimado por diferentes modelos, 2015



Fuente: INEGI, MCS 2015 y cálculos propios de las bases de datos del MCS 2014 y 2015.

ahora esta última cantidad se encuentra por arriba de la que correspondería al 2014 cuando sus valores son expresados en pesos del 2015, pero es solo ligeramente menor al que se obtiene de los valores originales. El resultado *ict_2015_A* se encuentra más cerca de lo esperado en ejercicios de esta naturaleza.

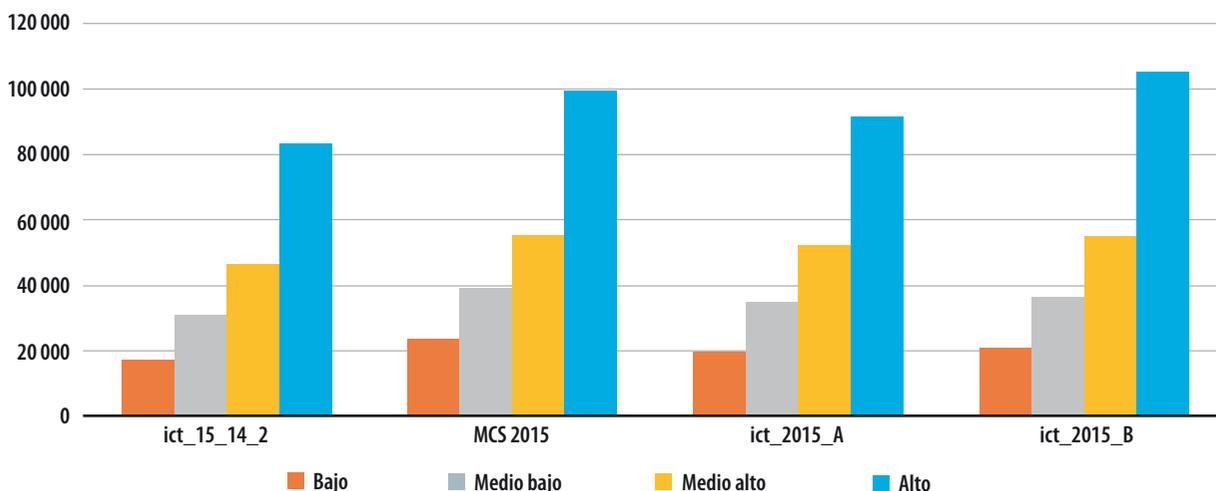
Cuando los promedios nacionales son calculados para cada estrato sociodemográfico, así como para cada uno de los tamaños de localidad, se tienen

comportamientos similares. Cabe destacar que el ICT promedio tanto para el estrato alto como para las localidades urbanas con más de 100 mil habitantes es aún mayor para *ict_2015_B* que lo que resultó para el MCS 2015. Dicha circunstancia se ilustra en las figuras 5.b y 6.b, respectivamente.

La figura 7.b, salvo por alguna excepción, exhibe un comportamiento similar para los promedios del ingreso corriente total cuando se le desagrega a nivel de entidad federativa.

Figura 5.b

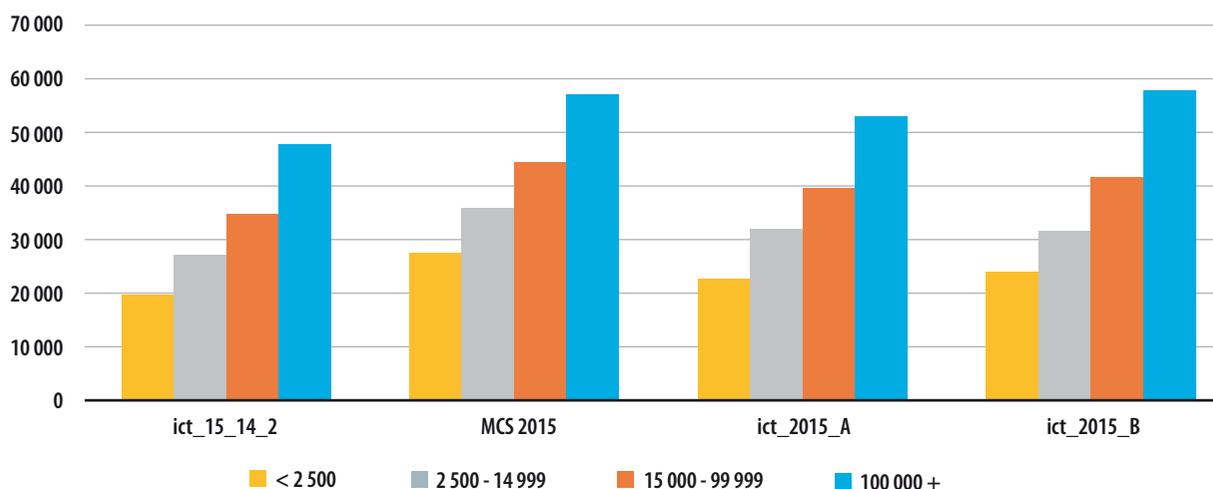
ICT promedio por hogar según estrato sociodemográfico y modelos ajustados, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 6.b

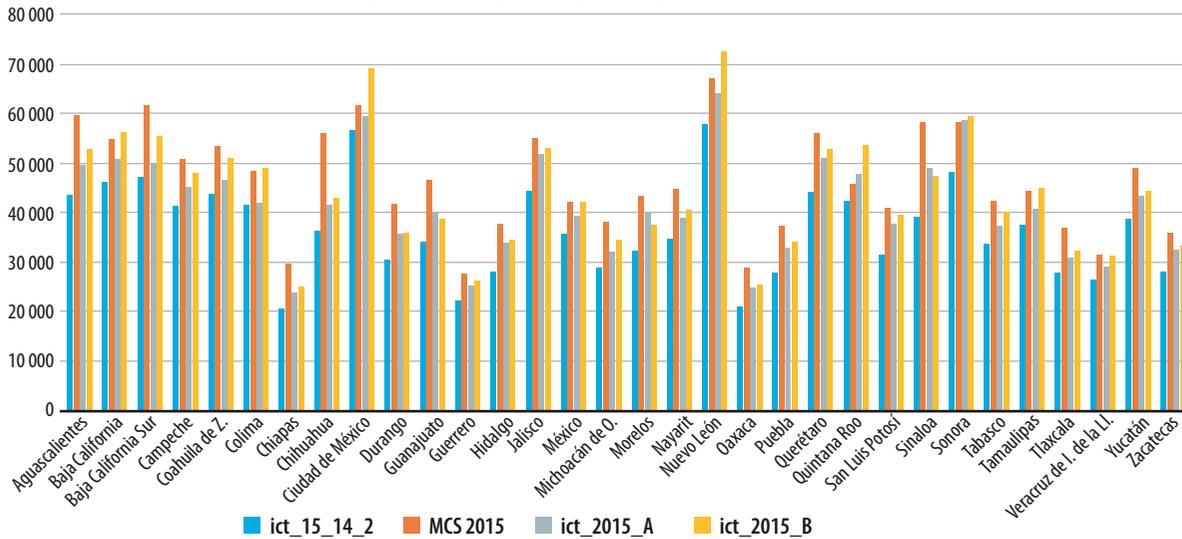
ICT promedio por hogar según tamaño de localidad, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

Figura 7.b

ICT promedio por hogar según entidad, 2015



Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

La tabla 1.b muestra la desagregación por deciles de ingreso para los mismos promedios para diversos años. A nivel de los deciles de hogares se tendrían tanto algunos incrementos en los promedios como también algunos decrementos. Esta circunstancia no parece ser inusual ya que algo semejante se observa, por ejemplo, entre los años 2012 y 2014. Por lo que toca a los coeficientes de Gini, queda claro que el nuevo resultado está con una magnitud muy similar a la de años anteriores aunque con un ligero incremento.

Bajo estas circunstancias, se obtiene la figura 10.b. El incremento en el valor del promedio nacional de los ingresos corrientes totales entre el 2014 y 2015 asciende, ahora, a solo 1.68%, ict_2015_A. Se observan, además, crecimientos moderados entre el 2014 y 2015 para algunos deciles de ingreso, pero en este caso también se exhiben decrementos para los deciles VI a IX. El mayor crecimiento se da en el primer decil, pero éste alcanza *apenas* 10.85 por ciento. Por su parte, el decil de ingresos mayores exhibe un crecimiento de 0.79 por ciento. Al

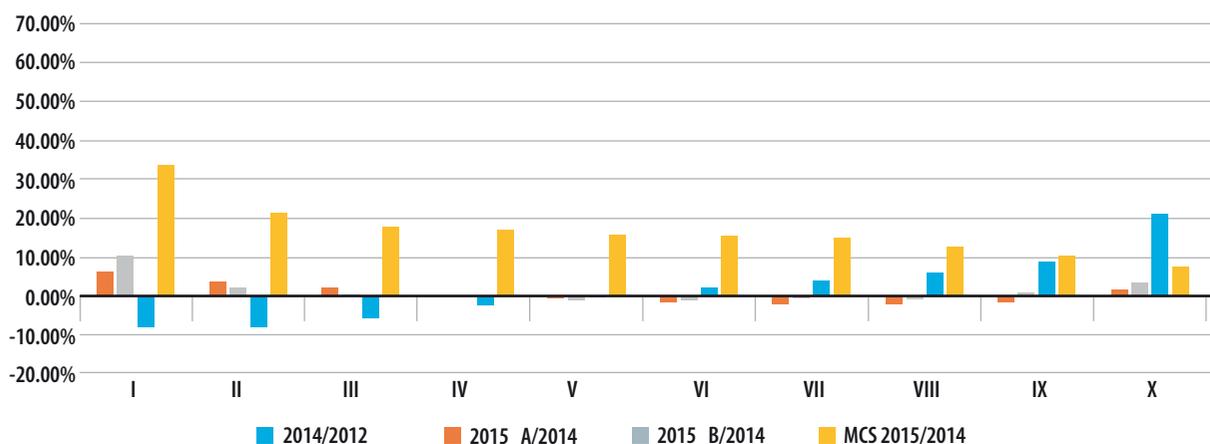
Tabla 1.b

Promedio del ICT por deciles de hogares y su coeficiente de Gini (precios constantes del 2015)

Deciles de hogares	2010	2012	2014	ict_2015_A	ict_2015_B
Ingreso corriente	35 780	40 959	41 000	41 549	44 349
I	5 944	5 757	6 116	6 746	5 637
II	11 386	10 999	11 409	11 638	10 522
III	15 838	15 318	15 622	15 665	14 768
IV	20 303	19 768	19 778	19 758	19 304
V	25 382	24 662	24 564	24 302	24 504
VI	31 511	30 637	30 178	29 901	30 820
VII	39 245	38 264	37 433	37 304	38 969
VIII	50 600	49 108	48 105	47 820	50 957
IX	70 588	67 419	66 510	66 825	72 446
X	156 457	147 662	150 283	155 527	182 210
Coefficiente de Gini	0.464	0.464	0.464	0.468	0.499

Figura 10.b

Crecimiento del ingreso corriente promedio, 2012-2014 y 2014-2015, según modelo



Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

parecer, nuevamente, las comparaciones se están llevando a cabo entre estructuras comparables por construcción.

En contraposición a lo observado en el caso anterior, para el caso de ict_2015_A el comportamiento de estos crecimientos porcentuales no se desvía significativamente del comportamiento exhibido por las mismas medidas en periodos anteriores. No es posible decir lo mismo para el caso ict_2015_B. Ahora podemos decir que en el primer caso no se estaría realizando una comparación indebida, pues están involucradas dos estructuras de ingreso comparables, en particular, en la parte inferior de la distribución. De hecho, una base de datos ajustada como la que se obtiene por esta segunda vía daría lugar a una mayor comparabilidad con todos los levantamientos previos y sin tener que recurrir a modificaciones a la información recopilada en éstos.

6. Conclusión y comentarios finales

6.1 El modelo lineal

Queda claro que la mayor ventaja el modelo utilizado tiene que ver con la facilidad de comunicación, pues se basa en una técnica ampliamente conocida. Otra ventaja, no utilizada del todo en este ejemplo, consiste en la disponibilidad de versiones de la me-

todología cuando los datos provienen de encuestas con diseños complejos. De esta manera, es posible hacer afirmaciones estadísticas más precisas en cuanto a la significancia o no de algún efecto considerado en el modelo. En nuestro caso, dado el enorme tamaño de muestra, no nos preocupaba obtener el modelo más parsimonioso sino uno que diera lugar a un buen pronóstico.

Su mayor desventaja, sin embargo, radica en su incapacidad para explicar un mayor porcentaje de la variabilidad del logaritmo del ingreso corriente total de los hogares. Lo anterior parece estar relacionado con posibles no linealidades en la relación entre el logaritmo de los ingresos corrientes totales y algunas de las variables numéricas incluidas.

6.2 La coyuntura o la distribución de los errores

En esta parte, de la que no parece surgir explicación alguna, se corrige la capacidad predictiva del modelo. Ella parece recoger tanto los aciertos como los errores incurridos durante el levantamiento de la información. Por ello, en alguna medida, permite acercarse a lo que hubiera ocurrido si un levantamiento tuviera lugar bajo las mismas condiciones de alguno otro. De esta manera, la comparabilidad entre diferentes ejercicios del Módulo de Condiciones Socioeconómicas es apoyada.

En conclusión, este enfoque o alguna variante suya tomará en cuenta principalmente los cambios ocurridos en la población objetivo por encima de los meramente operativos durante el levantamiento.

6.3 Otras consideraciones

Los resultados numéricos alcanzados hasta ahora mediante el ajuste de modelos lineales a datos del MCS a lo largo del periodo 2008 a 2015 nos permiten afirmar que este enfoque aporta, en general, estimaciones cuyo valor es menor a lo declarado por los hogares durante el levantamiento del 2015. De hecho, es posible afirmar que el sesgo de pronóstico es significativamente positivo, como se muestra en la tabla 6.

Algo semejante ocurre cuando se consideran los estratos sociodemográficos mostrando, sin embargo, un sesgo mayor para el estrato socioeconómico bajo, mismo que decrece a medida que se avanza en nivel, como se muestra en la tabla 7.

Para el caso de los tamaños de localidad considerados, se tiene algo semejante al caso de los estratos socioeconómicos. Ahora, el sesgo mayor se encuentra en las localidades de menor tamaño y decrece a medida que nos acercamos a los grandes centros urbanos. En ninguno de los casos, sin embargo, dicho sesgo podría ser considerado insignificante (ver tabla 8).

Un aspecto que sin duda llama la atención cuando se hace uso de modelos como los que se han discutido en este trabajo es la ilustración de que, en general, los deciles de ingreso son construcciones propias de cada levantamiento. En consecuencia, su uso con fines de comparación entre uno y otro levantamiento debe ser llevado a cabo con extrema cautela. Es posible ilustrar esta afirmación a partir de dos *deciliza*ciones alternativas para las observaciones del mismo periodo. Por una parte, formamos los de-

ciles a partir de la declaración del ingreso hecha por los informantes durante la recopilación de información. Por la otra, se forman nuevos deciles a partir del ingreso imputado mediante el uso de los modelos ajustados en alguno de los años. Las condiciones materiales que enfrentan los hogares tanto en términos de sus viviendas como de los acervos físicos y humanos acumulados permiten establecer el desplazamiento desde los deciles por declaración hacia los deciles por imputación.

No es posible concluir este documento sin antes afirmar que es necesario invertir un tiempo adicional en la adecuación de los modelos con el fin de que éstos aporten evidencia útil para la toma de decisiones. En general, la comparabilidad entre levantamientos requiere que las condiciones bajo las cuales la información es obtenida se mantengan lo más constantes posible. En la práctica, por supuesto, ello es imposible. De ahí que de cualquier manera se requiera echar mano de métodos y procedimientos que tomen en cuenta y eliminen sesgos involuntarios, como los que aparecen desde el momento en que es seleccionada la muestra en uno y otro levantamiento. Por ejemplo, la muestra no puede ser idéntica pues, en el caso de un cuestionario tan complicado como es el de la ENIGH, la no respuesta se elevaría de manera importante en los segundos levantamientos. De este modo, pueden tenerse diferencias importantes en la demografía de los hogares. Por supuesto, cuando la intención de mejorar la calidad de la información aportada por la encuesta da lugar a modificaciones operativas se hace, con fines de comparabilidad, aún más necesario recurrir a los mencionados métodos y procedimientos.

Con el fin de facilitar la medición de diferentes conceptos de pobreza, los valores de los ingresos por diversas fuentes fueron también imputados a partir del ICT que resulta de la metodología, distribuyéndolo proporcionalmente según lo declarado en cada hogar incluido en la muestra.

Tabla 6

Prueba de significancia para los residuos del pronóstico

	Linearized		
	Mean	Std. Err.	[95% Conf. Interval]
res_15_14	.1648661	.0037588	.1574978 .1722343

Fuente: cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

Tabla 7

Prueba de significancia para los residuos del pronóstico según estrato sociodemográfico

Estrato	Linearized		
	Mean	Std. Err.	[95% Conf. Interval]
Bajo	.2052357	.0115505	.1825938 .2278776
Medio bajo	.1793593	.0049975	.1695629 .1891557
Medio alto	.1203732	.0067004	.1072387 .1335077
Alto	.1138774	.0113066	.0917135 .1360413

Fuente: cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

Tabla 8

Prueba de significancia para los residuos del pronóstico según tamaño de localidad

Tamaño de localidad	Linearized		
	Mean	Std. Err.	[95% Conf. Interval]
1	.1390457	.0045667	.1300938 .1479976
2	.1872345	.0087	.1701802 .2042888
3	.1766315	.0097688	.1574822 .1957808
4	.201488	.0108191	.1802798 .2226962

Fuente: cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

Ajuste estadístico a la distribución del ingreso en el Módulo de Condiciones Socioeconómicas 2015 mediante imputaciones múltiples

Statistical Adjustment to Income Distribution in the MCS 2015 through Multiple Imputations

Delfino Vargas Chanes* y Servando Valdés Cruz**

El presente artículo se relaciona con el tema del ajuste estadístico al ingreso capturado en el Módulo de Condiciones Socioeconómicas (MCS) 2015 con fines de comparabilidad con la serie histórica del ingreso. Esta aproximación parte del supuesto de que la distribución del ingreso para el MCS 2015 está sobreestimada para los deciles I a VI y que los superiores no tienen ese problema; también, se supone que las carencias sociales están bien capturadas. La propuesta consiste en utilizar el método de imputaciones múltiples bajo el supuesto *Missing at Random (MAR)* eliminando los deciles I a VI de la distribución del ingreso de los hogares del 2015. Las bases de datos MCS 2014 y 2015 se unen de forma vertical, con 35 covariables homologadas relacionadas con los ingresos bajos. Primero, se evalúa el modelo de imputación a través de un modelo logístico para validar el supuesto *MAR*. El segundo paso consis-

This article addresses the adjustment of income distribution captured by the Module of Social Conditions of 2015 (MCS 2015) for the purpose of comparability with the historical series of income. This approach is based on the assumption that the income distribution for MCS 2015 is over estimated for deciles I to VI and that the upper deciles do not have that problem. It is also assumed that social lag variables are well captured. The purpose of this article consists of using Multiple Imputations method under *MAR (Missing at Random)* assumption, eliminating the deciles I to VI of the household income distribution of 2015. The MCS 2014 and 2015 databases are vertically merged, with 35 covariables related to low income. First, the imputation model is evaluated through a logistic model to validate the *MAR* assumption. The second step is to perform multiple imputations for the missing income and simulate 10 databases using

* Programa Universitario de Estudios del Desarrollo (PUED) de la Universidad Nacional Autónoma de México (UNAM), dvchanes@unam.mx y dvchanes@gmail.com

** PUED de la UNAM, servando.valdes.cruz@gmail.com

te en realizar imputaciones múltiples para el ingreso faltante y se simulan 10 bases de datos usando técnicas de Monte Carlo que completan el ingreso omitido (los deciles I a VI solo para el 2015) creando 10 versiones completas de datos. El tercer paso consiste en estimar 10 modelos de regresión que se resumen en uno solo (al final del proceso) que incluyen las 35 variables relacionadas con pobreza como predictores del modelo de imputación. Al final, con este modelo se imputan las observaciones faltantes de la cola inferior del ingreso para el 2015. A partir de la base de datos imputada se aplica la metodología del CONEVAL para estimar la pobreza. Los resultados indican que la pobreza a nivel nacional es de 42.4%, que representan a 51.4 millones de personas, una pobreza extrema de 8.4% en 10.2 millones de personas y un coeficiente de Gini de 0.492.

Palabras clave: imputación múltiple; pobreza; desigualdad; ajuste distribución-ingreso.

Monte Carlo techniques that complete omitted income (deciles I to VI only for 2015) by creating 10 complete versions of data. The third step consists of estimating 10 regression models that are summarized in a single model (at the end of the process) that includes the variables related to poverty, as predictors of the imputation model. Finally, with this model, the missing observations of the lower income distribution tail for the year 2015 are imputed. From the imputed database, the CONEVAL methodology is applied to estimate poverty. The results indicate that poverty at the national level is 42.4%, representing 51.4 million people, extreme poverty of 8.4% representing 10.2 million people and a Gini coefficient of 0.492.

Key words: multiple imputation; poverty; inequality; income-distribution adjustment.



Stock Market Broker Looking At Graphs On Multiple Screens /Andrey Popov/istock/Getty Images

Introducción

La estimación de la pobreza y la desigualdad han sido un tema de interés nacional por sus implicaciones en la política pública, ya que son elementos sustanciales para la asignación de presupuesto a los programas y acciones federales de desarrollo social. El uso de las encuestas que miden el gasto y el ingreso de los hogares hace posible, para otras instituciones, focalizar los programas sociales para incrementar la eficiencia de las acciones gubernamentales dirigidas a combatir la pobreza y el desarrollo social.

El Instituto Nacional de Estadística y Geografía (INEGI) es el encargado de realizar los operativos tanto del Módulo de Condiciones Socioeconómicas (MCS) y de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), mientras que el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) tiene como una de sus tareas procesar los resultados de esos operativos para medir la pobreza multidimensional.

El enfoque que se propone en el presente artículo utiliza un método estadístico llamado imputaciones múltiples, que permite la comparabilidad del ingreso del 2015 con años anteriores (2010 al 2014);¹ sin embargo, parte de supuestos (como todo modelo estadístico). El primero es que la tendencia general de las carencias sociales mantiene su comportamiento tendencial, es decir, que las carencias sociales observadas en el 2010, 2012, 2014 y 2015 han sido medidas de forma correcta y constituyen uno de los soportes fundamentales para estimar el ingreso del 2015. El segundo es que no todos los deciles de ingreso tienen que ser estimados, ya que los cambios observados en el MCS 2015 solo afectaron en los deciles inferiores de ingreso y, en menor medida, a los ingresos de los deciles superiores. De esta manera, el ingreso imputado para el MCS 2015 permite la com-

parabilidad del ingreso con la serie histórica del 2010 al 2015 y, en consecuencia, con la pobreza multidimensional.

El artículo incluye seis secciones. En la primera se muestra la tendencia general de la pobreza y justifica de alguna manera la necesidad de hacer comparables los resultados sobre el tema con las encuestas anteriores. En la segunda se justifican parte de los supuestos de la imputación al llamar la atención sobre la serie histórica de las carencias y su relación con la pobreza. En la tercera se explica el procedimiento de imputación múltiple. En la cuarta se presentan otros enfoques usados para calibrar el modelo. La quinta se dedica a presentar los resultados de la imputación múltiple. En la última se dan las conclusiones.

Tendencia de la pobreza multidimensional y las carencias sociales

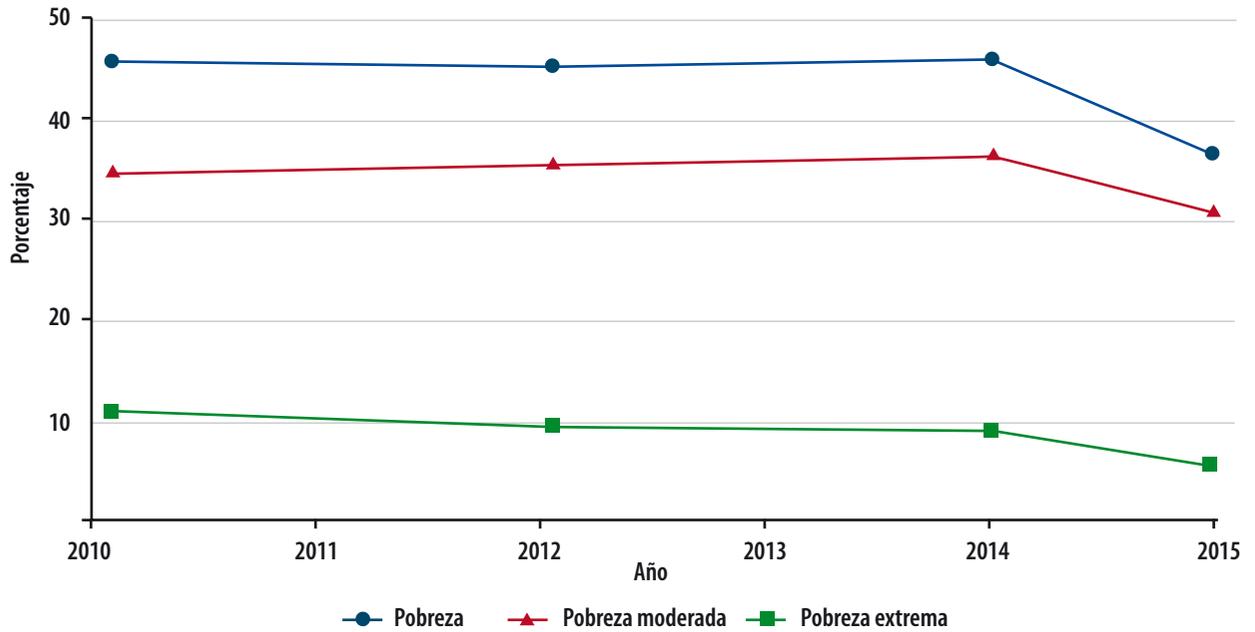
El CONEVAL utiliza el soporte normativo que le da la *Ley General de Desarrollo Social (LGDS)*, la cual utiliza dos enfoques: el de bienestar —que incluye los ingresos— y el de los derechos humanos —que incorpora las carencias sociales— (CONEVAL, 2011). La figura 1 muestra la evolución de la pobreza en las mediciones bianuales de pobreza multidimensional del 2010 al 2015. En esta figura, los datos del 2015 no fueron calibrados y rompen la serie histórica que no permiten la comparabilidad con años anteriores.

Por otro lado, la figura 2 muestra las seis carencias sociales del 2010 al 2015 (e. g., rezago educativo; acceso a servicios de salud, a los servicios básicos de la vivienda y a la alimentación; seguridad social; así como calidad y espacios en la vivienda). En general, se observa en esta figura una reducción de éstas, con excepción de las que se refieren a la alimentación y a los servicios básicos de la vivienda. Observamos consistencia en la medición de ellas, que dan muestra de una medición consistente en el tiempo observado y constituye uno de los supuestos del modelo de imputación.

¹ Consideramos pertinente usar el ingreso de los hogares y no el ingreso per cápita al suponer que el primero es el que se vio más inflado durante el levantamiento del MCS 2015 para los deciles inferiores. Todos los ingresos en este ejercicio se deflactaron al 2015.

Figura 1

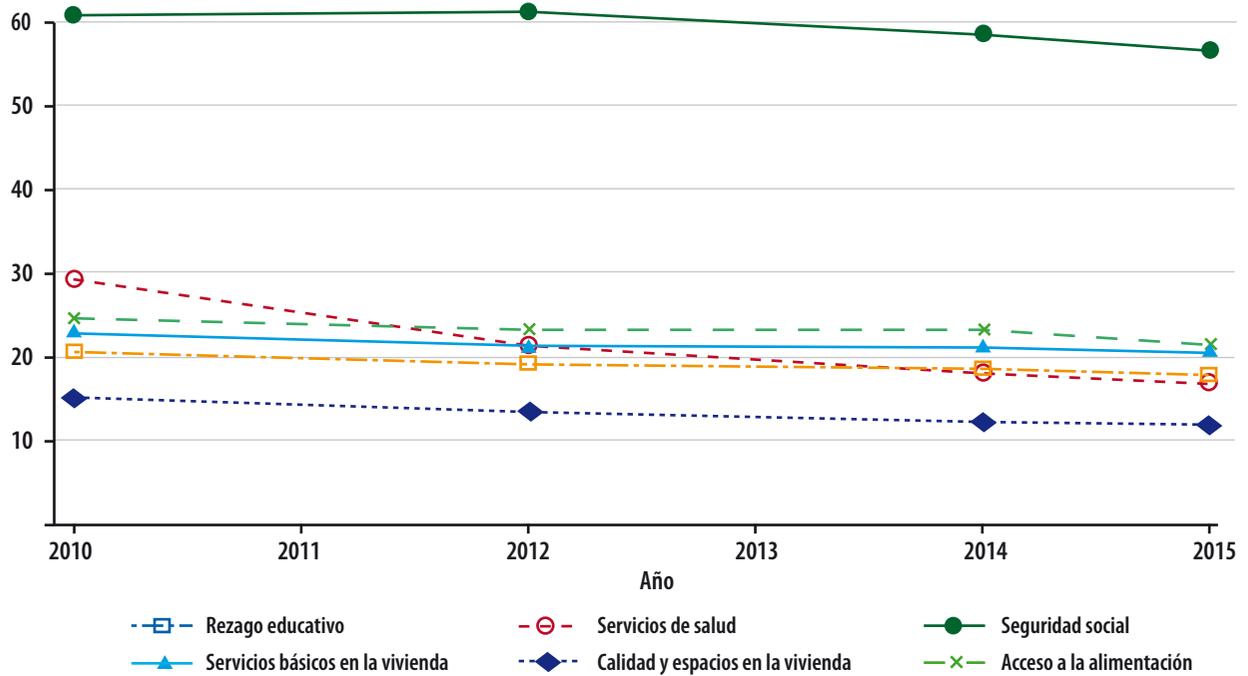
Evolución de la pobreza en México con ingreso inferior a la línea de bienestar y línea de bienestar mínimo, 2010-2015



Fuente: elaboración propia con datos del CONEVAL.

Figura 2

Evolución de la población con carencias en México, 2010-2015



Fuente: elaboración propia con datos del CONEVAL.

Deciles de ingreso

Otro de los supuestos usados para el presente ejercicio es que los deciles inferiores se vieron más afectados que los superiores. El cuadro 1 muestra la variación porcentual del ingreso promedio de los deciles de hogares. Por ejemplo, la tasa anual del 2010-2012 es -0.1%, muy similar a la del 2012-2014 que es de -0.9%; sin embargo, la tasa anual del 2014-2015 muestra un incremento de 14.8%; cuando se imputan los datos, la tasa anual es de 9.6 por ciento. Un razonamiento similar se puede hacer para cada decil, de manera que los deciles más afectados son del I al VI, donde se observan los mayores incrementos.

Procedimiento de imputación

Para reconstruir la comparabilidad de la serie histórica del ingreso, los datos se organizan de manera transversal usando los MCS 2014 y 2015; la base de

datos (BD) contiene las variables x_1 a x_{35}^2 usadas para el modelo de imputación (ver el Anexo 1). La variable de interés para realizar las correcciones es el ingreso del hogar deflactado al 2015.³ Las BD se codifican de manera homogénea para que las variables tengan el mismo sentido en los dos años; también, algunas de éstas contienen datos faltantes, pero el énfasis es imputar el ingreso de los hogares del 2015.

La habilidad para imputar datos faltantes depende del conocimiento que se tenga acerca del mecanismo que los produce. En la literatura se reconocen tres mecanismos: completamente al azar (*MCAR*, por sus siglas en inglés), al azar (*MAR*, por sus siglas en inglés) y no ignorable (*NI*); para más información, ver Rubin (1987). El primero supone

- 2 Las covariables incluyen aspectos sociodemográficos (sexo, edad, ocupación, escolaridad, jubilado), así como variables relacionadas con la pobreza (e. g., habla lengua indígena, carencia, rezagos sociales varios, tamaño del hogar, perceptores en el hogar, ocupados en el hogar, remesas, recibe ayuda de programas sociales).
- 3 La deflactación se hizo utilizando el índice nacional de precios al consumidor (INPC) de agosto de cada año; esta forma es la usada por el CONEVAL, que es congruente con lo propuesto en la metodología de medición multidimensional de la pobreza.

Cuadro 1

Variaciones porcentuales del ingreso promedio de los deciles de hogares en México, 2010-2015

Deciles de hogares	2010-2012	2012-2014	2014-2015 ^a	2014-2015 (I-VI) ^b
Ingreso corriente total	-0.1%	-0.9%	14.8%	9.6%
I	1.9%	9.4%	44.8%	32.9%
II	1.0%	3.4%	27.8%	-9.0%
III	0.8%	0.8%	23.7%	-10.1%
IV	1.5%	-1.4%	22.3%	-4.5%
V	1.3%	-1.7%	20.6%	3.4%
VI	2.0%	-2.8%	19.9%	12.8%
VII	1.3%	-2.9%	18.8%	18.8%
VIII	0.2%	-2.4%	15.9%	15.9%
IX	-1.5%	-2.0%	13.6%	13.6%
X	-1.0%	0.4%	8.1%	8.1%

^a Tasa de cambio del 2014 al 2015, datos sin corregir.

^b Tasa de cambio del 2014 al 2015, datos imputados en los deciles I-VI.

Fuente: estimaciones del CONEVAL con base en las ENIGH-MCS 2008, 2010, 2012, 2014 y 2015.

que éstos no se relacionan con ninguna covariable que explique la razón de su ausencia; es muy raro en la práctica y se puede inducir cuando el investigador asigna al azar los ítems faltantes; la imputación de datos bajo este esquema es muy sencilla. El segundo implica que existen covariables identificadas que explican la ausencia de los datos en la muestra; para ello, se requiere identificar un modelo de imputación que contiene variables que explican este mecanismo (Schafer y Olsen, 1998). El tercero supone que los datos faltantes se explican por un mecanismo endógeno, es decir, no hay covariables que expliquen la ausencia, que el incremento/decremento de la variable de interés incrementa/decrementa la probabilidad de estar ausente de la base de datos; este mecanismo es el más difícil de tratar de forma estadística. Por fortuna, los mecanismos *MCAR* y *NI* ocurren con menor frecuencia y el *MAR*, aparece en 90% de los casos (Enders, 2010).

Formalmente hablando, supongamos que una matriz de datos incluye los valores observados y los faltantes, y se denota por $Y = (Y_{obs}, Y_{falt})$, donde Y_{obs} se refiere a los observados y Y_{falt} es la matriz de los faltantes. La función de densidad de probabilidad se denota por:

$$p(Y|\theta) = \prod_{i=1}^n p(Y_i|\theta), \quad (1)$$

donde θ indica los parámetros que determinan la distribución de Y . Para entender mejor los mecanismos de datos faltantes, suponga que R es una matriz de indicadores con las mismas dimensiones de la matriz de datos original, donde cada elemento de R contiene 0 si el elemento es faltante y 1 si está presente. La función de densidad condicional se denota por:

$$p(Y, R|\theta, \phi) = p(Y|\theta) p(R|Y, \phi), \quad (2)$$

donde ϕ es la distribución condicional de R dado el conjunto de datos completos Y . En esta expresión, los datos completos observados se reemplazan e integran sobre la porción faltante y se expresan de la siguiente manera:

$$p(Y_{obs}, R|\theta, \phi) = \int p(Y_{obs}, Y_{falt}|\theta) p(R|Y_{obs}, Y_{falt}, \phi) dY_{falt}. \quad (3)$$

El mecanismo *MCAR* no depende de los valores observados ni de los faltantes, entonces se define como:

$$MCAR: p(R|Y_{obs}, Y_{falt}, \phi) = p(R|\phi), \quad (4)$$

Este mecanismo es equivalente a borrar al azar un conjunto de datos a partir de una matriz Y completa, donde cada observación tiene igual probabilidad de ser borrada.

El segundo mecanismo, el *MAR*, formalmente quiere decir que la distribución de los valores faltantes depende solo de los valores observados. Entonces:

$$MAR: p(R|Y_{obs}, Y_{falt}, \phi) = p(R|Y_{obs}, \phi). \quad (5)$$

En otras palabras, el mecanismo de los valores faltantes se puede encontrar en los valores observados.

Por último, el mecanismo *NI* se puede encontrar en los valores faltantes y no en los observados, por lo tanto:

$$NI: p(R|Y_{obs}, Y_{falt}, \phi) = p(R|Y_{falt}, \phi). \quad (6)$$

Los modelos de imputación y el analítico

Cuando se realizan imputaciones es importante distinguir entre ambos modelos. El de imputación es el que tiene covariables que explican la probabilidad de que un dato sea faltante; de esta manera, resulta muy útil conocer todas las covariables que explican el proceso de imputación. En tanto, el modelo analítico es el que consiste en analizar los valores imputados. De manera típica, bajo el modelo *MCAR*, los de imputación y el analítico son los mismos o casi iguales. Bajo el supuesto *MAR*, la calidad

de la imputación está asociada con selección de las variables del modelo de imputación (Graham y Schafer, 1999; Russell, Stern y Sinharay, 2000); por ejemplo, Graham y Schafer muestran que los estimadores de los parámetros tienen menor sesgo en relación con los parámetros poblacionales cuando se incrementa el número de covariables en el modelo de imputación. Por lo tanto, la estrategia que usamos en el documento consiste en incorporar un número considerable de covariables para mejorar la calidad de las imputaciones que estén relacionadas con el ingreso. En nuestro caso, utilizamos de preferencia covariables relacionadas con la población con ingresos bajos, pues partimos del supuesto de que los valores faltantes son los de la cola inferior del ingreso. El modelo de imputación es logístico, donde el valor de 1 se asigna para los valores faltantes y 0 cuando son observados (ver el Anexo 1, donde se muestran las covariables).

Imputación múltiple (IM)

Incorpora un proceso de simulación para completar los valores faltantes dado que un solo valor no refleja la variabilidad, la cual es el resultado del proceso de simulación de datos que resultan de un proceso de Monte Carlo; de esta manera, se obtienen varias versiones completas de datos y, después, se ajusta un modelo analítico para cada conjunto. El modelo analítico es el resumen de todos los modelos ajustados en uno solo, para el cual se aplican las reglas de Rubin (1987:76-77).

Uno de los métodos de Monte Carlo usado en IM consiste en aumentar los datos faltantes, que contiene el paso-I y el -P. El -I obtiene una muestra aleatoria de las observaciones a partir de la distribución marginal en la primera iteración:

$$Y_{falt}^{(t+1)} \sim p(Y_{falt} | Y_{obs}, \theta^{(t)}).$$

El paso-P obtiene una muestra aleatoria de parámetros de la distribución marginal que incorpora los valores observados e iniciales de los valores faltantes del -I en la primera iteración:

$$\theta^{(t+1)} \sim p(\theta | Y_{obs}, Y_{falt}^{(t+2)}).$$

El paso-I y el -P de la primera iteración proporcionan los valores iniciales $\{Y_{falt}^{(0)}, \theta^{(0)}\}$, así como las iteraciones posteriores para crear una Cadena de Markov con valores $\{Y_{falt}^{(1)}, \theta^{(1)}, Y_{falt}^{(2)}, \theta^{(2)}, \dots\}$ que convergen en distribución a $p(\theta, Y_{falt} | Y_{obs})$, con la finalidad de crear una cadena de observaciones completa $\{Y_{falt}^{(1)}, Y_{falt}^{(2)}, \dots, Y_{falt}^{(m)}\}$, lo cual es equivalente a correr m cadenas independientes o iteraciones de calentamiento, de tamaño t .

Un aspecto importante a considerar es probar la convergencia del proceso de cadenas múltiples de Monte Carlo (MCMC), así como el número de iteraciones que se requieren para que el proceso converja; estos aspectos dependen de la cantidad de datos faltantes. Sin embargo, el algoritmo requiere de un mínimo de iteraciones de calentamiento m para garantizar que una cadena $\theta^{(t+m)}$ sea independiente de $\theta^{(t)}$. Una vez que se cumpla esta condición, cada valor del parámetro estimado θ puede tomarse como una extracción independiente de $p(\theta | Y_{obs})$ y, en consecuencia, Y_{falt} se puede usar como valor imputado.

Se han propuesto varios métodos para investigar la convergencia de la distribución conjunta para un valor específico de m (Ritter y Tanner, 1992; Roberts, 1992). Desde un punto de vista práctico, la función de autocorrelación (ACF) se puede usar para determinar la convergencia del algoritmo para cada rezago- p de la serie estacional $\{k^{(t)}; t=1, 2, \dots, k\}$; la ACF se define como:

$$\rho_p = \frac{Cov(k^{(t)}, k^{(t+p)})}{V(k^{(t)})}.$$

La gráfica de ACF vs. p para un valor p finito muestra un *correlograma* que ayuda a identificar la dependencia lineal potencial de las iteraciones. Si muestra un decaimiento *súbito* para los valores de $p = 2$ a 4, sugieren una independencia serial, lo cual quiere decir que el algoritmo converge a una solución satisfactoria (Box y Jenkins, 1976). Existen otros

métodos para asegurar la convergencia que usan un enfoque Bayesiano que también se pueden considerar como criterios de convergencia, que no se usan en el presente trabajo (Casella y Berger, 2001; Gilks, Richardson y Spiegelhalter, 1996).

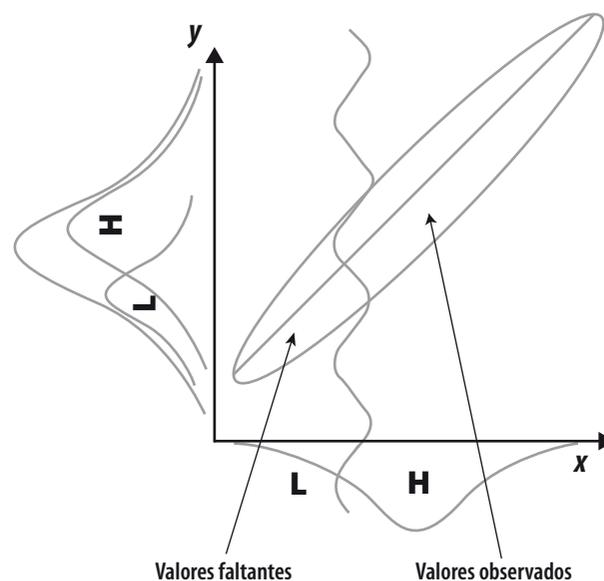
Otros métodos de imputación

Con la finalidad de calibrar el modelo de regresión obtenido a partir de la imputación múltiple se probaron dos enfoques más. El primero fue el de la imputación vía regresión, el cual completa los datos faltantes usando un enfoque Gaussiano (Gelman, Carlin, Stern y Rubin, 2004; Rubin, 1987; Shenker y Taylor, 1996). Este método parte de la idea de que los datos están completos en las covariables⁴ excepto en la variable de respuesta, que en nuestro caso es el ingreso. Para este ejercicio, también se omitieron los ingresos de los deciles I a VI del MCS 2015 del ingreso que se busca imputar.

La figura 3 muestra la distribución de valores observados y se encuentra en la parte *H* y los valores faltantes, en la *L*. El modelo de regresión estimado contiene las 35 variables previamente definidas (ver variables en el Anexo 1). El modelo estimado resulta ser muy similar (con diferencias en las milésimas) al ajustado vía regresión por imputación múltiple. Al estimar la incidencia de la pobreza usando la metodología del CONEVAL se obtienen prácticamente los mismos resultados.

El segundo enfoque fue el de imputación vía máxima verosimilitud con información completa (FIML, por sus siglas en inglés). Se probó ajustar otro método de estimación del modelo de regresión, bajo el supuesto *MAR*, usando el método de FIML (Arbuckle, 1996; Muthén, Kaplan y Hollis, 1987), el cual se basa en el trabajo de Little & Rubin (1987) y utiliza el supuesto de que el mecanismo de ignorabilidad es *MAR*; es decir, que los valores faltantes se pueden recuperar a partir de los datos observables. En realidad, el método FIML no pro-

Figura 3
Imputaciones vía regresión bajo el supuesto *MAR*



$$y_i = \alpha + \beta x_i + \zeta_i,$$

$$E(\zeta) = 0, V(\zeta) = \sigma_\zeta^2$$

$$E(x) = \mu_x, V(x) = \sigma_x^2$$

duce imputaciones, sino que utiliza la información disponible para obtener estimaciones de máxima verosimilitud para el modelo de regresión propuesto. Este enfoque genera una función de máxima verosimilitud basada en una matriz de covarianzas de todos los posibles subconjuntos de variables donde la información es completa. Mediante FIML se pueden usar diferentes patrones de datos incompletos. La BD es la misma que en los otros métodos (ver figura 4), donde igualmente se elimina el logaritmo del ingreso de los deciles I a VI. Se ajusta el modelo $\hat{Y} = X\hat{\beta}$, donde \hat{Y} es el valor estimado y $\hat{\beta}$ es el vector de parámetros estimados, y se procede a estimar los valores faltantes de los deciles I a VI. Como último paso, se calcula la pobreza usando la metodología del CONEVAL. Los resultados son los mismos que con los otros métodos, por lo tanto, enseguida se reportan los resultados obtenidos vía imputaciones múltiples.

⁴ La base de datos con las covariables completas se tomaron del modelo de imputaciones múltiples, omitiendo los deciles I a VI del logaritmo del ingreso.

Resultados

El procedimiento

La base de datos que usaremos es de los MCS 2015 y 2014, como lo hemos establecido en párrafos anteriores; el modelo de imputación utiliza las variables que se indican en el Anexo 1, las cuales se relacionan con la pobreza, ya que el modelo busca predecir el ingreso de los deciles con menores ingresos.

El primer paso consiste en ajustar un modelo de imputación para justificar el supuesto de *MAR* (Rubin, 1987; Schafer, 1997); para ello, se ajusta un modelo logístico.⁵ La variable de respuesta es $Y^*=1$ si el valor es faltante y $Y^*=0$ si está presente; se usaron las 35 covariables presentes en la BD, donde Y^* es el *logit*:

$$\text{logit}(Y^*) = \ln\left(\frac{Y^*}{1-Y^*}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

El modelo de imputación muestra un buen ajuste, ya que tiene 87.3% de casos correctamente clasificados con una sensibilidad de 81% y especificidad de 90.9 por ciento. Los coeficientes de las covariables fueron todos significativos, con excepción de cinco (ver el Anexo 2).

El modelo analítico es un modelo de regresión y tiene la siguiente forma:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_i,$$

donde $Y_i = \ln(\text{ingreso del hogar} + k)$ es el logaritmo del ingreso corriente total mensual del hogar. Las variables predictoras se muestran en el Anexo 1. Se busca la mejor transformación del ingreso mediante el $\log(y + k)$ y encontramos que el valor de

⁵ El valor 1 se asigna a los valores faltantes, que son el ingreso del hogar de los deciles I a VI y 0 si está presente; se usan 35 covariables que explican los valores faltantes (eliminados) en los deciles inferiores de la distribución del ingreso de hogares. El ajuste del modelo es satisfactorio y el supuesto de *MAR* se satisface.

la constante que minimiza el sesgo y corrige la *curtosis* es $k = 503.57$. Con dicha transformación se disminuye la asimetría y la *curtosis*, aun cuando el estadístico D de Lilliefors rechaza la normalidad ($p < 0.001$); no obstante, la asimetría y la *curtosis* son casi nulas (ver cuadro 2).

La base de datos cuenta con 35 variables homologadas para el 2014 y 2015; usando el ingreso por hogar se ajustan tres escenarios: en el primero se eliminan los ingresos de los deciles I al IV; en el segundo, los del I al V; y el tercero, los del I al VI para el MCS 2015. Ya que suponemos que el ingreso del hogar está sobreestimado en estos deciles, al ajustarlos nos permitirá que la serie sea comparable con las BD anteriores (ver figura 4).

En segundo lugar se aplica el método de imputaciones múltiples (Rubin, 1987; Schafer y Graham, 2002; Vargas Chanes y Lorenz, 2015) que usa métodos de Monte Carlo, donde se simulan 10 conjuntos completos que contienen los datos imputados.⁶ La garantía de que hay convergencia en las simulaciones y en la parte de imputación de los da-

⁶ Se han elegido $m=10$ imputaciones que garantizan una eficiencia relativa de 97 por ciento. La eficiencia relativa se calcula con la fórmula $EF = (1 - \gamma/m)^{-1}$, donde $\gamma = 0.30$ y $m=10$; es decir, se tiene 30% de información del ingreso faltante de los deciles I al VI al unir las dos bases de datos del 2014 y 2015.

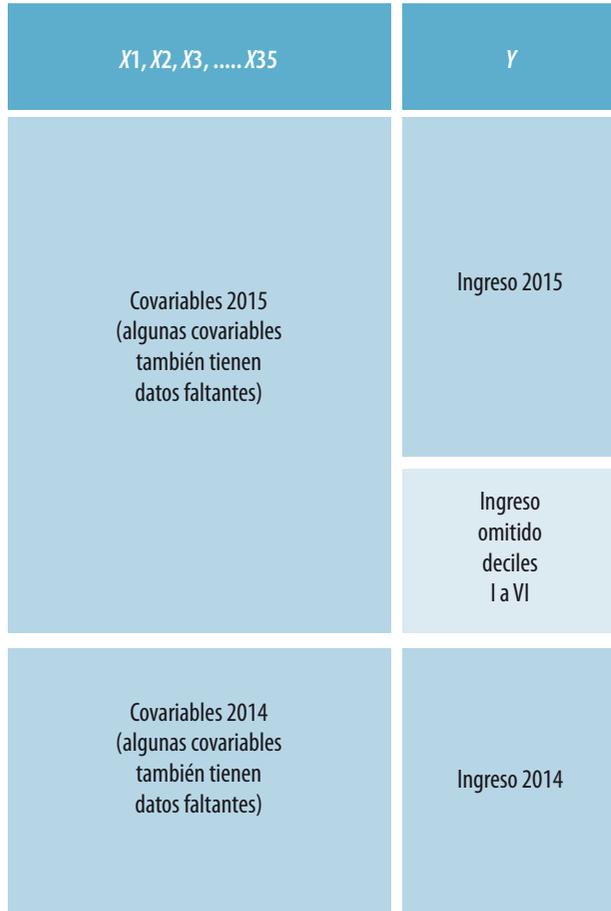
Cuadro 2

Prueba de normalidad de Lilliefors (Kolmogorov-Smirnov) y medidas de dispersión

Variable	Estadístico D	Valor de p	Asimetría	Curtosis
Ingreso sin transformar (ing2015)	0.275	<0.001	48.17	5 606.83
Log(Ingreso+k) (lning2015)	0.010	<0.001	0	0.58

Figura 4

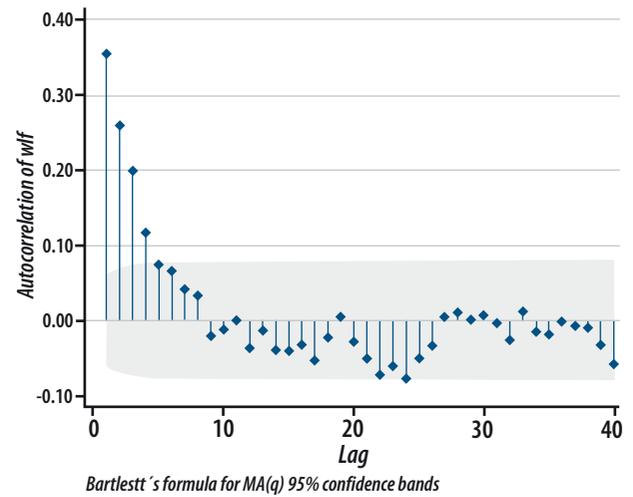
Esquema de preparación de archivos que incorporan los MCS 2014 y 2015



Nota: para el 2015, se omite el ingreso en los deciles I a VI.

Figura 5b

Función de autocorrelación de la WLF



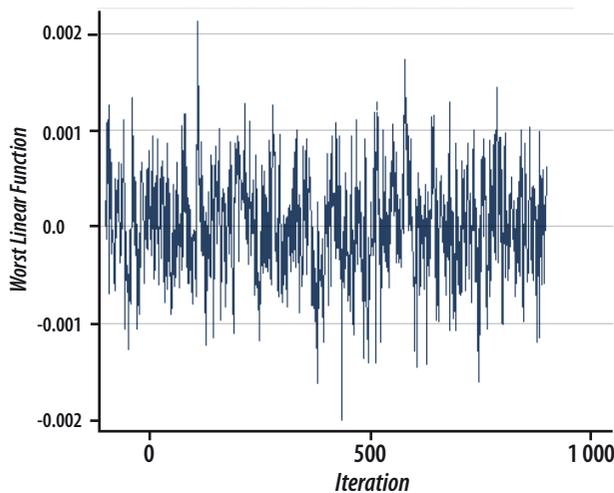
tos se juzga a partir de la peor función lineal (WLF, por sus siglas en inglés) y de la función de autocorrelación (ACF) de la WLF; por ejemplo, si la WLF (figura 5a) muestra *ruido blanco* y la ACF (figura 5b) presenta una caída de las autocorrelaciones para $p > 4$, entonces hay cierta garantía de que el proceso converge de manera satisfactoria.⁷

El siguiente paso consiste en ajustar un modelo de regresión para los 10 conjuntos de datos completos y se consolida en un modelo final usando las reglas de Rubin (1987:76-77). Este modelo usa las 35 variables previamente mencionadas (ver Anexo 3⁸) y permite estimar los valores del ingreso faltante para los tres escenarios plantados (e. g, omitiendo los deciles I-IV, I-V y I-VI del MCS 2015).

En suma, el ingreso se estima a través del modelo ajustado vía imputaciones múltiples, $\hat{Y} = X\hat{\beta}$, donde \hat{Y} es el valor estimado y $\hat{\beta}$ el vector de parámetros resumidos mediante las reglas de Rubin para los tres escenarios al imputar los deciles I-IV, I-V y I-VI. Se obtiene el anti-*log* de los ingresos estimados y se procede a calcular la pobreza empleando la metodología del CONEVAL; los ingresos mensuales de los hogares se calculan per cápita, con lo que se tie-

Figura 5a

Peor función lineal (WLF)



7 El número de iteraciones de calentamiento (*burn in iterations*) se fija en mil para este ejercicio.

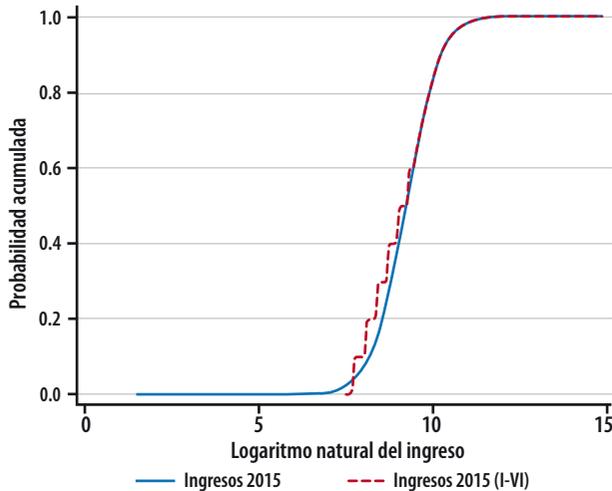
8 En este anexo se muestra el modelo ajustado para el escenario que omite los deciles I a VI.

nen los resultados que a continuación se muestran en el siguiente subapartado de esta sección. La figura 6 presenta la función de distribución acumulada

lada (FDA) del ingreso mensual por hogar para el MCS 2015 en sus dos versiones, la FDA sin corregir en línea continua y los deciles I al VI imputados en línea punteada.

Figura 6

Función de distribución acumulada del ingreso mensual por hogar para el MCS 2015 (sin corregir) y con datos imputados para el escenario de los deciles I-VI



Fuente: elaboración propia usando el MCS 2015.

Ingreso corriente y pobreza multidimensional

Los ingresos medios por deciles para los años 2010 al 2015 (originales e imputados) a precios constantes del 2015 se muestran en el cuadro 3. Se observa que en la última columna de ingresos medios totales por hogar (imputados para el 2015, deciles I-VI) tienen un incremento de 9.6% en relación con el 2014 y el primer decil tiene 32.9% de incremento en el ingreso. Los deciles II al IV muestran pérdidas de ingreso de 4 a 10%; el resto de los deciles superiores presentan incrementos en el ingreso de 8 a 18 por ciento.

Los resultados de la pobreza multidimensional del 2010 al 2015 se muestran en el cuadro 4; se presentan dos versiones del MCS 2015: la reportada por el INEGI en su primera versión y la corregi-

Cuadro 3

Ingreso promedio corriente total por deciles, Estados Unidos Mexicanos, 2010-2015^a

	2010	2012	2014	2015 ^b	2015 (I-IV) ^c	2015 (I-V) ^d	2015 (I-VI) ^e
Ingreso medio corriente total	13 111	13 098	12 983	14 909	14 527	14 340	14 225
I	1 512	1 541	1 686	2 441	2 408	2 294	2 240
II	3 336	3 368	3 484	4 452	3 385	3 240	3 171
III	4 792	4 830	4 868	6 023	4 645	4 466	4 378
IV	6 246	6 340	6 254	7 648	6 305	6 085	5 974
V	7 856	7 962	7 828	9 444	9 444	8 233	8 093
VI	9 753	9 952	9 671	11 600	11 600	11 596	10 912
VII	12 222	12 385	12 031	14 292	14 292	14 292	14 291
VIII	15 856	15 883	15 506	17 971	17 971	17 971	17 971
IX	21 984	21 659	21 229	24 107	24 107	24 107	24 107
X	47 553	47 069	47 279	51 123	51 123	51 123	51 123

^a Calculado a pesos constantes del 2015.

^b Valores observados no comparables.

^c Se imputan los deciles I-IV.

^d Se imputan los deciles I-V.

^e Se imputan los deciles I-VI.

Fuente: estimaciones propias con base en las ENIGH-MCS 2010, 2012, 2014 y 2015.

Cuadro 4

Medición de la pobreza, población general, Estados Unidos Mexicanos, 2010-2015

Pobreza multidimensional	Porcentaje						Millones de personas					
	2010	2012	2014	2015 ^a	2015 (I-IV) ^b	2015 (I-VI) ^d	2010	2012	2014	2015 ^a	2015 (I-IV) ^b	2015 (I-VI) ^d
Pobreza												
Población en situación de pobreza	46.1	45.5	46.2	36.7	40.0	41.7	52.8	53.3	55.3	44.5	48.6	51.4
Población en situación de pobreza moderada	34.8	35.7	36.6	30.8	32.3	33.5	39.8	41.8	43.9	37.4	39.2	41.2
Población en situación de pobreza extrema	11.3	9.8	9.5	5.9	7.7	8.2	13.0	11.5	11.4	7.1	9.4	10.2
Población vulnerable por carencias sociales	28.1	28.6	26.3	33.9	33.9	33.9	32.1	33.5	31.5	41.2	41.2	41.2
Población vulnerable por ingresos	5.9	6.2	7.1	5.5	6.2	6.7	6.7	7.2	8.5	6.6	7.5	8.4
Población no pobre y no vulnerable	19.9	19.8	20.5	23.9	23.2	22.6	22.8	23.2	24.6	29.0	28.2	27.2
Privación social												
Población con al menos una carencia social	74.2	74.1	72.4	70.6	70.6	70.6	85.0	86.9	86.8	85.7	85.7	85.7
Población con al menos tres carencias sociales	28.2	23.9	22.1	20.5	20.5	20.5	32.4	28.1	26.5	24.9	24.9	24.9
Indicadores de carencia social												
Rezago educativo	20.7	19.2	18.7	17.9	17.9	17.9	23.7	22.6	22.4	21.8	21.8	21.8
Carencia por acceso a los servicios de salud	29.2	21.5	18.2	16.9	16.9	16.9	33.5	25.3	21.8	20.5	20.5	20.5
Carencia por acceso a la seguridad social	60.7	61.2	58.5	56.6	56.6	56.6	69.6	71.8	70.1	68.7	68.7	68.7
Carencia por calidad y espacios en la vivienda	15.2	13.6	12.3	12.0	12.0	12.0	17.4	15.9	14.8	14.6	14.6	14.6
Carencia por acceso a los servicios básicos en la vivienda	22.9	21.2	21.2	20.5	20.5	20.5	26.3	24.9	25.4	24.9	24.9	24.9
Carencia por acceso a la alimentación	24.8	23.3	23.4	21.7	21.7	21.7	28.4	27.4	28.0	26.4	26.4	26.4
Bienestar												
Población con ingreso inferior a la línea de bienestar mínimo	19.4	20.0	20.5	12.2	18.3	19.8	22.2	23.5	24.6	14.8	22.2	24.9
Población con ingreso inferior a la línea de bienestar	52.0	51.6	53.2	42.1	46.2	48.4	49.3	60.6	63.8	51.1	56.1	59.9
Coefficiente de Gini	0.484	0.479	0.480	0.453	0.478	0.487	0.492					

^a Valores observados no comparables; ^b se imputan los deciles I-IV; ^c se imputan los deciles I-VI; ^d se imputan los deciles I-VI.

Fuente: estimaciones propias con base en las ENIGH-MCS 2008, 2010, 2012, 2014 y 2015.

da por el método de imputación múltiple usando tres escenarios (deciles I-IV, I-V y I-VI), los cuales dan muestra del grado de sensibilidad de los deciles imputados. Mientras más deciles se incorporen al modelo de imputación, la pobreza se incrementa, de 40.2% para el escenario I-IV hasta 42.4% para el del I-VI. También, la desigualdad se incrementa de 0.478 para el escenario I-IV hasta 0.492 para el del I-VI. Nos preguntamos, ¿cuál de los tres escenarios presenta una mejor corrección? La respuesta está en el supuesto del número de deciles a imputar. Hemos informado en el cuadro 1 que los deciles a corregir posiblemente son del I al VI, ya que en ellos se encuentran trabajadores de bajos ingresos, como: grupos étnicos, jornaleros o amas de casa, y de cierta forma constituyen un supuesto muy importante para elegir este escenario.

Los resultados para el escenario I-VI indican que la pobreza a nivel nacional es de 42.4% (51.4 millones de personas), una pobreza extrema de 8.4% (10.2 millones de personas) y un coeficiente de Gini de 0.492 (ver cuadro 4). Si tomamos este último escenario, los resultados son mixtos porque, por un lado, la pobreza se reduce (comparada con la incidencia del 2010 al 2014) pero, por el otro, la pobreza absoluta se mantiene como en el 2014. Por otra parte, el coeficiente de Gini se mantiene a la alza en este periodo, lo cual es un indicador de la inequidad en la distribución del ingreso.

Conclusiones

Existe un reto en la captación del ingreso por medio de encuestas; con frecuencia, éste se subestima o no se incluyen personas con ingresos altos. Hay varias formas de enfrentar este problema: los métodos de imputación pueden ser una vía útil para corregir esta dificultad y han sido extensamente utilizados. El presente artículo muestra una posible opción para hacer comparable la serie histórica del ingreso del 2010 al 2014 usando técnicas de imputación múltiple.

Todo modelo parte de supuestos y tendrá mejor desempeño si éstos se cumplen. Para realizar el presente ejercicio parte del supuesto de que el

ingreso se vio afectado en los deciles inferiores de la distribución y, por lo tanto, un modelo que incorpore variables relacionadas con los estratos inferiores del ingreso podrá tener mejor capacidad para corregir este problema. Llamamos a este modelo *de pobreza*, ya que se especializa en medir con mayor precisión a la población con ingresos bajos y, en consecuencia, también se supone que los ingresos de los deciles superiores no fueron muy afectados. Quizá los ingresos altos tengan otro problema, que es el truncamiento, pero este caso está fuera de los objetivos del presente ejercicio.

Se estiman tres modelos de regresión (no se muestran) a través de los métodos de imputaciones múltiples, imputación vía regresión multivariada y FIML; los tres generan los mismos parámetros puntuales del modelo de regresión. Se selecciona el modelo de imputación múltiple y se generan tres escenarios: en el primero se eliminan los ingresos de los deciles I al IV; el segundo, los del I al V; y el tercero, los del I al VI para el MCS 2015.

De ellos, el tercero explica que los deciles I al VI muestran mayor distorsión en la captura del ingreso y, por lo tanto, se pueden corregir mediante imputaciones múltiples. Los resultados indican que la pobreza a nivel nacional es de 42.4%, que representa 51.4 millones de personas, una pobreza extrema de 8.4%, es decir, 10.2 millones de personas, y un coeficiente de Gini de 0.492.

Fuentes

- Arbuckle, J. L. "Full Information Estimation in the Presence of Incomplete Data", in: Marcoulides, G. A. y R. E. Schumacker (eds.). *Advanced Structural Equation Modeling: Issues and Techniques*. Mahwah, New Jersey, Lawrence Erlbaum Associates, Inc., 1996, pp. 243-277.
- Box, G. E. P. y G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Englewood Cliffs, N. J., Prentice Hall, 1976.
- Casella, G. y R. L. Berger. *Statistical Inference* (2nd ed.). Belmont, CA, Duxbury Advanced Series, 2001.
- CONEVAL. "Metodología para la medición multidimensional de la pobreza en México", en: *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*. 2(1). México, INEGI, 2011, pp. 36-63.

- Enders, C. K. *Applied Missing Data Analysis*. New York, The Guilford Press, 2010.
- Gelman, A.; J. B. Carlin; H. S. Stern y D. B. Rubin. *Bayesian Data Analysis*. (2nd ed.). New York, Chapman & Hall, 2004.
- Gilks, W. R.; S. Richardson y D. J. Spiegelhalter. *Markov Chain Monte Carlo in Practice*. Washington, DC, Chapman & Hall-CRC, 1996.
- Graham, J. W. y J. L. Schafer. "On the Performance of Multiple Imputation for Multivariate Data with Small Sample", in: Hoyle, R. H. (ed.): *Statistical Strategies for Small Sample Research*. Thousands Oaks, CA., SAGE Publications, Inc., 1999, pp. 1-29.
- Muthén, B.; D. Kaplan y M. Hollis. "On Structural Equation Modeling with Data That Are Not Missing Completely at Random", in: *Psychometrika*. 52(3), 1987, pp. 431-462.
- Ritter, C. y M. A. Tanner. "The Gibbs Stopper and the Griddy Gibbs Sampler", in: *Journal of the American Statistical Association*. 87, 1992, pp. 861-868.
- Roberts, G. O. "Convergence Diagnosis of the Gibbs Sampler", in Bernardo, J. M.; J. O. Berger; A. P. Dawid y A. F. M. Smith (eds.). *Bayesian Statistics*. Oxford University Press, 1992.
- Rubin, D. B. *Multiple Imputation for Nonresponse in Surveys*. New York, USA, John Wiley & Sons, 1987.
- Russell, D. W.; H. S. Stern y S. Sinharay. *An Evaluation of Multiple Imputation as an Approach to Missing Data*. 2000.
- Schafer, J. L. y J. W. Graham. "Missing Data: Our View of the State of the Art", in: *Psychological Methods*. 7(2), 2002, pp. 147-177.
- Schafer, J. L. y M. K. Olsen. "Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective", in: *Multivariate Behavioral Research*. 33(4), 1998, pp. 545-571.
- Shenker, N. y J. M. G. Taylor. "Partially Parametric Techniques for Multiple Imputation", in: *Computational Statistics & Data Analysis*. 22, 1996, pp. 425-446.
- Vargas Chanes, D. y F. O. Lorenz. "Inference with Missing Data Using Latent Growth Curves", en: *Revista del Instituto Interamericano de Estadística*. 67(188 y 189), 2015, pp. 9-19.

Anexo 1

Variables del modelo

El siguiente cuadro contiene las variables independientes que constituyen los modelos de imputación y de regresión:

Continúa

Variable	Descripción
<i>tam_loc</i>	Tamaño de localidad
<i>rururb</i>	Rural/urbano
<i>edad</i>	Edad
<i>sexo</i>	Sexo
<i>ent</i>	Entidad federativa
<i>tamhogesc</i>	Tamaño del hogar escalado
<i>ic_rezedu</i>	Índice de carencia rezago educativo
<i>inas_esc</i>	Inasistencia a la escuela
<i>niv_ed</i>	Nivel educativo
<i>ic_asalud</i>	Índice de carencia de acceso a servicios de salud
<i>ic_segSOC</i>	Índice de carencia de acceso a la seguridad social
<i>ss_dir</i>	Acceso directo a la seguridad social
<i>jef_ss</i>	Acceso directo a la seguridad social de la jefatura del hogar
<i>cony_ss</i>	Acceso directo a la seguridad social de cónyuge de la jefatura del hogar
<i>hijo_ss</i>	Acceso directo a la seguridad social de hijos(as) de la jefatura del hogar

Concluye

Variable	Descripción
<i>s_salud</i>	Servicios médicos por otros núcleos familiares o por contratación propia
<i>icv_pisos</i>	Índice de carencia del material de piso de la vivienda
<i>icv_muros</i>	Índice de carencia del material de muros de la vivienda
<i>icv_techos</i>	Índice de carencia del material de techos de la vivienda
<i>icv_hac</i>	Índice de carencia por índice de hacinamiento de la vivienda
<i>ic_sbv</i>	Indicador de carencia de acceso a servicios básicos de la vivienda
<i>isb_agua</i>	Indicador de carencia de acceso a servicios de agua
<i>isb_dren</i>	Indicador de carencia de acceso a servicios de drenaje
<i>isb_luz</i>	Indicador de carencia de acceso a servicios de electricidad
<i>isb_combus</i>	Indicador de carencia por acceso a los servicios de combustible
<i>ic_ali</i>	Indicador de carencia por acceso a la alimentación
<i>iden</i>	Hogares con población de 0 a 17 años
<i>ins_ali</i>	Grado de inseguridad alimentaria
<i>vul_car</i>	Población vulnerable por carencias
<i>carencias</i>	Población con al menos una carencia
<i>carencias3</i>	Población con menos de tres carencias
<i>hli</i>	Hablante de lengua indígena
<i>discap</i>	Población con al menos una discapacidad física o mental
<i>deciles</i>	Deciles de ingreso total del hogar
<i>year</i>	Año

Anexo 2

Continúa

Resultados del modelo de imputación (logístico)

Descripción	Variable	Coef.	Std. err.	$P > z$
Tamaño de localidad	<i>tam_loc</i>	0.034	0.011	0.003
Rural/urbano	<i>rururb</i>	2.308	0.035	<0.001
Entidad federativa	<i>Ent</i>	0.004	0.001	<0.001
Edad	<i>edad</i>	0.001	0.001	0.032
Sexo	<i>sexo</i>	0.026	0.015	0.087
Tamaño del hogar escalado	<i>tamhogesc</i>	0.078	0.005	<0.001
Índice de carencia rezago educativo	<i>ic_rezedu</i>	-0.126	0.032	<0.001
Inasistencia a la escuela	<i>inas_esc</i>	0.027	0.027	0.317
Nivel educativo	<i>niv_ed</i>	-0.194	0.013	<0.001
Índice de carencia de acceso a servicios de salud	<i>ic_asalud</i>	-0.270	0.026	<0.001
Índice de carencia de acceso a la seguridad social	<i>ic_segsc</i>	-0.141	0.037	<0.001
Acceso directo a la seguridad social	<i>ss_dir</i>	-0.297	0.033	<0.001

Resultados del modelo de imputación (logístico)

Concluye

Descripción	Variable	Coef.	Std. err.	P > z
Acceso directo a la seguridad social de la jefatura del hogar	<i>jef_ss</i>	-0.514	0.021	<0.001
Acceso directo a la seguridad social de cónyuge de la jefatura del hogar	<i>cony_ss</i>	-1.338	0.034	<0.001
Acceso directo a la seguridad social de hijos(as) de la jefatura del hogar	<i>hijo_ss</i>	-1.031	0.029	<0.001
Servicios médicos por otros núcleos familiares o por contratación propia	<i>s_salud</i>	-0.134	0.030	<0.001
Índice de carencia del material de piso de la vivienda	<i>icv_pisos</i>	0.124	0.050	0.013
Índice de carencia del material de muros de la vivienda	<i>icv_muros</i>	-0.348	0.065	<0.001
Índice de carencia del material de techos de la vivienda	<i>icv_techos</i>	0.070	0.068	0.299
Índice de carencia por índice de hacinamiento de la vivienda	<i>icv_hac</i>	0.006	0.033	0.851
Indicador de carencia de acceso a servicios básicos de la vivienda	<i>ic_sbv</i>	0.038	0.048	0.433
Indicador de carencia de acceso a servicios de agua	<i>isb_agua</i>	0.066	0.039	0.091
Indicador de carencia de acceso a servicios de drenaje	<i>isb_dren</i>	0.301	0.039	<0.001
Indicador de carencia de acceso a servicios de electricidad	<i>isb_luz</i>	-0.019	0.114	0.868
Indicador de carencia por acceso a los servicios de combustible	<i>isb_combus</i>	0.435	0.041	<0.001
Indicador de carencia por acceso a la alimentación	<i>ic_ali</i>	-0.467	0.042	<0.001
Hogares con población de 0 a 17 años	<i>id_men</i>	0.396	0.023	<0.001
Grado de inseguridad alimentaria	<i>ins_ali</i>	0.355	0.016	<0.001
Hablante de lengua indígena	<i>hli</i>	0.180	0.036	<0.001
Población con al menos una discapacidad física o mental	<i>discap</i>	0.183	0.031	<0.001
Población vulnerable por carencias	<i>vul_car</i>	-5.041	0.025	<0.001
Población con al menos una carencia	<i>carencias</i>	2.795	0.034	<0.001
Población con menos de tres carencias	<i>carencias3</i>	0.088	0.037	0.018
		-1.839	0.055	<0.001

Tabla de la clasificación correcta del modelo de regresión logístico

Sensibilidad	81.53%
Especificidad	90.92%
Valor positivo predicho	84.97%
Valor negativo predicho	88.66%
Tasa de falsos positivos para un valor observado	9.08%
Tasa de falsos negativos para un valor faltante	18.47%
Tasa de falsos positivos	15.03%
Tasa de falsos negativos	11.34%
Clasificación correcta	87.29%

**Modelo de regresión de imputación múltiple.
Escenario que omite los deciles I-VI (resumen de las 10 imputaciones)**

Descripción	Variable	Coef.	Std. err.	P > t
Tamaño de localidad	<i>tam_loc</i>	-0.001	0.001	0.280
Rural/urbano	<i>rururb</i>	0.010	0.002	<0.001
Edad	<i>edad</i>	0.000	0.000	<0.001
Sexo	<i>sexo</i>	-0.002	0.001	0.110
Tamaño del hogar escalado	<i>tamhogesc</i>	-0.001	0.000	0.004
Índice de carencia rezago educativo	<i>ic_rezedu</i>	-0.005	0.002	0.009
Inasistencia a la escuela	<i>inas_esc</i>	-0.018	0.002	<0.001
Nivel educativo	<i>niv_ed</i>	0.006	0.001	<0.001
Índice de carencia de acceso a servicios de salud	<i>ic_asalud</i>	-0.002	0.001	0.189
Índice de carencia de acceso a la seguridad social	<i>ic_segsoc</i>	0.022	0.003	<0.001
Acceso directo a la seguridad social	<i>ss_dir</i>	0.011	0.002	<0.001
Acceso directo a la seguridad social de la jefatura del hogar	<i>jef_ss</i>	0.010	0.001	<0.001
Acceso directo a la seguridad social de cónyuge de la jefatura del hogar	<i>cony_ss</i>	0.028	0.002	<0.001
Acceso directo a la seguridad social de hijos(as) de la jefatura del hogar	<i>hijo_ss</i>	-0.002	0.001	0.180
Servicios médicos por otros núcleos familiares o por contratación propia	<i>s_salud</i>	0.003	0.002	0.094
Índice de carencia del material de piso de la vivienda	<i>icv_pisos</i>	-0.018	0.003	<0.001
Índice de carencia del material de muros de la vivienda	<i>icv_muros</i>	-0.005	0.004	0.217
Índice de carencia del material de techos de la vivienda	<i>icv_techos</i>	0.004	0.004	0.315

**Modelo de regresión de imputación múltiple.
Escenario que omite los deciles I-VI (resumen de las 10 imputaciones)**

Descripción	Variable	Coef.	Std. err.	P > t
Índice de carencia por índice de hacinamiento de la vivienda	<i>icv_hac</i>	-0.001	0.002	0.553
Indicador de carencia de acceso a servicios básicos de la vivienda	<i>ic_sbv</i>	0.025	0.003	<0.001
Indicador de carencia de acceso a servicios de agua	<i>isb_agua</i>	-0.020	0.002	<0.001
Indicador de carencia de acceso a servicios de drenaje	<i>isb_dren</i>	-0.022	0.003	<0.001
Indicador de carencia de acceso a servicios de electricidad	<i>isb_luz</i>	-0.043	0.007	<0.001
Indicador de carencia por acceso a los servicios de combustible	<i>isb_combus</i>	-0.019	0.003	<0.001
Indicador de carencia por acceso a la alimentación	<i>ic_ali</i>	0.019	0.003	<0.001
Hogares con población de 0 a 17 años	<i>id_men</i>	-0.001	0.002	0.403
Grado de inseguridad alimentaria	<i>ins_ali</i>	-0.016	0.001	<0.001
Población vulnerable por carencias	<i>vul_car</i>	-0.049	0.002	<0.001
Población con al menos una carencia	<i>carencias</i>	0.030	0.002	<0.001
Población con menos de tres carencias	<i>carencias3</i>	-0.001	0.002	0.763
Hablante de lengua indígena	<i>hli</i>	-0.026	0.002	<0.001
Población con al menos una discapacidad física o mental	<i>discap</i>	-0.015	0.002	<0.001
Deciles de hogar	<i>deciles</i>	0.287	0.000	<0.001
Año	<i>year</i>	0.085	0.001	<0.001
Constante	<i>_cons</i>	163.327	2.930	<0.001

Ajuste por factor de crecimiento con base en el indicador de gasto de la ENIGH

Adjustment by means of a factor growth based on an **ENIGH's** expenditure indicator

Edgar Vielma Orozco,* Octavio Heredia Hernández** y José Vences Rivera***

Los resultados del Módulo de Condiciones Socioeconómicas (MCS) 2015 presentaron una discontinuidad en la tendencia que se venía observando en algunas variables de ingreso. En este documento se desarrolla un modelo para estimar el factor de ajuste entre la serie histórica y la nueva magnitud de ingreso monetario. Con este factor se llevan a cabo imputaciones a las observaciones del MCS 2015 de tal forma que se aparenten los resultados que se hubieran observado en caso de que las circunstancias de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)-MCS 2014 se hubieran conservado en el MCS 2015. La estimación del factor de ajuste se basa en la relación histórica entre el ingreso de los hogares y el gasto, ambos a nivel nacional, modelada a partir del ajuste de una regresión lineal simple a las estimaciones de estos parámetros obtenidas con la ENIGH en sus diferentes ediciones.

Palabras clave: encuestas de ingresos; modelos de imputación; distribución de los ingresos; deciles de ingresos; fuentes de ingreso; México; oficinas nacionales de estadística; comparabilidad; series históricas; variables de control; gasto; factor de crecimiento; base de datos.

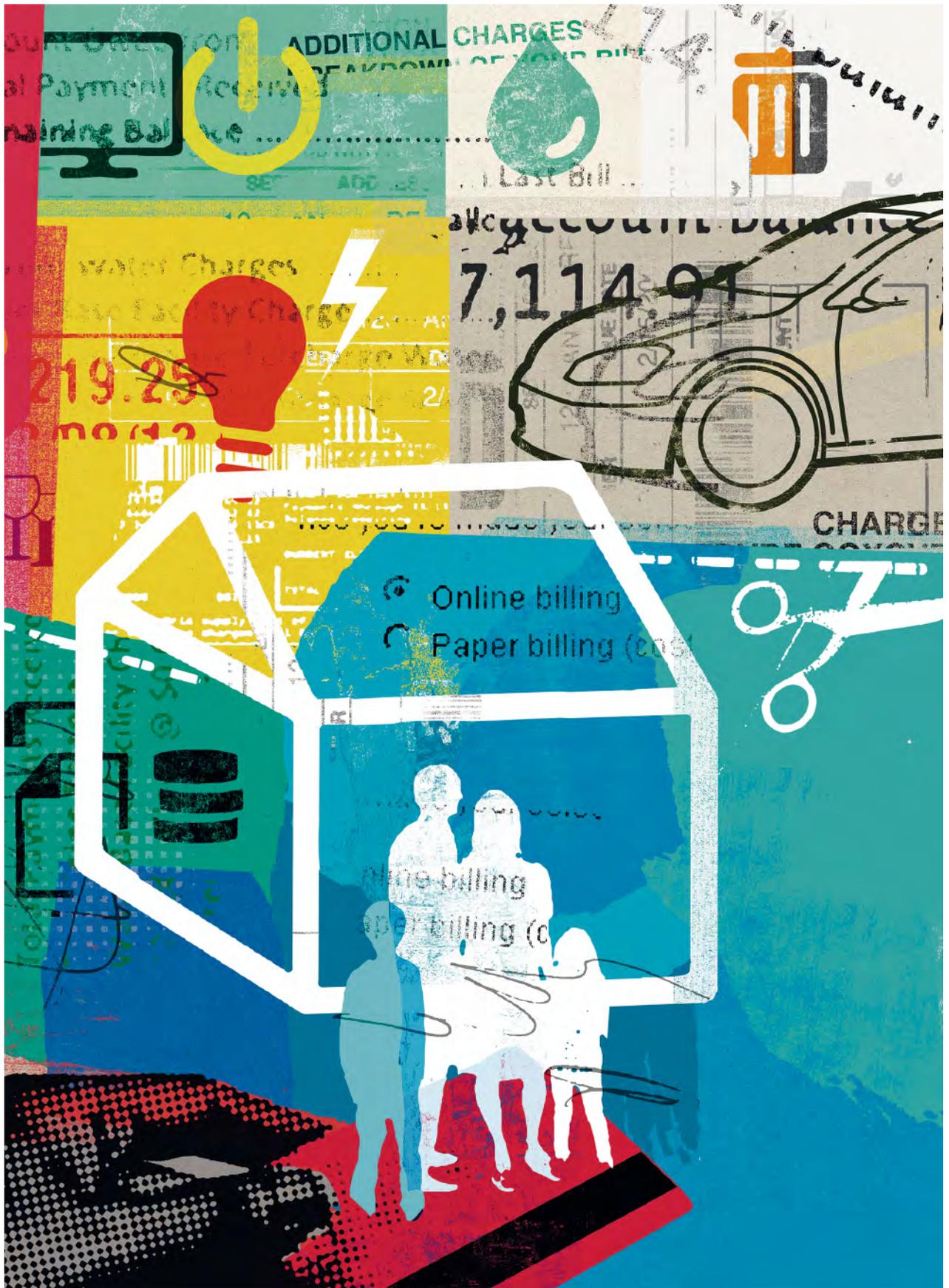
* Instituto Nacional de Estadística y Geografía (INEGI), edgar.vielma@inegi.org.mx

** INEGI, octavio.heredia@inegi.org.mx

*** INEGI, jose.vences@inegi.org.mx

The results of the 2015 Socio-Economic Conditions Module (MCS by its Spanish acronym) survey presented a discontinuity in the tendency observed in some of the income variables. This document develops a model to estimate the adjustment factor between the historical series and the new magnitude of monetary income. With this factor, imputations to the observations of the 2015 MCS are done in such a way that they simulate the results that would have been observed if the circumstances of the 2014 MCS-ENIGH had been kept for the 2015 MCS. The estimation of the adjustment factor is based on the historical relationship between the households income and expenditure, both at a national level, modeled based on the adjustment of a simple linear regression to the estimations of these parameters obtained with the different ENIGH surveys.

Key words: income surveys; imputation models; income distribution; income deciles; income sources; Mexico; National Statistical Offices (NSO); comparability; historical series; control variables; expenditure; growth factor; database.



Collage of family on credit card and household bills/Stuart Kinlough/Getty Images

1. Introducción

La medición del ingreso a través de censos o encuestas es un reto complejo y dinámico que enfrentan las oficinas nacionales de estadística (ONE) de todo el mundo. El caso de México no es la excepción por lo que, a través de los años, el Instituto Nacional de Estadística y Geografía (INEGI) ha tenido que enfrentarlo con diversas tácticas. Una de ellas es la disponibilidad de un equipo de campo altamente capacitado y, en la mayoría de los casos, con varios años de experiencia. Sin embargo, el tener control en elementos propios del levantamiento no se extiende a tener controladas las actitudes de los informantes y, mucho menos, el ambiente que los rodea.

El control, o al menos la caracterización de los informantes, es muy complejo. En términos generales, se entiende que existe un sesgo de los informantes a reportar un ingreso menor al que realmente tienen. El problema, sin embargo, no es el subreporte, sino la heterogeneidad con la que ocurre. Esto es, el sesgo de subreporte no es constante a través del tiempo, de estrato socioeconómico o por localización geográfica. La heterogeneidad se ve fomentada por eventos con ciertos rasgos aleatorios, como delincuencia, elecciones y ciclos económicos locales.

Las ONE requieren estar en constante revisión de las actitudes y los factores que afectan a los informantes y generar acciones que permitan mitigar, en la mejor medida posible, la heterogeneidad de las respuestas. De acuerdo con las prácticas recomendadas por los organismos internacionales de estadística, las acciones de revisión se llevan a cabo cada vez que se realiza un levantamiento o ciclo estadístico.

La Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH)-Módulo de Condiciones Socioeconómicas (MCS) 2014 llevó a cabo un proceso regular de revisión del cual se desprendieron algunas iniciativas, entre las que destacan: 1) ajustar el cuestionario al nuevo entorno generado por la reforma fiscal, 2) incrementar la consistencia entre la

información de ingreso reportada y el entorno del entrevistado y 3) generación en línea de reportes sobre errores detectados por parte de los entrevistadores. Aunque estas acciones no representan en sí un cambio de diseño conceptual, estadístico, de estrategias operativas de captación y del procesamiento de la información, el INEGI captó un ingreso con un crecimiento no visto con anterioridad.

Por lo anterior, el Instituto detonó un análisis de tipo cualitativo con el fin de tener un mejor entendimiento de los resultados obtenidos, que consistió en entrevistas bajo la técnica de *rapport ethnográfico* para sondear las experiencias de campo del personal que participó en la capacitación a nivel nacional, en su réplica en las entidades federativas y en el operativo de campo. El estudio detectó que los entrevistadores no consideraron que había cambiado la metodología porque siguieron preguntando lo mismo y haciendo el mismo procedimiento que en ejercicios anteriores, pero manifestaron que en la capacitación se les pidió ser más analíticos respecto al entorno socioeconómico del hogar.

En este documento se presentan los resultados de un modelo que adapta los resultados del MCS 2015 a la tendencia histórica que se venía observando. El modelo está calibrado con los datos históricos de la ENIGH para caracterizar la distancia entre el ingreso y otros factores y, con esto, generar un factor de ajuste, neutralizando la ganancia de una visión más analítica en los entrevistadores.

En el 2016 se hizo el levantamiento de la ENIGH 2016 la cual, a diferencia del MCS 2015, permite obtener una medición tanto del ingreso como del gasto de los hogares. Derivado de que la ENIGH 2016 no presenta cambio alguno ni de diseño conceptual y estadístico ni de operativo de captación de información con respecto al MCS 2015, se espera que el ingreso sea comparable entre ambos proyectos al implementar la ENIGH 2016 las mismas mejoras que el MCS 2015. De manera adicional, cabe señalar que no hubo ninguna instrucción en la capacitación relacionada con ser más analítico respecto

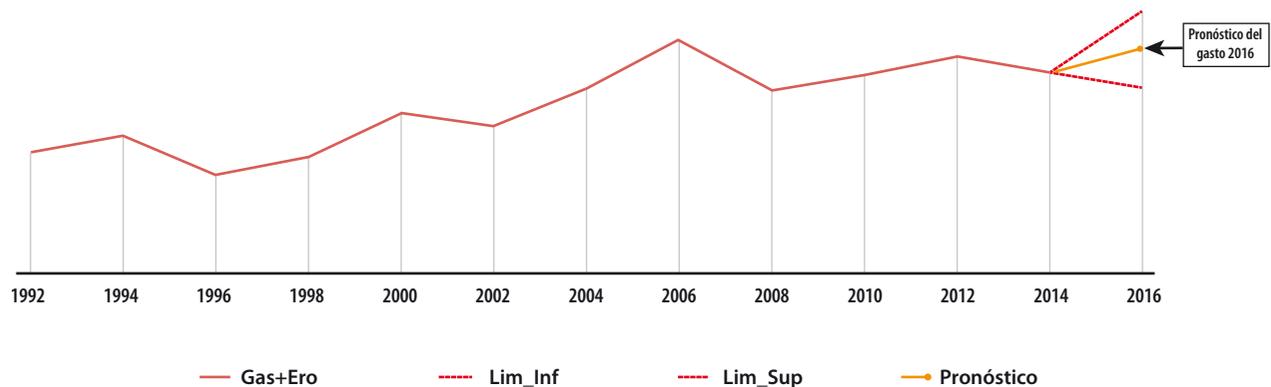
al gasto de los hogares, razón por la cual se prevé que la medición del gasto preserve una continuidad con la serie histórica de levantamientos anteriores. Dado que los resultados de la ENIGH 2016 aún no estaban publicados, fue necesario estimar un gasto (pronóstico) que pueda ser implementado en el modelo, dada la consistencia que existe en esta variable a través del tiempo (ver gráfica A).

Utilizando el gasto pronosticado 2016, el cual se prevé sea muy cercano al valor que se espera de la ENIGH 2016, se puede construir un ingreso ajustado para el 2016. El presente modelo desarrolla ese ajuste que permitirá neutralizar cualquier cambio atípico del ingreso esperado en ese año e, incluso,

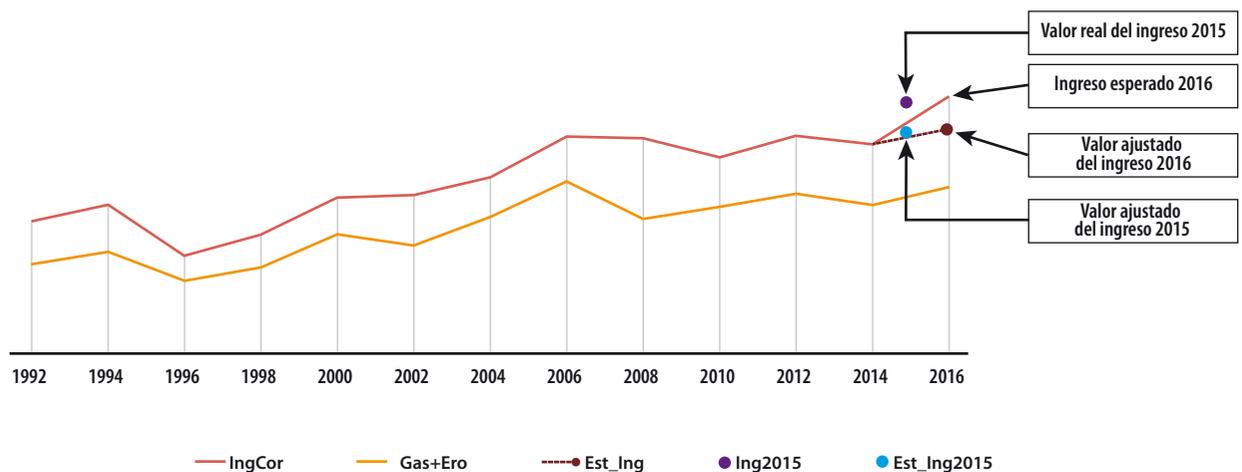
interpolarse un valor del ingreso 2015 con la finalidad de tener una comparabilidad del ingreso 2015 (ajustado) con el resto de la serie histórica. Es importante mencionar que existe una alta correlación (0.95) entre el gasto y el ingreso (ver gráfica B).

La estructura general de este documento es la siguiente: en la segunda sección se describen algunos antecedentes de los problemas de ingreso que existen tanto en México como en otros países; en la tercera se muestra el modelo conceptual y en la cuarta, el modelo empírico con las variables utilizadas; en la quinta sección se describen los resultados; y la sexta presenta las conclusiones del modelo.

Gráfica A



Gráfica B



2. Antecedentes

2.1 Antecedentes históricos de las encuestas de ingresos

Es importante conocer que la complejidad y perfeccionamiento que han adquirido las actuales encuestas de ingresos forman parte de un proceso histórico cuyos antecedentes se remontan a poco más de cien años en México y el resto del mundo.

A nivel mundial, los primeros estudios de gastos en los hogares nacieron con las manifestaciones iniciales de la Revolución Industrial, el surgimiento de la clase obrera y los centros urbanos industrializados. La clase obrera urbana comenzó rápidamente a ser el grupo de población más extenso en muchas ciudades y países, además del principal motor de la economía mundial. A la par, las relaciones económicas mundiales se interconectaban cada vez más, de manera que, si alguna economía se afectaba o sufría pérdidas, las demás se veían también afectadas. Con estos fenómenos comenzó a surgir también la pobreza urbana, afectando grandes sectores de la población y las ciudades.

La manifestación de la pobreza en las calles llamó la atención de científicos y estudiosos, que empezaron a analizar el gasto que efectuaban las familias para conocer si sus salarios e ingresos eran o no suficientes para cubrir sus necesidades básicas. Las encuestas sobre ingresos y gastos más antiguas se han encontrado en el Reino Unido (1904), Alemania (1907), Australia y Nueva Zelanda (1910), Austria y Suecia (1912). En 1917, los Países Bajos realizaron una encuesta de ingresos y gastos, mientras que en 1918, Estados Unidos de América y Noruega llevaron a cabo la propia; por último, Japón desarrolló en 1920 su propia encuesta.¹

Hubo dos acontecimientos mundiales que propiciaron el desarrollo de este tipo de levantamientos en todo el mundo de manera más homogénea

e institucionalizada: el primero fue la creación de la Organización Internacional del Trabajo (OIT) en 1919 como parte del *Tratado de Paz de Versalles* y que puso fin a la Primera Guerra Mundial, reflejando la convicción de que la justicia social era esencial para alcanzar la paz; en el *Tratado* se determinó que la OIT sería el organismo encargado de unificar y universalizar las estadísticas sobre temas laborales, principalmente estadísticas de accidentes laborales, de paros y huelgas y sobre salarios y costo de la vida.²

El segundo fue la crisis económica de 1929, que tuvo un fuerte impacto en todo el mundo, generando altos niveles de desempleo en las principales ciudades; con ella se vieron seriamente afectados los salarios y el poder adquisitivo de las familias. Esta situación obligó a intensificar la generación de información relativa al costo de la vida sobre todo porque, a pesar de que autoridades, patrones y empleados sabían que la capacidad de compra de los salarios se había deteriorado, era difícil determinar con exactitud en qué cantidad había aumentado el costo de la vida y cuánto se había deteriorado el poder adquisitivo de los salarios.³

A partir de esa época, y bajo la batuta de la OIT, se levantaron en el mundo encuestas sobre ingresos y gastos como insumo importante para medir el índice de precios, importante indicador que determina los niveles de inflación o, dicho de otra manera, de los cambios en los precios de los artículos de consumo más importantes y la capacidad de compra de los salarios.

En México se han encontrado antecedentes de este tipo de encuestas desde 1914, cuando el Departamento del Trabajo aplicó diferentes cuestionarios y metodologías para conocer los principales gastos de las familias obreras.⁴ El levantamiento de la información se hacía de diferentes maneras y

1 Morse, R. *Métodos de encuesta sobre las condiciones de vida de las familias. Ingresos-gastos-consumo*. 1942. Recuperado de http://staging.ilo.org/public/libdoc/ilo/ILO-SR/ILO-SR_N23_span.pdf

2 International Labour Office. *Report on the International Conference of Representatives of Labour Statistical Departments, held at Geneva, 29 October to 2 November 1923*. 1924. Recuperado de http://www.ilo.org/public/libdoc/ilo/ILO-SR/ILO-SR_N4_engl.pdf

3 Cárdenas, E. *La hacienda pública y la política económica, 1929-1958*. México, El Colegio de México-Fideicomiso Historia de las Américas-Fondo de Cultura Económica, 1994.

4 Gómez, A. & A. Mussachio. "Un nuevo índice de precios para México, 1886-1929", en: *El Trimestre Económico*. Vol. 67, núm. 265, enero-marzo 2000, pp. 47-91.

con varios cuestionarios. Una parte de la información se captaba directamente con familias obreras seleccionadas de forma previa; se les preguntaba sobre sus gastos durante un tiempo determinado, por lo general una semana. Otros datos se obtenían mediante esqueletos de cuestionarios que se enviaban a las presidencias municipales para que registraran los precios que habían tenido en su municipio determinados artículos de consumo; esta información se solicitaba mes con mes y se publicaba en el *Boletín del Departamento del Trabajo* mediante cuadros que mostraban la evolución de los precios a nivel nacional.

Fue hasta 1956 que se levantó la primera encuesta con métodos científicos de muestreo; estuvo a cargo de la Dirección General de Estadística (DGE), que entonces pertenecía a la Secretaría de Economía (SE): se llamó Ingresos y Egresos de la Población de México; en 1958, la misma DGE levantó la encuesta Ingresos y Egresos de las Familias en la República Mexicana. Más adelante, en 1960, se realizó la llamada Las 16 Ciudades Principales de la República Mexicana, Ingresos y Egresos Familiares. Por su parte, el Banco de México, llevó a cabo en 1963 el levantamiento de la encuesta Ingresos y Gastos Familiares en México y, en 1968, otra denominada los Ingresos y Gastos de las Familias.

Durante el periodo 1969-1970, nuevamente la DGE, pero ahora perteneciente a la Secretaría de Industria y Comercio (SIC), realizó la encuesta Ingresos y Egresos de la República Mexicana. En 1975 fue la Secretaría del Trabajo y Previsión Social (STPS) la instancia que efectuó la encuesta Ingresos y Gastos Familiares. En 1977 la DGE, pero ahora como parte de la Secretaría de Programación y Presupuesto (SPP), desarrolló la ENIGH, antecedente inmediato de la ENIGH actual que ha llevado a cabo el INEGI desde 1984.

A partir de que el INEGI tomó la batuta de esta encuesta, trabajó por ampliar los objetivos y homogeneizar la metodología, tomando en cuenta las recomendaciones internacionales y los requerimientos de información de diferentes usuarios, pero siempre cuidando la comparabilidad histórica.

Desde entonces, la ENIGH ha ganado reconocimiento como una de las encuestas más consolidadas de México, y ha sido referente internacional de proyectos similares.

En el 2002, la Secretaría de Desarrollo Social (SEDESOL) creó una metodología oficial para la medición de la pobreza y convino que se aplicara a la ENIGH. A partir de esta fecha, la ENIGH aporta la información necesaria para medir la evolución de la pobreza en nuestro país. Esta responsabilidad se afianzó en el 2004 con la *Ley General de Desarrollo Social* en la que se estableció la creación del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) encargado de medir la evolución de la pobreza a partir de la información generada por el INEGI.

A la ENIGH se le han añadido otros instrumentos para complementar información, tal es el caso del MCS, elemento indispensable para la medición de la pobreza desde el 2008. Más adelante, el 4 de septiembre de 2012 se publicó en el *Diario Oficial de la Federación* el acuerdo por el que se determinó como Información de Interés Nacional a la ENIGH. Así, a lo largo de los años, la Encuesta se ha fortalecido como uno de los proyectos estadísticos de gran importancia para la determinación de la política social.

2.2 Estudios sobre la calidad de las encuestas de ingresos

De todas las estadísticas recabadas por el gobierno, quizá ninguna es más ubicua y con mayor impacto en el diseño de políticas públicas que aquella relacionada con el ingreso. Es por esta razón que existe una amplia literatura dedicada a la investigación de la calidad en la medición de las encuestas de ingresos.

Un primer enfoque del tema es la comparación del ingreso entre diferentes tipos de fuentes. Por un lado, las *macro*, que se refieren sobre todo a los sistemas nacionales de cuentas de los países y, por el otro, a las fuentes *micro* que principalmente consisten en encuestas, registros administrativos y censos.

En este sentido, el estudio más completo del tema es el realizado por la Organización para la Cooperación y el Desarrollo Económicos (OCDE) y la Oficina Europea de Estadística (EUROSTAT) en el 2011,⁵ en el cual se busca cuantificar la diferencia entre las fuentes macro y micro a lo largo de un grupo de varios países, incluido México. En dicho estudio se utiliza un indicador de diferencia media (*average gap indicator*), el cual abarca un valor entre 0 y 100%, donde 0% indicaría un emparejamiento perfecto entre la información de fuentes micro con macro y 100% mostraría una diferencia muy grande, ya sea porque no se cuenta con alguna de las fuentes o que una es el doble de la otra. Los resultados muestran que para los 20 países analizados el promedio de este índice fue de 36%, desde un rango de 16% en Japón hasta 70% en México.

En el mismo estudio se analizan las tasas de cobertura de los diferentes componentes que integran el ingreso, así que, para los 19 países analizados, los dos componentes con la menor alineación entre fuentes micro y macro fueron los *ingresos por cuenta propia* y los *intereses e ingresos repartidos recibidos de empresas*; en la mayoría de los países estos componentes muestran tasas de cobertura inferiores a 80 por ciento. Cabe señalar el caso de México, donde la tasa de cobertura para los *ingresos por cuenta propia* fue menor a 20 por ciento.

Un segundo enfoque consiste en conocer la magnitud y naturaleza de los errores de los encuestados a la hora de proporcionar información sobre sus ingresos, así como entender la base cognitiva de dichos errores.

Dentro de este enfoque, destaca la Oficina del Censo de los Estados Unidos de América, que cuenta con varias investigaciones al respecto. En una revisión⁶ de varios estudios en 1997, hecha por la misma agencia del gobierno, se indica que, aunque es notorio que las encuestas de ingreso tienden a subestimar la cantidad del ingreso respecto

al punto de referencia (*benchmark*) que son las cuentas nacionales o registros administrativos, no existe evidencia contundente para afirmar que se debe a una tendencia generalizada de los encuestados a reportar ingresos más bajos de manera intencionada. También se referencia que, aunque es común que en las encuestas de ingresos la tasa de no respuesta sea entre 7-8% y que los reportes con datos imputados sean próximos a uno de cada cinco, esto no es suficiente para explicar las diferencias entre los ingresos recabados en encuestas y aquellos obtenidos por otras fuentes.

En esta misma revisión se establece que el reporte de ingresos en las encuestas es generalmente un proceso de dos etapas. En la primera se reportan las fuentes de ingreso y en la segunda, las cantidades de ingreso por cada fuente. Siendo que los errores pueden caer en cualquiera de las dos, éstos se suelen analizar de manera separada.

De los errores en la primera etapa, diversos estudios encuentran una tendencia a subestimar las fuentes de ingreso cuya magnitud es altamente variable. La revisión por la Oficina del Censo encontró que, mientras que las fuentes de *sueldos y salarios* son apenas afectadas por la subestimación, las de *transferencias* lo son un poco más, y que la mayor subestimación se encuentra en las fuentes de *ingresos por activo*.

En cuanto a los errores en la etapa de reportar la cantidad de ingreso de cada fuente, varios estudios concluyen que la cantidad de ingreso es, en general, bien reportada, aunque existen pequeñas subestimaciones en la cantidad de ingresos en varias fuentes que no sean *sueldos y salarios*. La revisión por la Oficina del Censo señala, por un lado, que las cantidades por ingresos de *sueldos y salarios* contienen un error muy pequeño, el cual no es de signo claro; por otro lado, que las cantidades de *ingresos por activo* tienen una pequeña subestimación y que las de ingreso por *transferencias*, una mayor.

Por último, esta revisión señala entre las causas más comunes para estos dos tipos de errores por

5 OCDE-EUROSTAT. *A cross-country comparison of household income, consumption and wealth between micro sources and national accounts aggregates*. Paris, OCDE, 2011.

6 C. Moore, Jeffrey. *Income Measurement Error in Surveys: A Review*. Suitland, Oficina del Censo de los Estados Unidos, 1997.

parte de los encuestados las siguientes: dificultad de éstos para comprender los conceptos y categorías de los ingresos, falta de conocimientos sobre sus ingresos, pasar por alto ingresos con una importancia baja, diferentes estrategias para calcular los montos de ingresos, tendencia a recordar más los ingresos más recientes, confusión entre distinguir cuando el ingreso fue ganado y cuándo recibido, la disponibilidad o no de un registro de sus ingresos y, finalmente, la sensibilidad y tabú que representa para varias personas el hablar de forma transparente sobre sus ingresos.

2.3. Resultados del MCS 2015

El MCS 2015 registró un incremento del ingreso corriente (IC) en términos reales del orden de 15.64% respecto al levantamiento ENIGH-MCS 2014. Hay tres componentes que concurren detrás de esto: I) el crecimiento en el número de hogares, II) el incremento en el promedio de perceptores por hogar y III) el aumento en el promedio del ingreso por perceptor. Los primeros dos exceden, con mucho, los incrementos bienales registrados en levantamientos anteriores, así como lo que registran otras fuentes; en particular, la forma como aumentan los hogares se relaciona con un proceso de expansión compensatoria ante el hecho de que el tamaño del hogar en el MCS 2015 resultó menor que en el levantamiento previo, de modo que se necesitaron más hogares para ajustar a las proyecciones del Consejo Nacional de Población (CONAPO), que tienen como referente todas las encuestas en hogares del INEGI. En cuanto al incremento de perceptores por hogar es, sobre todo, un reflejo del aumento de perceptores ocupados (es decir, con ingresos del trabajo).

Al respecto, aunque la expansión pudo tener algún efecto, no basta para explicar lo observado. Por su parte, el incremento en el promedio de ingresos por perceptor resulta de la combinación de un retroceso en la captación de fuentes de ingresos en el 2014 y un aumento en las fuentes de ingreso por perceptor del 2015, acompañado de una mayor declaración de ingresos. De los tres factores

considerados (hogares, perceptores por hogar e ingresos por perceptor) es el tercero el que terminó siendo dominante en el agregado, así como en la mayoría de los deciles, salvo en el IX y el X; pero donde el incremento del ingreso por perceptor ocurrió de manera más pronunciada fue en los primeros deciles.⁷

2.3.1 Estadística descriptiva del ingreso en el MCS 2015

En la tabla 1 se muestra el ingreso total por fuentes de ingresos según el año de levantamiento. Como se observa, para el periodo 2010-2012 el incremento fue de 1.4% y para el 2012-2014 fue de 2.6%, lo cual contrasta fuertemente con el incremento de 15.6% para el periodo 2014-2015. Cabe mencionar que este incremento se da además en un periodo de un solo año, es decir, en la mitad del tiempo que en los incrementos anteriores, lo cual pone de manifiesto lo extraordinario de la captación del ingreso en el MCS 2015.

Cuando el análisis se realiza por ingreso promedio en vez de ingreso total, se observa de nuevo una variación muy grande del ingreso en el MCS 2015, en comparación con otros años. Para lo anterior, se presenta la tabla 2, en la cual se ve que la variación no se reparte de igual manera entre las diversas fuentes de ingreso, ya que va desde una variación positiva en 66% en cuanto a *Renta de la propiedad*, hasta una negativa de -11% para el rubro de *Otros ingresos*.

Aunque la variación por fuente de ingreso fue muy heterogénea entre las diversas fuentes de ingreso, también se observa que la estructura porcentual que éstos representan no sufrió una alteración importante, en parte debido a que los cambios de mayor variación no se dieron en las más importantes. Para lo anterior, se muestra en la gráfica 1 la composición del ingreso por fuente en los diferentes levantamientos.

⁷ INEGI. *Análisis y conclusiones de los resultados del Módulo de Condiciones Socioeconómicas 2015 llevado a cabo por el INEGI en el contexto del Grupo Técnico INEGI- CONEVAL*. Ciudad de México, INEGI, 2016.

Tabla 1

Ingreso total trimestral por principales fuentes de ingreso según año de levantamiento (precios constantes del 2015)^a

Fuentes de ingreso	Año de levantamiento (miles de pesos)				Variación (porcentaje)		
	2010	2012	2014	2015	2010-2012	2012-2014	2014-2015
Ingreso corriente	1 266 332 790	1 284 468 404	1 318 162 696	1 524 262 517	1.4	2.6	15.6
Ingreso del trabajo	867 921 751	895 761 799	908 604 869	1 022 789 609	3.2	1.4	12.6
Remuneraciones por trabajo subordinado	730 774 271	740 384 957	778 967 567	858 777 632	1.3	5.2	10.2
Ingresos por trabajo independiente	104 175 265	119 428 945	92 637 747	118 533 617	14.6	-22.4	28.0
Ingresos de otros trabajos	32 972 215	35 947 897	36 999 555	45 478 360	9.0	2.9	22.9
Renta de la propiedad	55 227 256	45 565 331	63 163 701	108 847 937	-17.5	38.6	72.3
Transferencias	177 972 618	185 718 344	183 278 356	225 206 403	4.4	-1.3	22.9
Estimación del alquiler de la vivienda	163 573 381	155 918 863	161 664 054	166 083 256	-4.7	3.7	2.7
Otros ingresos corrientes	1 637 784	1 504 067	1 451 716	1 335 312	-8.2	-3.5	-8.0

^a INEGI. Presentación sobre el Módulo de Condiciones Sociodemográficas 2015. Ciudad de México, INEGI, 2016.

Tabla 2

Ingreso promedio trimestral por hogar por principales fuentes de ingreso según año de levantamiento (precios constantes del 2015)

Fuentes de ingreso	Año de levantamiento (pesos)				Variación (porcentaje)		
	2010	2012	2014	2015	2010-2012	2012-2014	2014-2015
Ingreso corriente	42 725	40 959	41 000	45 887	-4.1	0.1	11.9
Ingreso del trabajo	29 283	28 564	28 261	30 790	-2.5	-1.1	8.9
Remuneraciones por trabajo subordinado	24 656	23 610	24 229	25 853	-4.2	2.6	6.7
Ingresos por trabajo independiente	3 515	3 808	2 881	3 568	8.4	-24.3	23.8
Ingresos de otros trabajos	1 112	1 146	1 151	1 369	3.0	0.4	19.0
Renta de la propiedad	1 863	1 453	1 965	3 277	-22.0	35.2	66.8
Transferencias	6 005	5 922	5 701	6 780	-1.4	-3.7	18.9
Estimación del alquiler de la vivienda	5 519	4 972	5 028	5 000	-9.9	1.1	-0.6
Otros ingresos corrientes	55	48	45	40	-13.2	-5.9	-11.0

Otra forma de exponer el aumento de los ingresos es estudiar cómo se distribuye el incremento del ingreso en el promedio de los diferentes deciles; para ello, se cuenta con la tabla 3, en la que se constata que el mayor incremento en los ingresos se dio en el promedio del decil I, donde el incremento fue de 33.6%, seguido del segundo, con 21.2 por ciento. Lo anterior explica la reducción del coeficiente de Gini a 0.445, el cual se había mantenido constante en los tres ejercicios anteriores en

alrededor de 0.46. Cabe mencionar que, en levantamientos anteriores, el incremento máximo para cualquier decil en un periodo de dos años había sido de 6.2% para el primer decil en el periodo 2012-2014, lo cual es superado para este levantamiento por todos los deciles en solo un año.

También se hace necesario mencionar que la variación del ingreso promedio tampoco se repartió de la misma manera entre las diferentes entidades

Gráfica 1

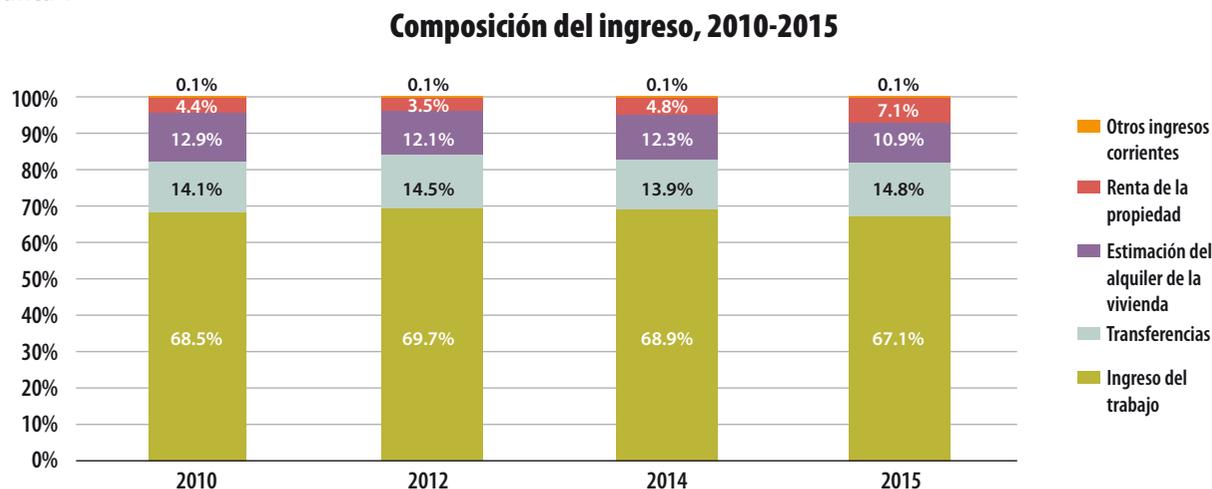


Tabla 3

Ingreso promedio trimestral por hogar ordenado por deciles de hogares según año de levantamiento y su coeficiente de Gini (precios constantes del 2015)

Deciles de hogares	Año de levantamiento				Variación		
	2010	2012	2014	2015	2010-2012	2012-2014	2014-2015
Total	42 725	40 959	41 000	45 887	-4.1	0.1	11.9
I	5 944	5 757	6 116	8 169	-3.1	6.2	33.6
II	11 386	10 999	11 409	13 830	-3.4	3.7	21.2
III	15 838	15 318	15 622	18 387	-3.3	2.0	17.7
IV	20 303	19 768	19 778	23 147	-2.6	0.1	17.0
V	25 382	24 662	24 564	28 478	-2.8	-0.4	15.9
VI	31 511	30 637	30 178	34 866	-2.8	-1.5	15.5
VII	39 245	38 264	37 433	43 030	-2.5	-2.2	15.0
VIII	50 600	49 108	48 105	54 139	-2.9	-2.0	12.5
IX	70 588	67 419	66 510	73 252	-4.5	-1.3	10.1
X	156 457	147 662	150 283	161 568	-5.6	1.8	7.5

Coeficiente de Gini	0.469	0.464	0.464	0.445
---------------------	-------	-------	-------	-------

federativas, ya que se tienen los casos de Ciudad de México y Quintana Roo, donde se redujo alrededor de 5%, hasta el caso opuesto en Chihuahua, donde se incrementó 44.5% (ver gráfica 2).

3. Modelo conceptual

3.1 Modelo general de ajuste del ingreso a través de un factor de crecimiento

El modelo se conceptualiza de manera general bajo las siguientes ideas:

Consideremos dos tiempos consecutivos t_0 y t_1 , $t_0 < t_1$, y un parámetro de interés asociado al ingreso

so de los hogares a nivel nacional.⁸ Denotemos como I^{t_0} e I^{t_1} a los valores teóricos de ese parámetro de interés en t_0 y t_1 , respectivamente, entonces se podría observar que estos dos valores se encuentran relacionados tanto por un factor de crecimiento real⁹ $\Delta_{t_1}^e$ como por el efecto de inflación, es decir,

$$I^{t_1} = \Delta_{t_1}^e I^{t_0 C}, \quad (1)$$

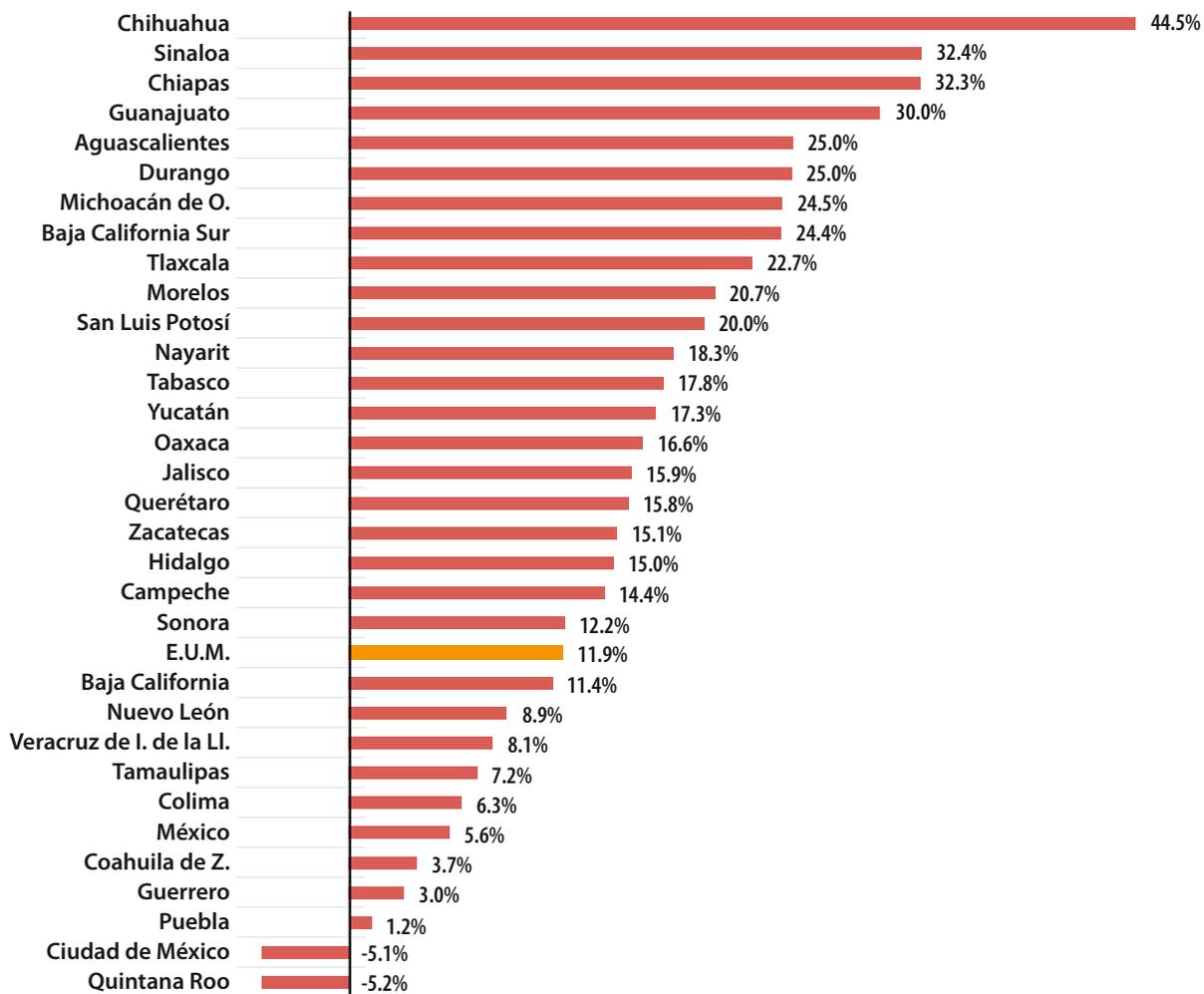
donde $I^{t_0 C}$ es el valor I^{t_0} ajustado por el efecto de inflación. Observe que si $\Delta_{t_1}^e \geq 1$, se tendría un crecimiento favorable.

8 Por ejemplo, el ingreso total, el promedio por hogar o el promedio per cápita.

9 Cabe mencionar que cada parámetro tiene asociado un diferente factor de crecimiento.

Gráfica 2

Variación del ingreso promedio del hogar por entidad federativa, 2014-2015



En la práctica, los valores de los parámetros I^{t_0} e I^{t_1} no son conocidos y, por lo tanto, se recurre al uso de encuestas que sirvan para obtener las respectivas estimaciones. Consideremos dos encuestas que en tiempos consecutivos t_0 y t_1 realizan la medición del ingreso de los hogares para obtener las estimaciones necesarias y que comparten el mismo diseño conceptual, estadístico y metodológico a lo largo de sus procesos. Entonces, las correspondientes estimaciones del ingreso en cada ejercicio a nivel nacional, definidas como \hat{I}^{t_0} e \hat{I}^{t_1} , se encontrarían, en principio, también relacionadas por el factor de crecimiento real $\Delta_{t_1}^e$ y por el efecto de inflación:

$$\hat{I}^{t_1} = \Delta_{t_1}^e \hat{I}^{t_0C} \quad (2)$$

Sin embargo, la igualdad de la ecuación (2) se puede vulnerar ante la presencia de factores exógenos que afecten, en forma de ganancia o pérdida, los ingresos de una encuesta a otra, más allá del impacto económico real y los efectos de la inflación. En general, esta aparición de factores exógenos propicia las siguientes desigualdades en la ecuación (2):

$$\frac{\Delta_{t_1}^e \hat{I}^{t_0C}}{\hat{I}^{t_1}} < 1 \text{ y } \frac{\Delta_{t_1}^e \hat{I}^{t_0C}}{\hat{I}^{t_1}} > 1. \quad (3) \text{ y } (4)$$

Es decir, hay un factor $\Delta_{t_1}^c$ que involucra a aquellos incrementos o decrementos no explicados¹⁰ tal que:

$$\hat{I}^{t_1} = \Delta_{t_1}^c \Delta_{t_1}^e \hat{I}^{t_0C} \quad (5)$$

De manera que para obtener un valor del parámetro de interés \tilde{I}^{t_1} que satisfaga la relación:

$$\tilde{I}^{t_1} = \Delta_{t_1}^e \hat{I}^{t_0C},$$

se debería de aplicar el factor $\frac{1}{\Delta_{t_1}^c}$ a la estimación \hat{I}^{t_1} .

¹⁰ Dichos incrementos no explicados podrían ser a consecuencia de una capacitación con entrevistadores más analíticos, una mayor sensibilización del informante por la técnica de entrevista, así como por los esfuerzos gubernamentales de transparentar el valor de los apoyos, entre otros. Los decrementos podrían ser consecuencia de un cambio en la declaración de ingreso del informante.

Más aún, si se asume que la relación a nivel nacional descrita en la ecuación (1) se cumple para valores del parámetro de interés correspondientes a ciertos dominios de estudio o cortes menores al nacional, entonces para esos dominios se debería satisfacer que:

$$I_d^{t_1} = \Delta_{t_1}^e I_d^{t_0C}, \quad (6)$$

donde $I_d^{t_0}$ e $I_d^{t_1}$ son los valores del parámetro de interés en el dominio d en el tiempo t_0 y t_1 , respectivamente, $I_d^{t_0C}$ es el valor $I_d^{t_0}$ con el efecto de la inflación y $\Delta_{t_1}^e$ es el factor de crecimiento real a nivel nacional.

Así, las desigualdades en (3) y (4) también podrían observarse al considerar estimaciones de los parámetros de interés, es decir, se podría identificar un factor $\Delta_{t_1,d}^c$ a nivel dominio d tal que:

$$\hat{I}_d^{t_1} = \Delta_{t_1,d}^c \Delta_{t_1}^e \hat{I}_d^{t_0C}, \quad (7)$$

donde $\hat{I}_d^{t_0}$ e $\hat{I}_d^{t_1}$ son las estimaciones del parámetro de interés en el dominio d en el tiempo t_0 y t_1 , respectivamente; $\hat{I}_d^{t_0C}$ es el valor $\hat{I}_d^{t_0}$ con el efecto de la inflación y $\Delta_{t_1}^e$ es el factor de crecimiento real a nivel nacional.

De la ecuación (7) es posible observar que para eliminar el efecto de factores no explicados se podría aplicar el factor $\frac{1}{\Delta_{t_1,d}^c}$ a la estimación $\hat{I}_d^{t_1}$ y que $\Delta_{t_1,d}^c$ se podría estimar siempre que se conozca $\Delta_{t_1}^e$.

Cabe mencionar que si $\hat{I}_d^{t_1}$ es la estimación del total del ingreso de los hogares del dominio d , entonces la aplicación del factor $\frac{1}{\Delta_{t_1,d}^c}$ se puede hacer a nivel hogar para todos los hogares de ese dominio; más aún, el factor $\frac{1}{\Delta_{t_1,d}^c}$ se puede hacer a nivel persona para todos los hogares de ese dominio; se puede decir que el factor de ajuste para una persona r del hogar contenido en el dominio d , $\frac{1}{\Delta_{t_1,dr}^c}$, cumple lo siguiente:

$$\frac{1}{\Delta_{t_1,dr}^c} = \frac{1}{\Delta_{t_1,d}^c} \quad (8)$$

3.2. Relación entre el ingreso y otras variables

Como se mencionó en el apartado anterior del modelo conceptual, es necesario encontrar un factor de crecimiento real $\Delta_{t_1}^e$, por lo que se estudiará la relación teórica entre los ingresos, gastos y ahorros, para ver en qué medida el crecimiento de los gastos y el ahorro pueden relacionarse con el crecimiento real del ingreso ($\Delta_{t_1}^e$).

Partiendo de una perspectiva de la teoría económica, el concepto de ingreso disponible es la cantidad máxima que una persona, hogar u otra unidad puede permitirse gastar en bienes o servicios de consumo sin tener que financiar sus gastos reduciendo su patrimonio (dinero, disponiendo de otros activos financieros o no financieros o incrementando sus pasivos). De esto se deduce que los conceptos de renta, consumo y ahorro se encuentran estrechamente relacionados entre sí, por lo que se requiere revisar la teoría económica de estos conceptos.

3.2.1 Relación teórica del consumo y el ahorro con el ingreso

Los individuos reparten su ingreso entre consumo y ahorro. Esta decisión crucial afecta el bienestar de las familias a través del tiempo. Quienes prefieran consumir más hoy —y, por lo tanto, ahorrar menos— tendrán que consumir menos en el futuro. Quienes consuman menos hoy —y, en consecuencia, ahoren más— podrán tener un consumo mayor en el futuro. A nivel económico agregado, los efectos combinados de las decisiones de consumo y ahorro de cada familia ayudan a determinar la tasa de crecimiento de la economía, la balanza comercial, el nivel de ingreso nacional disponible, el nivel de empleo y otras variables macroeconómicas.

El análisis se basa en gran medida en la teoría de consumo y ahorro para el ciclo de vida.¹¹ Cada indi-

11 Modigliani, Franco. "Life Cycle, Individual Thrift, and the Wealth of Nations", en: *American Economic Review*. Vol. 76, núm. 3. American Economic Association, Pittsburgh, 1986, pp. 297-313.

viduo recibe un flujo de ingreso a través de su vida y debe definir una trayectoria de consumo que sea coherente con dichos ingresos. En varios modelos formales, la vida se divide en una serie de periodos, que pueden ser años o intervalos de varios años, dependiendo del objetivo del modelo. En cualquier lapso dado, una familia puede consumir menos o más que el ingreso de ese periodo. Si consume menos y ahorra más, el monto ahorrado será usado para consumir más en algún momento futuro. Si consume más que su ingreso actual, deberá *desahorrar* (tener ahorro negativo) en el presente, con lo que reduciría su consumo futuro.

Lo que nos dice esta teoría es que las familias deciden su consumo presente en función de sus expectativas sobre su ingreso futuro, así como de la tasa de interés que pueden obtener sobre el ahorro y la tasa de interés que deben pagar cuando se endeuda. Por lo tanto, este proceso de toma de decisiones es intertemporal, es decir, las familias toman en cuenta la forma en que sus decisiones presentes afectarán sus oportunidades de consumo futuras.

3.2.2 Representación teórica de la relación del consumo y el ahorro con el ingreso

A continuación, se presentan las restricciones a las que se enfrentan los individuos en la elección intertemporal de recursos.¹² Los individuos, en cada momento, disponen de una determinada cantidad de recursos, entre los que deciden cuáles ahorrar y cuáles consumir. Esa decisión es la que hace que los consumidores se enfrenten a la llamada restricción presupuestaria intertemporal, la cual valora los recursos totales disponibles para consumir actualmente y en el futuro.

Ahora, analizamos la decisión de un consumidor en dos momentos: en el periodo 1 presenta una renta Y_1 y un consumo C_1 , de forma que para el segundo periodo obtendría una renta Y_2 y un consumo C_2 . Un detalle importante es que el consumidor tiene la

12 Larraín B., Felipe. *Macroeconomía en la economía global*. Buenos Aires, Pearson Education, 2006.

posibilidad de pedir préstamos y ahorrar, con lo cual el consumo de un periodo podría ser menor al otro. Ese ahorro o préstamo lo denotamos con S .

En el primer periodo se observa que el ahorro es igual a la renta menos el consumo:

$$S = Y_1 - C_1 \quad (i)$$

pero en el segundo periodo, el ahorro es igual a la diferencia entre el consumo y la renta entre el tipo de interés real (r):

$$S = (C_2 - Y_2) / (1 + r) \quad (ii)$$

De aquí se despeja el consumo y se obtiene:

$$C_2 = (1 + r)S + Y_2 \quad (iii)$$

Para hallar la restricción presupuestaria de ese consumidor, se sustituye la ecuación (i) en S en la ecuación (iii):

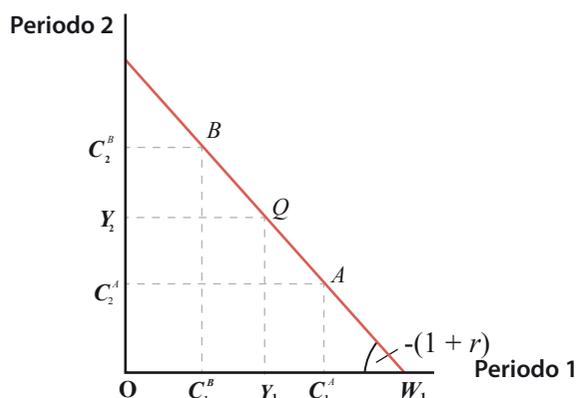
$$C_2 = (1 + r)(Y_1 - C_1) + Y_2 \quad (iv)$$

La restricción presupuestaria (ver gráfica 3) que aparece en (iv) puede representarse gráficamente de manera más clara si se reformula como:

$$C_2 = Y_2 + (1 + r)Y_1 - (1 + r)C_1 \quad (iv')$$

Gráfica 3

Representación gráfica de la restricción presupuestaria^a



^a Larrain B., Felipe. *Macroeconomía en la economía global*. Buenos Aires, Pearson Education, 2006.

Claramente, la línea que relaciona C_2 con C_1 es una recta con pendiente $-(1+r)$ que atraviesa el punto (Y_1, Y_2) . Esta línea representa todas las combinaciones posibles de consumo (C_1, C_2) que sean coherentes con la restricción presupuestaria intertemporal. La familia puede escoger cualquier par de consumos a lo largo de esta línea. Si elige el punto A de la gráfica 3, tendrá que *desahorrar* o endeudarse en el primer periodo, puesto que $C_1^A > Y_1$. Claramente, la familia será deudora neta al final de este lapso. En consecuencia, C_2^A tiene que ser inferior a Y_2 (como resulta evidente al ver el diagrama), ya que tiene que pagar la deuda. Así, la restricción presupuestaria refleja el *trade-off* intertemporal fundamental entre el consumo presente (periodo 1) y el consumo futuro (periodo 2). Con una trayectoria de ingreso disponible, si una familia decide aumentar el consumo presente, solo podrá hacerlo a expensas del consumo futuro. A la inversa, en un punto como B , la familia está limitando su consumo de hoy $C_1^B < Y_1$, con el fin de poder consumir más en una fecha posterior.

3.2.3 Relación conceptual del gasto y las erogaciones financieras con el ingreso corriente total

Buscando utilizar algunas variables captadas por la ENIGH como *proxys* de los conceptos teóricos de ingreso, consumo y ahorro, se presentan las definiciones conceptuales de algunas variables que se construyen con la información captada por esta encuesta.

Ingreso corriente total

Son varios los factores que inciden en la formación del nivel de bienestar de los hogares: el ingreso corriente total en efectivo o en bienes y servicios que reciben los integrantes del hogar es uno de los de mayor relevancia, ya que del flujo y monto de estos ingresos depende, en buena medida, el acceso regular por parte del hogar a los bienes y servicios que constituyen su canasta de consumo en un periodo dado. El ingreso corriente total se define

como el máximo de recursos que el hogar puede acceder de manera regular y que están disponibles para su consumo de bienes y servicios en un lapso dado, manteniendo inalterado el patrimonio del hogar.

Gasto monetario

El gasto es el desembolso que un hogar tiene que realizar para atender sus necesidades y cumplir con sus compromisos. El gasto corriente total da cuenta de los realizados de forma regular en el hogar para la adquisición de su canasta de consumo, además del gasto regular no destinado al consumo. Bajo este concepto no se considera el gasto que modifica el patrimonio del hogar o el de cualquiera de los integrantes de éste. En específico, el gasto corriente monetario es la suma de los gastos regulares que directamente hacen los hogares en bienes y servicios para su consumo.

Erogaciones financieras monetarias

Las erogaciones financieras y de capital monetarias no son parte constitutiva del gasto corriente de los hogares y pueden modificar el valor patrimonial de los mismos, son desembolsos destinados a la compra de bienes muebles e inmuebles, objetos valiosos, activos físicos y no físicos que modifican el patrimonio del hogar. Además, incluyen una gran variedad de operaciones que van desde la realización de depósitos bancarios hasta la adquisición de bienes inmuebles, pasando por la compra de divisas, el pago de pasivos diversos (tarjetas de crédito, deudas ante terceros, etc.) y la compra de acciones y bonos, entre otras.

Por lo tanto, es posible interpretar que los datos arrojados por la ENIGH deberían cumplir, en cierta medida, alguna igualdad del tipo:

$$S = Y_1 - C_1 \quad (i)$$

Con el objetivo de encontrar un factor de crecimiento real Δ_t^e para el ingreso de los hogares,

se utiliza la variable del ingreso corriente total (*IngCor*) como una variable *proxy* del ingreso teórico (*Y*), la de los gastos monetarios como una *proxy* del consumo teórico (*C*) y la de erogaciones financieras monetarias como una *proxy* del ahorro (*S*). La decisión de tomar estas variables como *proxys* se debe a que cada una de ellas corresponde, de acuerdo con las definiciones utilizadas en la ENIGH, como las más cercanas a los conceptos teóricos de ingreso, consumo y ahorro.

4. Modelo empírico

4.1 Objetivo y supuestos del modelo empírico

A partir del presente ejercicio se busca construir una base de datos ajustada del MCS 2015 que reduzca el efecto de los crecimientos no explicados en la medición del ingreso del 2015. Parte de la discusión del crecimiento del ingreso del 2015 está centrada en que, para dicho año, se puso mayor énfasis y cuidado en la capacitación del personal operativo y el levantamiento de información, evitando caer en situaciones que en ejercicios previos tuvieron áreas de oportunidad.

Con la intención de analizar el crecimiento antes mencionado se estudian diversas hipótesis sobre los factores que pudieron haber influido. En particular, surge la inquietud de saber si a partir de los gastos monetarios y las erogaciones financieras monetarias (como ahorros) captadas por la ENIGH (tradicional), es posible determinar una relación que logre explicar el comportamiento de una y otra. Sabemos que hubo acciones que pudieron mejorar la captación del ingreso (son hipótesis), no obstante, nunca se ha hablado de una acción para mejorar la captación de los gastos y los ahorros, situación que se podría corroborar en la ENIGH 2016.

Asumimos que la metodología de medición de los gastos sigue siendo la misma, lo que permite que, a partir de la construcción de un parámetro de control referente a la encuesta y de la homologación de las

distribuciones del ingreso de la población, se pueda hacer una adecuada comparabilidad entre ejercicios.

En este escenario, surgen las siguientes preguntas: ¿los gastos y los ahorros han tenido un cambio abrupto también como los ingresos?, o bien, ¿se puede hablar de un comportamiento estable conforme pasa el tiempo?, ya que nunca ha estado en tela de juicio si se ha captado de manera diferente o no el gasto y los ahorros de los hogares.

Si la primera pregunta se contestara de manera afirmativa en el 2016, tendríamos el mismo problema que en el MCS 2015 donde podríamos concluir que tanto el ingreso como los gastos y los ahorros tuvieron un cambio respecto a ejercicios anteriores. Pero si la segunda pregunta es la que se responde así, entonces podríamos hablar de una variable de control que nos permitirá estimar cuánto fue lo que se debió captar en el ingreso para el periodo 2016, ajustando 2015 con la distribución de 2014 y un factor de crecimiento anclado en los gastos, como un indicador del cambio que se debió observar entre la ENIGH-MCS 2014 y el MCS 2015.

La metodología empleada en la construcción de este ajuste respeta el diseño estadístico tanto de la ENIGH-MCS 2014 como del MCS 2015. No se imputan valores relacionados con las estructuras demográficas que estima cada encuesta, sino que se aborda de forma directa la imputación de ingresos (*imputación basada en modelos*, ver anexo 7.1), desde las claves de fuente de ingreso donde fue captada la información. Una vez que los ingresos son imputados, nuevamente se recalculan los promedios trimestrales y se reconstruye el concentrado del hogar, con el fin de explotar los resultados.

A lo largo del documento, se emplea la siguiente terminología:

2014: cifras publicadas provenientes de la ENIGH-MCS 2014.

2015: cifras publicadas provenientes del MCS 2015.

2015A: cifras ajustadas provenientes del MCS 2015.

4.2 Determinación del factor de crecimiento real del ingreso

Se busca estimar el ingreso corriente tradicional de la ENIGH 2016 considerando las tendencias de él y de los gastos monetarios, así como las erogaciones financieras monetarias que se han observado históricamente en los proyectos de la ENIGH, de 1992 al 2014. Bajo este pronóstico 2016, se asumirá un punto medio de dicho crecimiento para el 2015.

A partir de los ingresos corrientes y del agregado de los gastos monetarios y erogaciones financieras monetarias de las ENIGH 1992-2014, construidos de manera tradicional, se hizo un análisis de cómo estas variables se han comportado con el tiempo, el cual se desarrolló desde el ejercicio de 1992 (nuevos pesos) hasta el del 2014, sin considerarse los ejercicios correspondientes a 1984, 1989 y el 2005, ya que se trata de levantamientos que no concuerdan con la aplicación bienal (ver esta serie en la gráfica 4).

En primera instancia, se observa que el ingreso y el gasto tienen un comportamiento muy similar en tendencia, lo cual se refleja en un coeficiente de correlación de 0.9554.

Derivado de la alta correlación, se construye un modelo lineal por mínimos cuadrados del ingreso corriente (*IngCor*) a partir del acumulado del gasto monetario y las erogaciones financieras monetarias (*GasEroMon*). Se busca crear un modelo para predecir el ingreso, dada la información del gasto. En particular, se tiene el interés de pronosticar el ingreso 2016, una vez observado el gasto de dicho ejercicio. Lo que nos ayudaría a también predecir el ingreso 2015, aun cuando el proyecto no haya medido gastos por tratarse solo del MCS (ver gráfica 5).

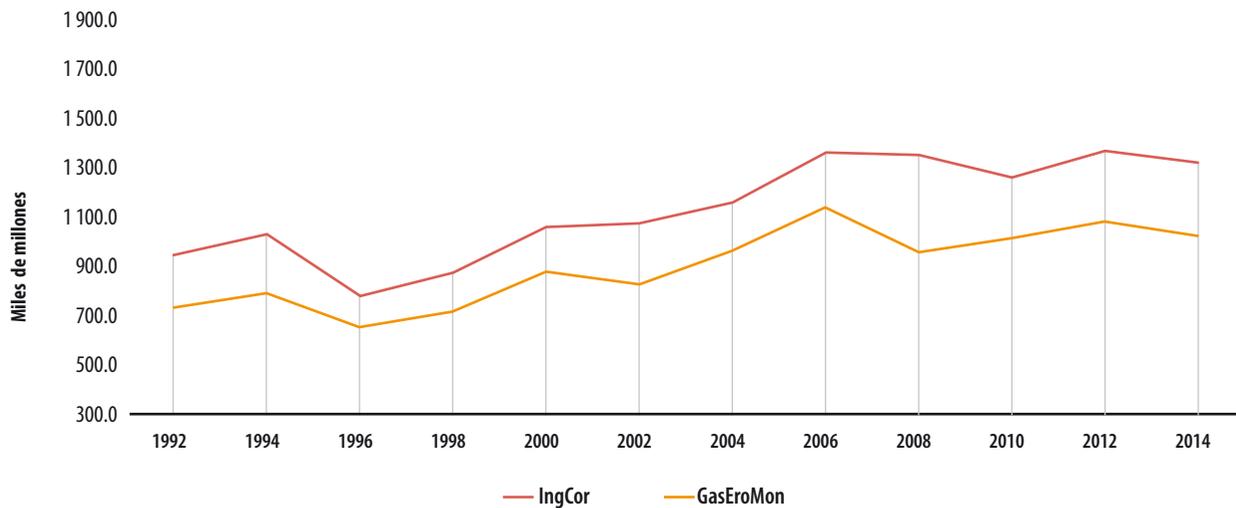
La estimación de modelo lineal está dada por:

$$\widehat{y}_{IngCor} = 1.0142 * 10^{10} + 1.2481694 * x_{GasEroMon} \quad (I)$$

Dado que los resultados de la ENIGH 2016 no han sido publicados, es necesario estimar un gasto ficticio (provisional) que pueda ser implementado en el modelo (I).

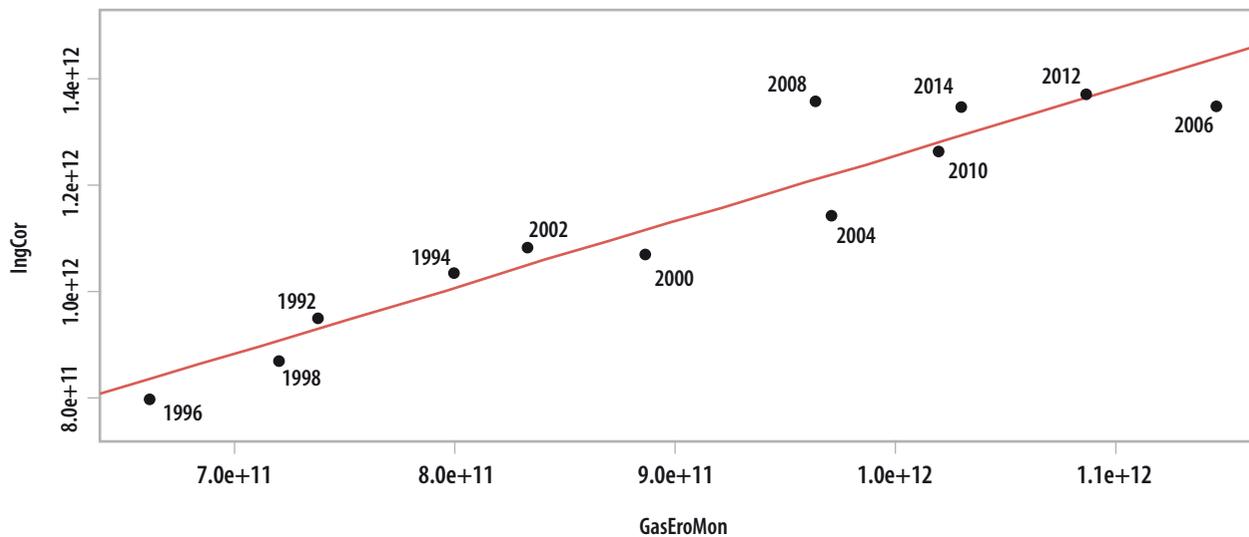
Gráfica 4

Ingreso corriente total (*IngCor*) y gasto monetario más erogaciones financieras monetarias (*GasEroMon*). Precios constantes del 2016



Gráfica 5

Ajuste del modelo lineal del ingreso corriente total (*IngCor*) y del gasto monetario más erogaciones financieras monetarias (*GasEroMon*). Precios constantes del 2016



Dicha estimación se realiza de manera provisoria por medio de la relación lineal que existe entre el gasto y los años de realización de los proyectos ENIGH (ver anexo 7.2), resultando en:

$$x_{\widehat{GasEroMon}_{2016}} = 1,117,011,119,366.3$$

Con lo anterior, y bajo el modelo lineal (I), el pronóstico del ingreso corriente del ejercicio 2016 resultó en:

$$\widehat{IngCor} = 1,404,361,305,011.5$$

Este ingreso representa un crecimiento de 5.7% con respecto al ingreso observado en la ENIGH 2014 de 1,328,629,427,636.3.

Ese 5.7% será considerado como factor de crecimiento del 2014 al 2016 y, por consiguiente, será utilizado en su valor medio para el crecimiento del 2014 al 2015, por lo tanto el $\Delta_{t_1}^e$ será de 2.81 por ciento.

4.3 Aplicación del modelo empírico

4.3.1 Características generales del ajuste de los ingresos

La idea general es estimar de alguna forma el factor $\Delta_{t_1}^e$ asociado al crecimiento real nacional, para entonces calcular $\frac{1}{\Delta_{t,d}^c}$ descrito en la sección 3.1, para algunos dominios, considerando los levantamientos del MCS 2015 y la ENIGH-MCS 2014. Posteriormente, cada factor se aplicará al valor reportado por rubros de ingreso de todos los hogares pertenecientes a ciertos cortes, de manera que se obtendrá un nuevo vector de valores de ingresos para cada rubro considerado. Al final, se contará con una base de datos ajustada del MCS 2015 donde se elimina, en lo posible, el efecto de factores exógenos del ingreso. Algunos aspectos importantes de la metodología son:

- El parámetro considerado es el ingreso corriente total de los hogares.
- La estimación de $\Delta_{t_1}^e$ se obtiene a partir de la relación entre el ingreso total de los hogares y el agregado del gasto monetario y erogaciones financieras monetarias. Como se vio en la sección 4.2, este factor se estima en $\Delta_{15}^e = 1.0281$.
- Un aspecto importante a considerar fue que en el MCS 2015 la cantidad de hogares que no reportaron ingresos por transferencias fue mucho menor con respecto al ejercicio ENIGH-MCS 2014, por lo que se optó agregar en el MCS 2015 de manera aleatoria para el rubro de transferencias ingresos con valor cero, de modo que la cantidad de ceros entre ejercicios fuera aproximadamente la misma.
- Dado que las diferencias entre el ingreso en el decil X del MCS 2015 y el ingreso en el decil X

del ENIGH-MCS 2014 no son significativas (ver Anexo 7.3), el ajuste al ingreso del MCS 2015 por factor de crecimiento con base en un indicador del gasto se aplicará solo en los deciles del I al IX.

- La modificación de los ingresos considera que el factor $\Delta_{t_1}^e$ estimado a nivel nacional puede ser aplicado en cortes menores al nacional, por ejemplo, entidades federativas, es decir, se calculará para ajustar los ingresos del MCS 2015 un factor $\frac{1}{\Delta_{t,d}^c}$ para ciertos dominios de interés. Una alternativa sería hacer la modificación a nivel nacional partiendo de la ecuación (5) pero, de manera desafortunada, se estaría asumiendo que todo creció de la misma forma, independientemente de las zonas geográficas o la concentración poblacional, por lo que parece conveniente diferenciar dependiendo de ciertas características. En específico, aquí se considera lo siguiente:

– Además, los dominios a considerar toman en cuenta desagregaciones formadas por las entidades federativas ($i = 1, \dots, 32$), por tipo de localidad rural o urbana ($j = 1, 2$) y por estrato socioeconómico ($k = 1, \dots, 4$).

Por lo tanto, los dominios son 256, formados por los hogares en los cruces entidad federativa, tipo de localidad y estrato socioeconómico.

- Aun cuando el factor $\Delta_{t_1}^e$ se estima a partir del ingreso total de los hogares, se asume que la ecuación (7) se cumple cuando las estimaciones corresponden a los totales del ingreso de los hogares de las siguientes fuentes de ingreso: sueldo por trabajo subordinado, remuneraciones en especie, ingresos por trabajo independiente, otros ingresos por trabajo sea o no subordinado, ingresos provenientes por sociedades, jubilaciones, donativos, remesas, transferencias de instituciones y transferencias de otros hogares. Así, la búsqueda de un factor de ajuste se hace para cada rubro, es decir, se buscará un factor $\frac{1}{\Delta_{t,d}^c}$, $l = 1, \dots, 10$, asociado a cada una de las diez fuentes de ingreso al interior del dominio d . Asimismo, se tendrá el factor de ajuste $\Delta_{t_1,d}^c$, $l = 1, \dots, 10$ para el periodo contrario.

4.3.2 Construcción del factor de crecimiento real y la base de datos ajustada

Consideremos la siguiente notación:

Datos originales

x_{ijklr}^{14c} es el ingreso reportado de la fuente de ingreso l en el r -ésimo registro perteneciente a un hogar contenido en la entidad i , tipo de localidad j y estrato socioeconómico k correspondiente al ENIGH-MCS 2014 en precios constantes del 2015.

x_{ijklr}^{15} es el ingreso reportado de la fuente de ingreso l en el r -ésimo registro perteneciente a un hogar contenido en la entidad i , tipo de localidad j y estrato socioeconómico k correspondiente al MCS 2015.

Datos ajustados

x_{ijklr}^{15A} es el ingreso reportado de la fuente de ingreso l en el r -ésimo registro perteneciente a un hogar contenido en la entidad i , tipo de localidad j y estrato socioeconómico k correspondiente al MCS 2015 en su versión ajustada.

Se define \hat{I}_{ijkl}^{14C} como el ingreso total expandido de los hogares, asociado a la fuente de ingreso l en la entidad i , tipo de localidad j y estrato socioeconómico k correspondiente a la ENIGH-MCS 2014 en precios del 2015, es decir:

$$\hat{I}_{ijkl}^{14C} = \sum_r W_{ijkr}^{14} x_{ijklr}^{14C},$$

donde w_{ijkr}^{14} es el factor de expansión asociado al hogar contenido en la entidad i , tipo de localidad j y estrato socioeconómico k .

De igual manera, sea \hat{I}_{ijkl}^{15} el ingreso total expandido de los hogares asociados a la fuente de ingreso l en la entidad i , tipo de localidad j y estrato socioeconómico k correspondiente al MCS 2015:

$$\hat{I}_{ijkl}^{15} = \sum_r W_{ijkr}^{15} x_{ijklr}^{15}.$$

Para cada fuente de ingreso l , entidad i , tipo de localidad j y estrato socioeconómico k consideremos el siguiente coeficiente f_{ijkl} :

$$f_{ijkl} = \begin{cases} \frac{\Delta_{15}^e * \hat{I}_{ijkl}^{14C}}{\hat{I}_{ijkl}^{15}}, & \text{siempre que } \hat{I}_{ijkl}^{14C} \neq 0 \text{ y } \hat{I}_{ijkl}^{15} \neq 0. \\ 1, & \text{en otro caso.} \end{cases}$$

Entonces, se considera la siguiente estimación de $\frac{1}{\Delta_{15}^c}$, como:

$$\frac{1}{\Delta_{15}^c} = f_{ijkl}$$

Luego, el factor de ajuste del r -ésimo registro perteneciente a un hogar para la fuente de ingreso l en la entidad i y tipo de localidad j estará dado por:

$$x_{ijklr}^{15A} = \frac{1}{\Delta_{15}^c} x_{ijklr}^{15}. \quad (9)$$

Cada factor $\frac{1}{\Delta_{15}^c}$ se deberá aplicar en todas las bases de datos públicas, a todos y cada uno de los registros de ingreso del MCS 2015, asociados a la fuente l , del r -ésimo registro perteneciente a un hogar en la entidad i , tipo de localidad j y estrato socioeconómico k . Es decir, la aplicación de factor de ajuste para una fuente de ingreso en un hogar determinado consiste en aplicar la modificación a todas las claves de ingreso o gasto que se usan para construir la totalidad del ingreso en ese rubro que presenta ese hogar. Así, se reconstruye la base de datos del MCS 2015 en su versión ajustada (2015A).

5. Resultados

5.1 MCS 2015 ajustado

Principales estimaciones del 2015A

Conforme a los resultados del ajuste del MCS 2015 (2015A), el ingreso corriente promedio por hogar registró un incremento de 2.5% respecto al 2014, en comparación con el 11.9% que presenta el incremento original 2014-2015. Por deciles de

Tabla 4

Ingreso corriente total promedio trimestral por hogar, según decil, 2010-2012-2014-2015-2015A. Precios del 2015

Deciles de hogares	Total nacional					Variación porcentual			
	2010	2012	2014	2015	2015A	2010-2012	2012-2014	2014-2015	2014-2015A
Total	42 725	40 959	41 000	45 887	42 021	-4.1	0.1	11.9	2.5
I	5 944	5 757	6 116	8 169	6 247	-3.1	6.2	33.6	2.1
II	11 386	10 999	11 409	13 830	11 402	-3.4	3.7	21.2	-0.1
III	15 838	15 318	15 622	18 387	15 527	-3.3	2.0	17.7	-0.6
IV	20 303	19 768	19 778	23 147	19 762	-2.6	0.1	17.0	-0.1
V	25 382	24 662	24 564	28 478	24 533	-2.8	-0.4	15.9	-0.1
VI	31 511	30 637	30 178	34 866	30 164	-2.8	-1.5	15.5	0.0
VII	39 245	38 264	37 433	43 030	37 260	-2.5	-2.2	15.0	-0.5
VIII	50 600	49 108	48 105	54 139	47 296	-2.9	-2.0	12.5	-1.7
IX	70 588	67 419	66 510	73 252	66 152	-4.5	-1.3	10.1	-0.5
X	156 457	147 662	150 283	161 568	161 861	-5.6	1.8	7.5	7.7

Tabla 5

Ingreso corriente total trimestral por hogar, según decil, 2010-2012-2014-2015-2015A. Precios del 2015

Deciles	Total nacional					Variación porcentual			
	2010	2012	2014	2015	2015A	2010-2012	2012-2014	2014-2015	2014-2015A
Total	1 266 332 790	1 284 468 404	1 318 162 696	1 524 262 517	1 395 840 892	1.4	2.6	15.6	5.9
I	17 616 405	18 054 994	19 662 135	27 136 934	20 750 627	2.5	8.9	38.0	5.5
II	33 747 047	34 493 364	36 680,440	45 939 106	37 876 513	2.2	6.3	25.2	3.3
III	46 941 548	48 036 726	50 224,306	61 078 947	51 577 579	2.3	4.6	21.6	2.7
IV	60 175 078	61 990 692	63 587 807	76 890 425	65 646 571	3.0	2.6	20.9	3.2
V	75 230 196	77 337 598	78 975 556	94 598 336	81 494 465	2.8	2.1	19.8	3.2
VI	93 393 786	96 077 197	97 023 728	115 817 953	100 200 002	2.9	1.0	19.4	3.3
VII	116 317 057	119 993 092	120 349 274	142 938 032	123 770 657	3.2	0.3	18.8	2.8
VIII	149 973 705	154 001 584	154 660 631	179 837 550	157 108 950	2.7	0.4	16.3	1.6
IX	209 216 483	211 422 397	213 831 421	243 328 564	219 744 410	1.1	1.1	13.8	2.8
X	463 721 485	463 060 760	483 167 398	536 696 669	537 671 118	-0.1	4.3	11.1	11.3

hogares, destaca el decil I con un valor de 2.1%, mientras que el incremento original fue de 33.6% (ver tabla 4).

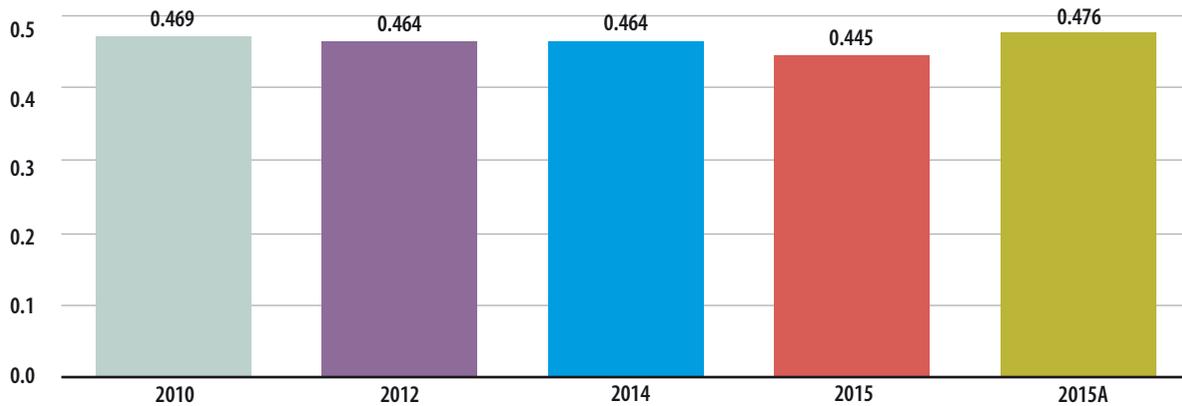
En cuanto al ingreso corriente por hogar para el 2015A, se registró un incremento de 5.9% respecto al 2014, en comparación con 15.6% que presenta el incremento original 2014-2015. Por deciles de hogar se observa que, en todos, hubo un incremento en

favor del 2015A, en particular destaca el decil I con un valor de 5.5%, mientras que el incremento original fue de 38% (ver tabla 5).

En el 2015A, el coeficiente de Gini con las cifras ajustadas para medir la concentración del ingreso fue de 0.476. El rango de variación de este coeficiente es de 0 a 1; entre más se acerca a 1 significa mayor desigualdad en la distribución del ingreso

Gráfica 6

Índice de Gini, 2010-2012-2014-2015-2015A



y viceversa, entre más cercano esté del 0, implica una distribución del ingreso más igualitaria (ver gráfica 6).

El ingreso corriente total aumentó 5.89% (2014-2015A) en contraste con 15.64% (2014-2015). Al hacer un análisis por fuente de ingreso, se observa que en todos los casos se registró un incremento en el periodo 2014-2015A, salvo en *Otros ingresos corrientes*.

En particular, destaca el incremento del ingreso por *Ingreso del trabajo*, con 3.17 por ciento. Cabe señalar que, con la cifra del 2015 sin el ajuste, tal incremento fue de 12.57% (ver tabla 6).

Al considerar el ámbito geográfico, se observa que, para el 2015A, la participación del ámbito urbano en el ingreso corriente trimestral es de 87.7%, frente al 12.3% del área rural.

Tabla 6

Ingreso corriente total trimestral de los hogares por las principales fuentes del ingreso 2010-2012-2014-2015-2015A. Miles de pesos, precios constantes del 2015

Principales fuentes del ingreso	Total nacional									
	2010	2012	2014	2015	2015A	Variación 2010-2012	Variación 2012-2014	Variación 2014-2015	Variación 2014-2015A	Participación 2015A
Ingreso corriente	1 266 332 790	1 284 468 404	1 318 162 696	1 524 262 517	1 395 840 892	1.4%	2.6%	15.6%	5.9%	
Ingreso del trabajo	867 921 751	895 761 799	908 604 869	1 022 789 609	937 388 846	3.2%	1.4%	12.6%	3.2%	67.2%
Remuneraciones por trabajo subordinado	730 774 271	740 384 957	778 967 567	858 777 632	798 827 540	1.3%	5.2%	10.3%	2.6%	
Ingresos por trabajo independiente	104 175 265	119 428 945	92 637 747	118 533 617	101 074 931	14.6%	-22.4%	27.9%	9.1%	
Ingresos de otros trabajos	32 972 215	35 947 897	36 999 555	45 478 360	37 486 374	9.0%	2.9%	22.9%	1.3%	
Renta de la propiedad	55 227 256	45 565 331	63 163 701	108 847 937	97 867 742	-17.5%	38.6%	72.3%	54.9%	7.0%
Transferencias	177 972 618	185 718 344	183 278 356	225 206 403	193 165 737	4.4%	-1.3%	22.9%	5.4%	13.8%
Estimación del alquiler de la vivienda	163 573 381	155 918 863	161 664 054	166 083 256	166 083 256	-4.7%	3.7%	2.7%	2.7%	11.9%
Otros ingresos corrientes	1 637 784	1 504 067	1 451 716	1 335 312	1 335 312	-8.2%	-3.5%	-8.0%	-8.0%	0.1%

Tabla 7

Hogares y su ingreso corriente trimestral por deciles de hogares según tamaño de localidad, 2015A. Miles de pesos

Deciles de hogares	Total		Tamaño de localidad					
			De 2 500 y más habitantes		De menos de 2 500 habitantes		Participación porcentual	
	Hogares	Ingreso	Hogares	Ingreso	Hogares	Ingreso	Urbano	Rural
Ingreso corriente	33 218 037	1 395 840 892	25 932 179	1 223 667 989	7 285 858	172 172 903	87.7%	12.3%
I	3 321 803	20 750 627	1 613 476	10 196 704	1 708 327	10 553 923	49.1%	50.9%
II	3 321 803	37 876 513	2 003 085	22 976 350	1 318 718	14 900 163	60.7%	39.3%
III	3 321 803	51 577 579	2 324 006	36 159 395	997 797	15 418 184	70.1%	29.9%
IV	3 321 803	65 646 571	2 494 300	49 359 840	827 503	16 286 731	75.2%	24.8%
V	3 321 803	81 494 465	2 649 086	65 081 701	672 717	16 412 764	79.9%	20.1%
VI	3 321 803	100 200 002	2 786 338	84 150 025	535 465	16 049 977	84.0%	16.0%
VII	3 321 803	123 770 657	2 880 818	107 389 195	440 985	16 381 462	86.8%	13.2%
VIII	3 321 803	157 108 950	2 971 983	140 646 775	349 820	16 462 175	89.5%	10.5%
IX	3 321 803	219 744 410	3 098 934	205 507 681	222 869	14 236 729	93.5%	6.5%
X	3 321 810	537 671 118	3 110 153	502 200 325	211 657	35 470 793	93.4%	6.6%

Por otro lado, en la medida en que aumenta el número de decil, la brecha entre los dos ámbitos se va abriendo; en particular, es notorio que para el decil I la participación en los dos ámbitos tiende a ser la misma (ver tabla 7).

Por entidad federativa, el ingreso promedio por hogar en el 2015A fue mayor en Nuevo León con 63 734 pesos, seguida de Ciudad de México, Baja California Sur y Aguascalientes con 57 941, 57 193 y 55 234 pesos, respectivamente. En el 2014, las que tuvieron mayor promedio por hogar son Ciudad de México, Nuevo León, Sonora y Coahuila de Zaragoza.

Por otro lado, en el 2015A, las entidades federativas con menores ingresos promedio por hogar al trimestre son Guerrero (24 261 pesos), Oaxaca (25 525), Chiapas (25 707) y Veracruz de Ignacio de la Llave (28 361). Cabe hacer mención que en el 2014 figuraron las mismas entidades, pero en diferente orden: Chiapas, Oaxaca, Guerrero y Veracruz de Ignacio de la Llave (ver tabla 8 y mapas 1 y 2).

5.2 Resultados de pobreza 2015

Con el ajuste de cifras que se realizó (ver tabla 9), el porcentaje de la población en situación de pobreza por ingresos es de 51.6% (en el 2014, según el CONEVAL, fue de 53.2%).

En las gráficas 7-11 se presentan los diferentes grupos de pobreza. En todas se observa una caída notable del 2015 respecto al 2014; una situación similar se muestra con los datos ajustados, pero con menos pronunciamiento; en particular, la pobreza que experimentó menos diferencia es la moderada multidimensional al pasar de 36.6 a 36% en el periodo de referencia, mientras que la pobreza multidimensional pasó de 46.2 a 44.1% en el mismo periodo.

Para el 2015, con los datos ajustados, los tres estados que registran ingreso menor a la línea de bienestar son Chiapas, Oaxaca y Guerrero. Por su parte, aunque en Chiapas su porcentaje bajó un

Continúa

Tabla 8

Ingreso corriente trimestral y su promedio por hogar según entidad federativa 2010-2012-2014-2015-2015A. Precios constantes del 2015

Entidad federativa	2010		2012		2014		2015		2015A		Variación % de promedios 2010-2012	Variación % de promedios 2012-2014	Variación % de promedios 2014-2015	Variación % de promedios 2014-2015A
	Total	Promedio												
Nacional														
Hogares	29 638 940		31 359 562		32 150 400		33 218 037		33 218 037					
Ingreso	1 266 332 790	42 725	1 284 468 404	40 959	1 318 162 696	41 000	1 524 262 517	45 887	1 395 840 892	42 021	-4.1	0.1	11.9	2.5
Aguascalientes	14 694 649	50 062	14 119 873	44 149	15 827 426	47 529	20 307 882	59 434	18 872 868	55 234	-11.8	7.7	25.0	16.2
Baja California	51 995 409	57 882	47 322 123	49 258	48 000 482	49 017	57 583 712	54 616	54 966 145	52 133	-14.9	-0.5	11.4	6.4
Baja California Sur	10 029 793	53 370	10 949 069	52 094	10 643 988	49 327	14 170 602	61 361	13 207 976	57 193	-2.4	-5.3	24.4	15.9
Campeche	8 880 507	40 598	10 212 176	42 933	10 648 233	44 388	12 354 679	50 782	11 462 254	47 114	5.8	3.4	14.4	6.1
Coahuila de Zaragoza	36 691 372	49 181	37 262 644	46 528	41 459 760	51 407	44 351 683	53 301	41 063 894	49 350	-5.4	10.5	3.7	-4.0
Colima	8 084 467	43 978	8 640 537	43 599	9 271 454	45 443	10 425 429	48 293	9 432 344	43 693	-0.9	4.2	6.3	-3.9
Chiapas	24 975 850	22 022	29 218 683	24 118	28 057 903	22 417	37 222 278	29 648	32 274 058	25 707	9.5	-7.1	32.3	14.7
Chihuahua	38 585 948	38 184	46 341 588	42 886	41 334 240	38 778	61 403 572	56 015	55 416 841	50 554	12.3	-9.6	44.5	30.4
Ciudad de México	176 089 134	68 106	149 443 467	56 884	167 424 324	64 907	164 607 768	61 622	154 775 451	57 941	-16.5	14.1	-5.1	-10.7
Durango	13 808 265	32 730	14 687 057	33 552	15 159 534	33 364	19 641 012	41 709	17 523 223	37 211	2.5	-0.6	25.0	11.5
Guanajuato	49 105 960	35 695	56 622 156	41 158	51 870 744	35 739	68 309 232	46 457	61 210 714	41 629	15.3	-13.2	30.0	16.5
Guerrero	22 679 653	26 965	21 981 889	23 848	24 302 018	26 782	26 583 349	27 584	23 380 642	24 261	-11.6	12.3	3.0	-9.4
Hidalgo	21 221 024	31 186	23 694 700	32 263	25 506 279	32 792	29 856 155	37 707	26 312 407	33 231	3.5	1.6	15.0	1.3
Jalisco	89 657 305	47 758	90 792 369	46 034	101 479 760	47 369	116 734 331	54 917	108 521 183	51 053	-3.6	2.9	15.9	7.8
México	176 275 848	46 601	175 041 196	43 505	169 070 578	39 867	186 528 471	42 114	174 811 539	39 468	-6.6	-8.4	5.6	-1.0
Michoacán de Ocampo	38 419 265	34 084	38 512 626	32 859	36 513 522	30 611	48 036 424	38 112	41 820 133	33 180	-3.6	-6.8	24.5	8.4

Tabla 8

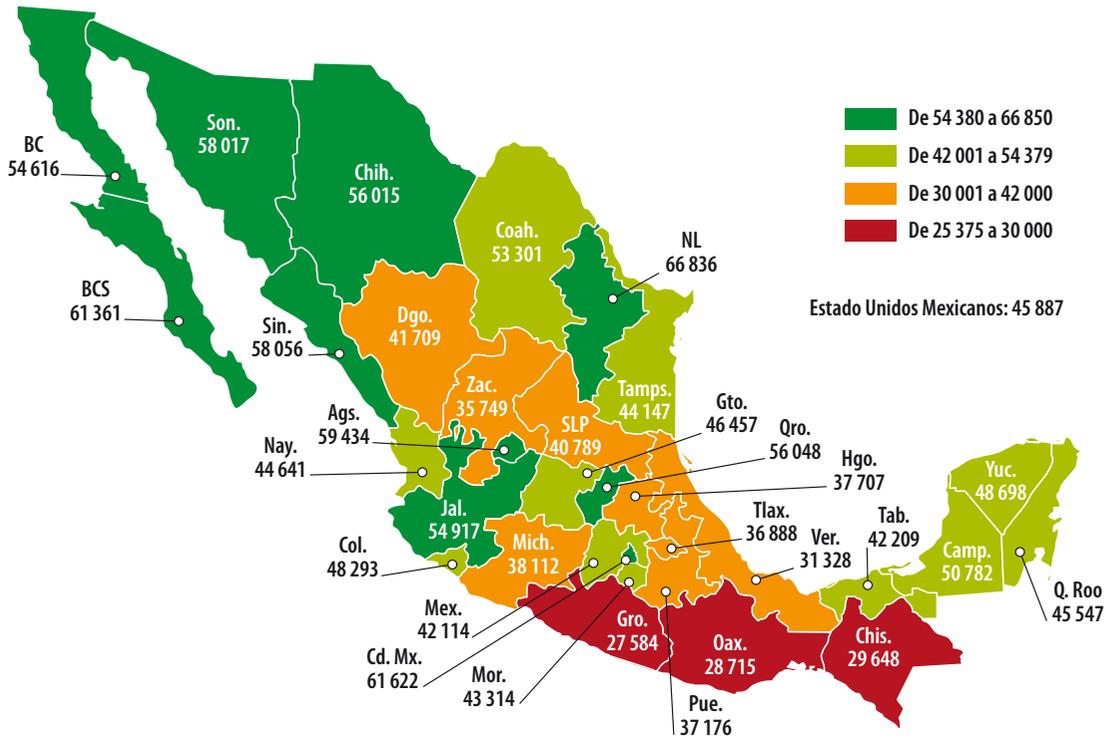
Ingreso corriente trimestral y su promedio por hogar según entidad federativa 2010-2012-2014-2015-2015A. Precios constantes del 2015

Concluye

Entidad federativa	2010		2012		2014		2015		2015A		Variación % de promedios 2010-2012	Variación % de promedios 2012-2014	Variación % de promedios 2014-2015	Variación % de promedios 2014-2015A
	Total	Promedio												
Morelos	18 345 773	37 483	19 498 574	37 761	18 943 235	35 894	23 989 154	43 314	20 933 126	37 796	0.7	-4.9	20.7	5.3
Nayarit	11 925 016	38 920	11 185 038	34 639	12 890 016	37 753	15 750 767	44 661	14 447 429	40 965	-11.0	9.0	18.3	8.5
Nuevo León	86 116 750	67 127	84 040 942	62 218	84 022 504	61 376	95 989 485	66 836	91 534 358	63 734	-7.3	-1.4	8.9	3.8
Oaxaca	25 200 294	26 150	28 505 054	27 158	25 427 144	24 629	31 694 385	28 715	28 173 588	25 525	3.9	-9.3	16.6	3.6
Puebla	45 657 234	32 680	45 359 708	30 067	55 436 684	36 723	58 163 243	37 176	51 219 140	32 738	-8.0	22.1	1.2	-10.9
Querétaro	21 736 787	45 671	24 695 445	48 030	25 363 440	48 394	30 913 391	56 048	27 784 790	50 376	5.2	0.8	15.8	4.1
Quintana Roo	18 714 720	48 008	17 759 870	43 276	21 107 306	48 032	21 448 986	45 547	19 595 242	41 610	-9.9	11.0	-5.2	-13.4
San Luis Potosí	23 330 638	35 197	24 225 463	34 364	24 427 739	33 991	30 782 761	40 789	27 026 146	35 811	-2.4	-1.1	20.0	5.4
Sinaloa	31 928 603	42 818	33 595 313	42 583	35 479 201	43 848	48 233 394	58 056	44 723 805	53 832	-0.5	3.0	32.4	22.8
Sonora	36 674 861	49 985	43 443 889	53 234	42 713 895	51 726	48 961 363	58 017	45 514 757	53 933	6.5	-2.8	12.2	4.3
Tabasco	19 005 447	32 327	22 923 123	37 267	22 456 115	35 834	27 008 573	42 209	24 908 549	38 927	15.3	-3.8	17.8	8.6
Tamaulipas	37 941 859	40 196	40 760 337	41 851	41 570 946	41 170	46 148 327	44 147	41 376 287	39 582	4.1	-1.6	7.2	-3.9
Tlaxcala	8 445 721	28 954	9 199 961	31 047	9 197 370	30 055	11 647 059	36 888	9 745 716	30 867	7.2	-3.2	22.7	2.7
Veracruz de Ignacio de la Llave	68 509 647	32 761	70 684 188	31 376	66 122 690	28 977	71 839 021	31 328	65 035 202	28 361	-4.2	-7.6	8.1	-2.1
Yucatán	20 116 140	38 877	20 595 532	38 303	23 038 998	41 510	28 362 848	48 698	25 135 677	43 157	-1.5	8.4	17.3	4.0
Zacatecas	11 488 851	29 138	13 153 815	31 578	13 395 168	31 061	15 213 182	35 749	13 635 410	32 042	8.4	-1.6	15.1	3.2

Mapa 1

Promedio de ingreso por hogar según entidad federativa, 2015



Mapa 2

Promedio de ingreso por hogar según entidad federativa, 2015A

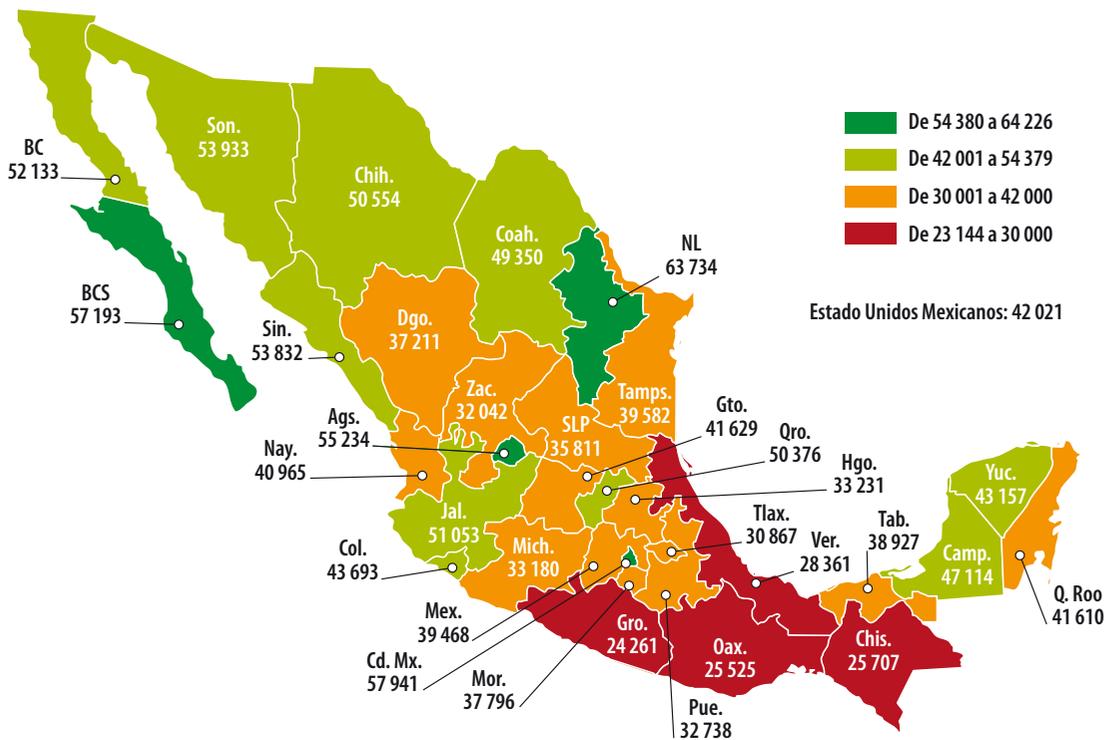


Tabla 9

Continúa

Porcentaje de la población por entidad federativa y tipo de pobreza, 2015A

Año	2014						2015A						Diferencias			
	pobreza	pobreza_e	vul_car	vul_ing	plb	pobreza	pobreza_e	vul_car	vul_ing	plb	Δ pobreza	Δ pobreza_e	Δ vul_car	Δ vul_ing	Δ plb	
Entidad federativa																
Aguascalientes	34.8%	2.1%	25.3%	8.9%	43.7%	32.5%	2.3%	26.4%	10.6%	43.1%	-6.5%	8.7%	4.5%	18.9%	-1.3%	
Baja California	28.6%	3.1%	38.3%	6.6%	35.2%	28.6%	2.1%	34.4%	7.6%	36.2%	0.2%	-32.3%	-10.3%	15.4%	3.0%	
Baja California Sur	30.3%	3.9%	34.4%	5.6%	35.9%	27.3%	2.4%	35.5%	6.6%	33.9%	-9.8%	-39.6%	3.2%	17.4%	-5.5%	
Campeche	43.6%	11.1%	32.1%	4.0%	47.6%	42.2%	7.5%	31.8%	4.0%	46.3%	-3.1%	-32.6%	-0.7%	0.9%	-2.8%	
Coahuila de Zaragoza	30.2%	3.7%	24.2%	11.1%	41.3%	25.2%	2.7%	24.8%	14.2%	39.4%	-16.5%	-28.1%	2.5%	27.7%	-4.6%	
Colima	34.3%	3.4%	33.2%	6.4%	40.7%	33.6%	3.2%	33.5%	7.7%	41.3%	-2.2%	-6.9%	0.8%	20.7%	1.4%	
Chiapas	76.2%	31.8%	15.3%	2.5%	78.7%	73.8%	31.7%	15.9%	2.6%	76.3%	-3.2%	-0.5%	4.4%	2.6%	-3.0%	
Chihuahua	34.4%	5.4%	25.8%	12.0%	46.4%	30.0%	1.9%	26.0%	11.9%	41.9%	-12.8%	-64.9%	0.7%	-0.4%	-9.6%	
Ciudad de México	28.4%	1.7%	28.0%	8.0%	36.4%	30.8%	1.2%	23.8%	9.8%	40.6%	8.6%	-31.7%	-14.9%	22.7%	11.7%	
Durango	43.5%	5.3%	24.3%	10.5%	53.9%	41.8%	3.9%	25.7%	10.1%	51.9%	-3.9%	-27.4%	5.7%	-3.2%	-3.8%	
Guanajuato	46.6%	5.5%	27.7%	8.6%	55.1%	41.5%	4.2%	32.3%	5.9%	47.4%	-10.8%	-24.1%	16.7%	-31.0%	-14.0%	
Guerrero	65.2%	24.5%	26.2%	2.6%	67.9%	66.7%	23.0%	25.2%	2.2%	68.9%	2.3%	-6.1%	-3.7%	-18.1%	1.5%	
Hidalgo	54.3%	12.3%	25.8%	5.1%	59.4%	51.8%	9.9%	29.0%	4.9%	56.7%	-4.6%	-19.5%	12.3%	-3.9%	-4.5%	
Jalisco	35.4%	3.2%	29.6%	7.9%	43.3%	32.2%	3.0%	31.5%	7.3%	39.5%	-9.1%	-5.8%	6.2%	-7.2%	-8.7%	
México	49.6%	7.2%	23.7%	9.3%	58.9%	44.7%	5.8%	23.9%	10.9%	55.6%	-9.9%	-19.8%	1.0%	16.8%	-5.7%	
Michoacán de Ocampo	59.2%	14.0%	25.0%	4.1%	63.3%	56.4%	11.6%	26.7%	4.5%	60.9%	-4.9%	-17.1%	7.0%	11.2%	-3.8%	
Morelos	52.3%	7.9%	25.7%	6.1%	58.4%	48.6%	8.8%	26.7%	6.1%	54.8%	-7.0%	12.7%	3.6%	0.1%	-6.3%	
Nayarit	40.5%	8.5%	29.4%	6.5%	47.0%	37.4%	5.9%	32.9%	7.9%	45.3%	-7.7%	-30.2%	11.7%	21.1%	-3.7%	

Tabla 9

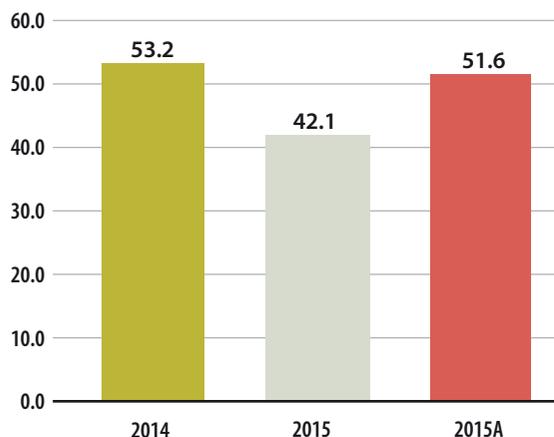
Porcentaje de la población por entidad federativa y tipo de pobreza, 2015A

Entidad federativa	2014						2015A						Diferencias						Concluye
	pobreza	pobreza_e	vul_car	vul_ing	plb	pobreza	pobreza_e	vul_car	vul_ing	plb	Δ pobreza	Δpobreza_e	Δ vul_car	Δ vul_ing	Δ plb				
Nuevo León	20.4%	1.3%	28.0%	9.1%	29.5%	19.9%	1.4%	27.9%	10.5%	30.5%	-2.3%	8.8%	-0.4%	15.6%	3.3%				
Oaxaca	66.8%	28.3%	23.3%	2.1%	68.8%	67.4%	24.7%	20.9%	2.5%	69.8%	0.9%	-12.8%	-10.1%	17.0%	1.4%				
Puebla	64.5%	16.2%	19.8%	5.1%	69.7%	60.8%	11.6%	22.1%	5.2%	66.0%	-5.8%	-28.2%	11.2%	1.5%	-5.2%				
Querétaro	34.2%	3.9%	33.1%	7.8%	42.0%	33.1%	3.7%	31.4%	8.7%	41.8%	-3.1%	-3.1%	-5.2%	11.6%	-0.4%				
Quintana Roo	35.9%	7.0%	34.2%	6.3%	42.1%	37.7%	6.8%	32.2%	6.6%	44.2%	5.0%	-1.9%	-5.8%	5.0%	5.0%				
San Luis Potosí	49.1%	9.5%	24.3%	7.6%	56.7%	46.3%	9.5%	24.7%	7.6%	53.9%	-5.7%	-0.2%	1.6%	-0.8%	-5.1%				
Sinaloa	39.4%	5.3%	30.4%	6.9%	46.3%	36.3%	3.6%	31.5%	7.3%	43.6%	-7.8%	-31.2%	3.8%	6.1%	-5.7%				
Sonora	29.4%	3.3%	32.1%	7.3%	36.6%	28.6%	2.9%	30.7%	7.9%	36.5%	-2.5%	-10.9%	-4.3%	8.4%	-0.4%				
Tabasco	49.6%	11.0%	37.5%	2.3%	51.9%	45.6%	7.7%	41.6%	2.3%	47.9%	-8.0%	-30.6%	11.1%	-2.6%	-7.8%				
Tamaulipas	37.9%	4.3%	24.4%	11.2%	49.1%	39.5%	4.2%	24.3%	10.1%	49.6%	4.2%	-2.8%	-0.3%	-9.9%	1.0%				
Tlaxcala	58.9%	6.5%	22.6%	7.7%	66.6%	56.5%	5.4%	20.7%	10.2%	66.7%	-4.1%	-17.5%	-8.4%	33.5%	0.2%				
Veracruz de Ignacio de la Llave	58.0%	17.2%	24.8%	5.0%	63.0%	58.3%	14.5%	25.7%	4.0%	62.4%	0.6%	-15.7%	3.6%	-20.1%	-1.1%				
Yucatán	45.9%	10.7%	27.7%	7.0%	52.8%	44.8%	8.9%	26.5%	7.6%	52.4%	-2.3%	-16.7%	-4.3%	9.5%	-0.7%				
Zacatecas	52.3%	5.7%	24.4%	7.3%	59.7%	51.4%	5.5%	23.8%	7.3%	58.7%	-1.8%	-3.9%	-2.4%	-0.0%	-1.6%				
Nacional	46.2%	9.5%	26.3%	7.1%	53.2%	44.1%	8.1%	26.5%	7.5%	51.6%	-4.4%	-15.2%	1.1%	5.5%	-3.1%				

Nota: pobreza_e = pobreza extrema, vul_car = población vulnerable por carencias, vul_ing = población vulnerable por ingreso y plb = población con un ingreso menor a la línea de bienestar.

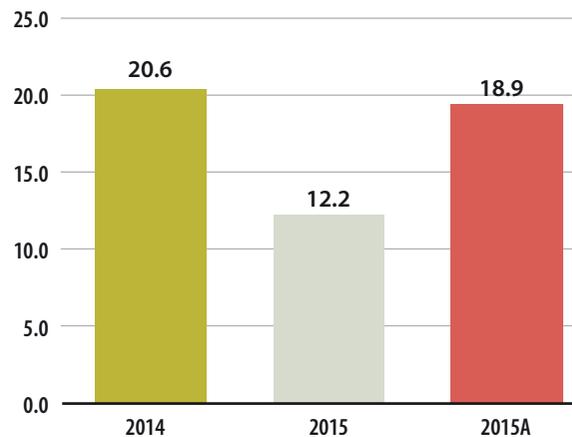
Gráfica 7

Porcentaje de la población con ingreso menor a la línea de bienestar 2014, 2015 y 2015A



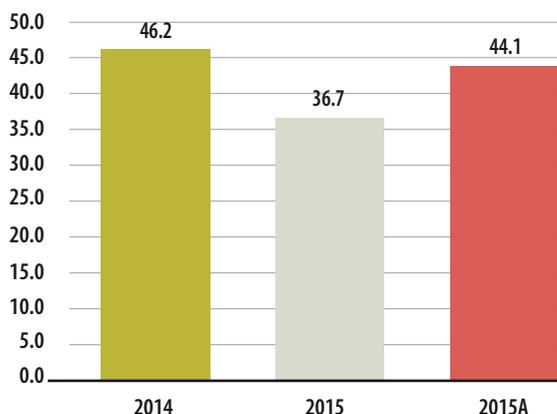
Gráfica 8

Porcentaje de la población con ingreso menor a la línea de bienestar mínimo 2014, 2015 y 2015A



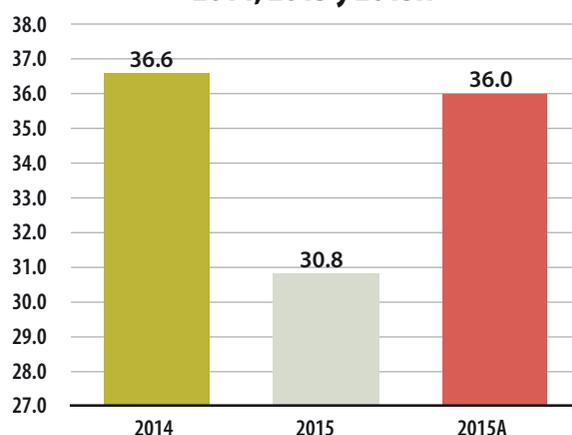
Gráfica 9

Porcentaje de la población en situación de pobreza multidimensional 2014, 2015 y 2015A



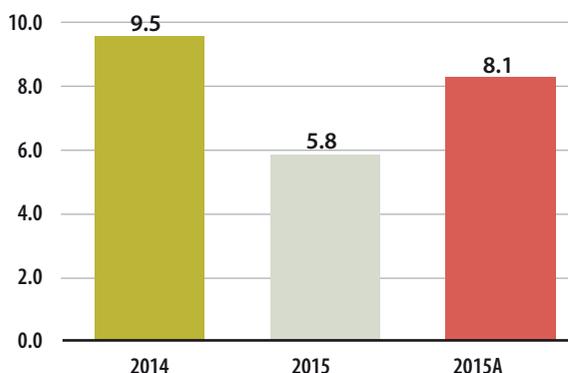
Gráfica 10

Porcentaje de la población en situación de pobreza moderada multidimensional 2014, 2015 y 2015A



Gráfica 11

Porcentaje de la población en situación de pobreza extrema multidimensional 2014, 2015 y 2015A



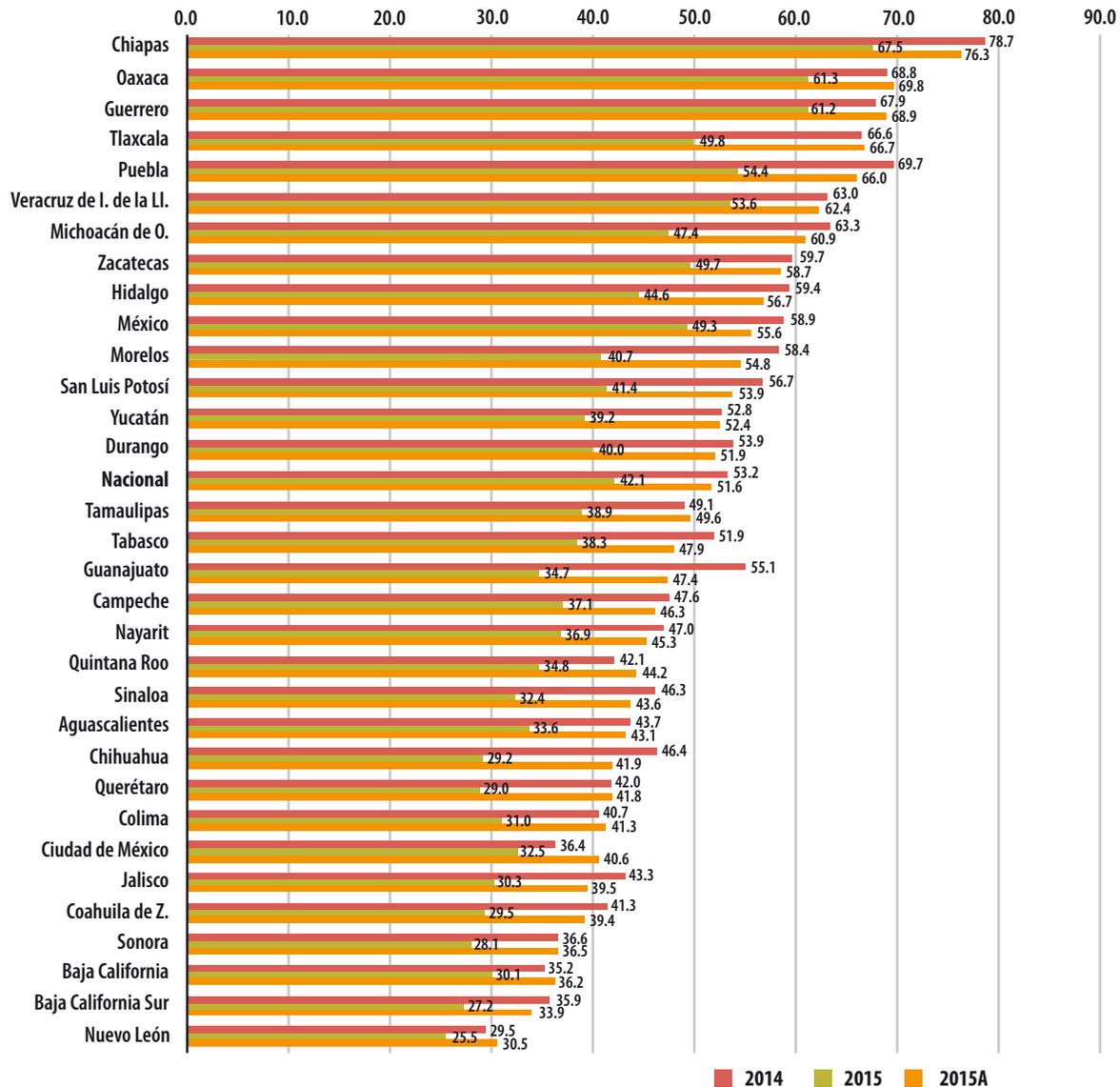
poco respecto al 2014, en el caso de Oaxaca y Guerrero, aumentó de forma ligera.

En el otro extremo, con la menor población con ingreso menor a la línea de bienestar se ubican Nuevo León, Baja California Sur y Baja California; de ellos, Nuevo León y Baja California aumentaron su porcentaje, mientras que Baja California Sur logró bajar la pobreza respecto al 2014 (ver gráfica 12).

Referente al 2015 con los datos ajustados, los tres estados con mayor población con ingreso menor a la línea de bienestar mínimo son Chiapas, Oaxaca y

Gráfica 12

Porcentaje de la población con ingreso menor a la línea de bienestar 2014, 2015 y 2015A (ordenado 2015A)



Guerrero. En comparación con el 2014, vemos que bajaron de manera importante su porcentaje.

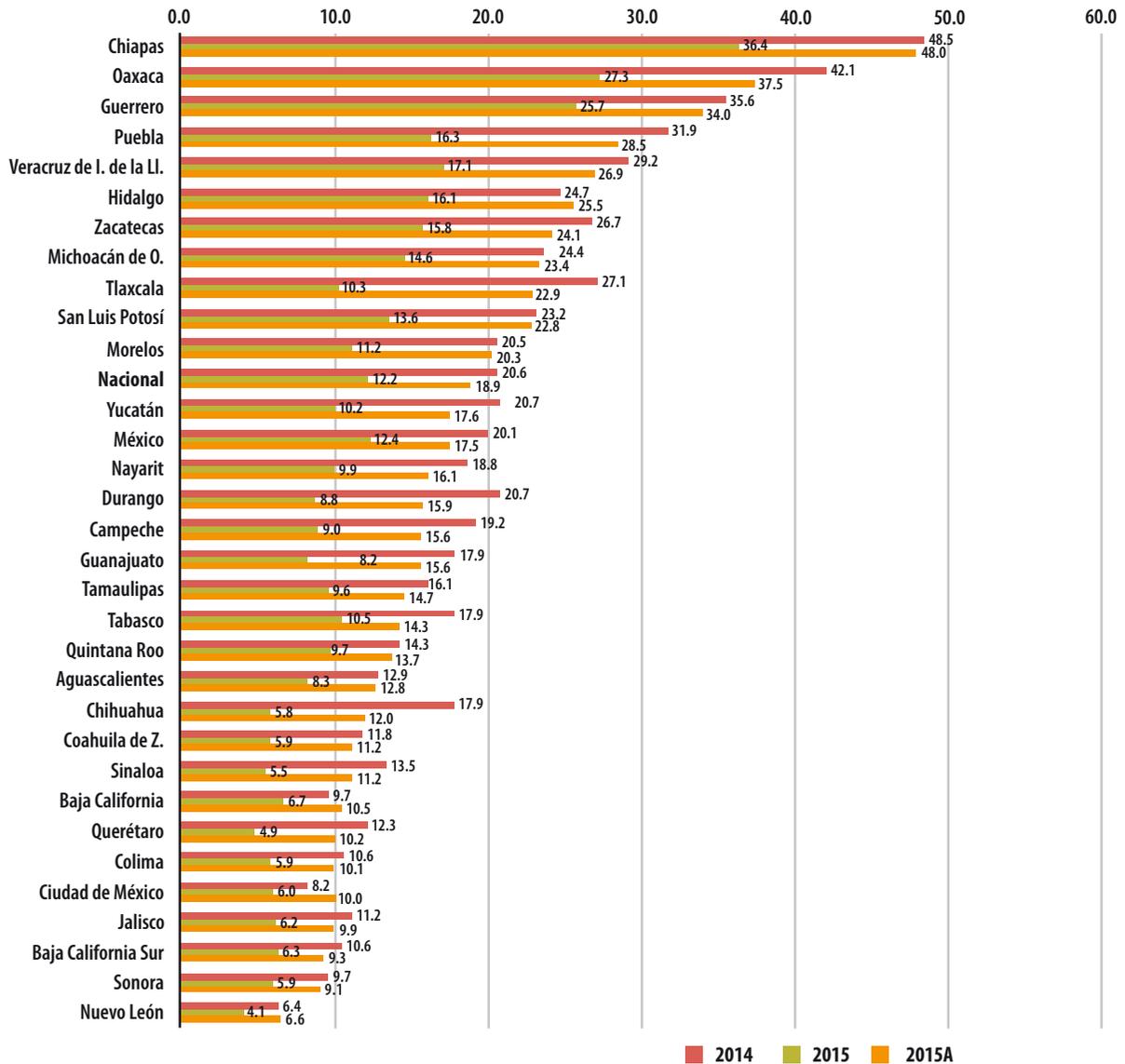
En contraste, con bajos porcentajes de población con ingreso menor a la línea de bienestar mínimo se ubicaron Nuevo León, Sonora y Baja California Sur; los dos últimos lograron disminuir el porcentaje correspondiente, mientras que Nuevo León aumentó ligeramente (ver gráfica 13).

Para el 2015, con datos ajustados, los tres estados que registraron una mayor pobreza multidimensional son Chiapas, Oaxaca y Guerrero; mientras que Chiapas bajó su porcentaje en comparación con el 2014, Oaxaca y Guerrero incrementaron su porcentaje ligeramente.

En contraste, con bajos porcentajes de pobreza multidimensional se ubicaron Nuevo León,

Gráfica 13

Porcentaje de la población con ingreso menor a la línea de bienestar mínimo 2014, 2015 y 2015A, por entidad federativa (ordenado 2015A)



Coahuila de Zaragoza y Baja California Sur. Estas entidades disminuyeron el porcentaje de pobreza multidimensional respecto al 2014 (ver gráfica 14).

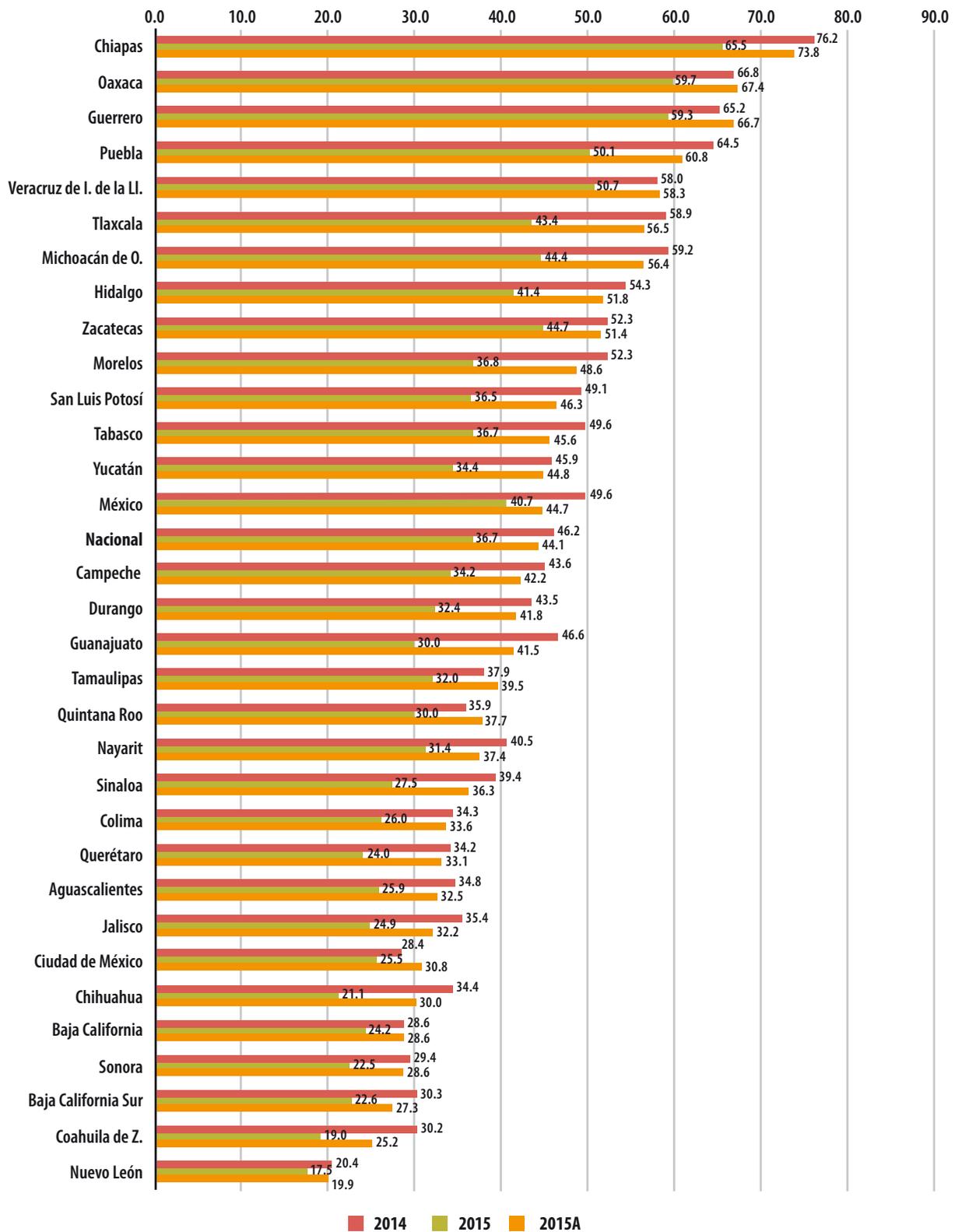
En el rubro de pobreza moderada multidimensional se observa que en el 2015, con los datos ajustados, los tres estados que registraron un mayor porcentaje son Tlaxcala, Puebla y Zacatecas. De ellos, solo Tlaxcala y Zacatecas tuvieron una reducción de su porcentaje respecto al 2014, mientras que en Puebla hubo un aumento.

Por otro lado, con bajos porcentajes de pobreza moderada multidimensional se ubicaron Nuevo León, Coahuila de Zaragoza y Baja California Sur. Estas entidades tuvieron una reducción del porcentaje respecto al 2014 (ver gráfica 15).

Para el 2015, con los datos ajustados, los tres estados que registraron un mayor porcentaje de pobreza extrema multidimensional son Chiapas, Oaxaca y Guerrero. Los dos últimos tuvieron una disminución porcentual importante con respecto al 2014; en contraparte, Chiapas quedó prácticamente igual.

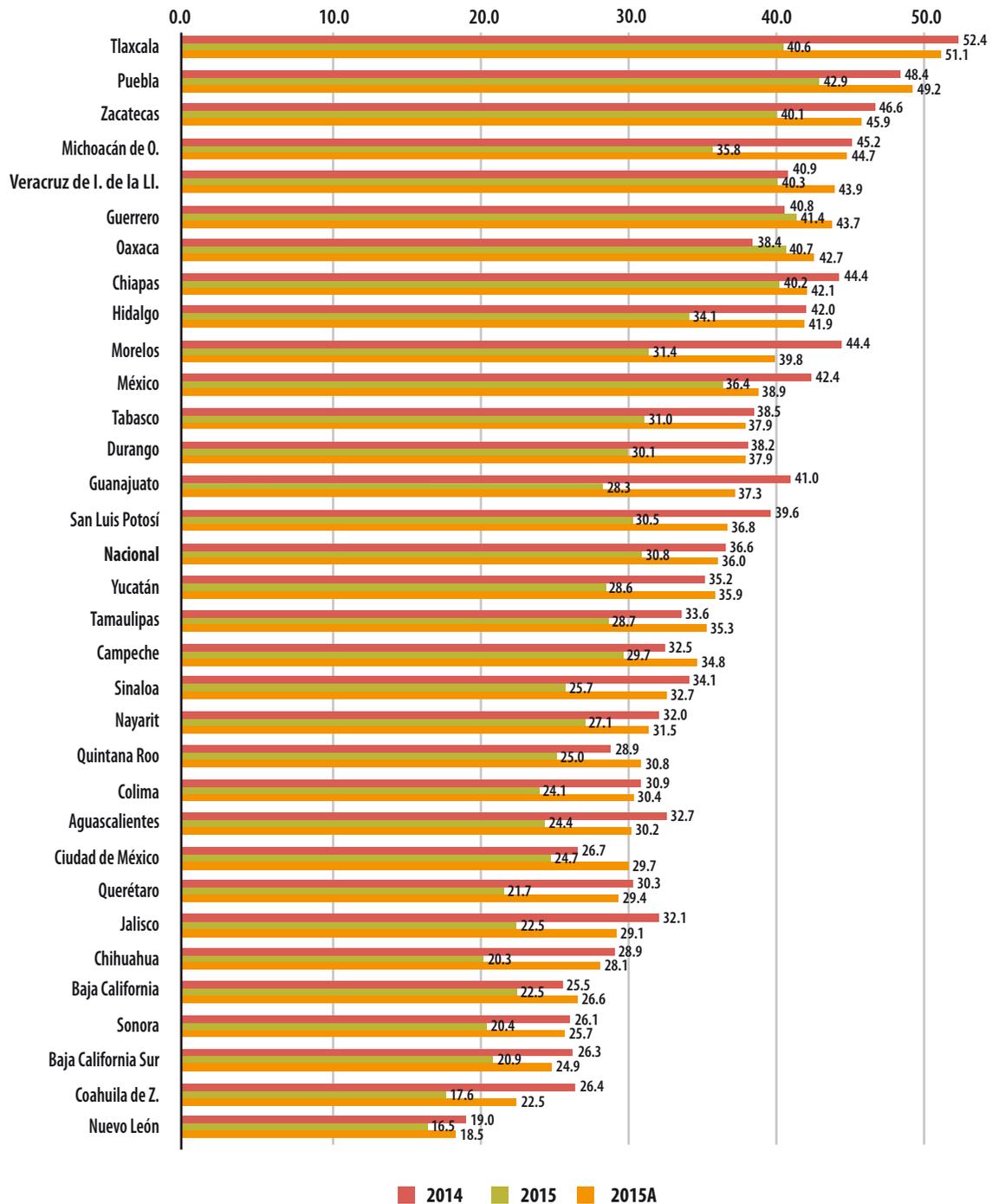
Gráfica 14

Porcentaje de la población en situación de pobreza multidimensional 2014, 2015 y 2015A por entidad federativa (ordenado 2015A)



Gráfica 15

Porcentaje de la población en situación de pobreza moderada multidimensional 2014, 2015 y 2015A por entidad federativa (ordenado 2015A)

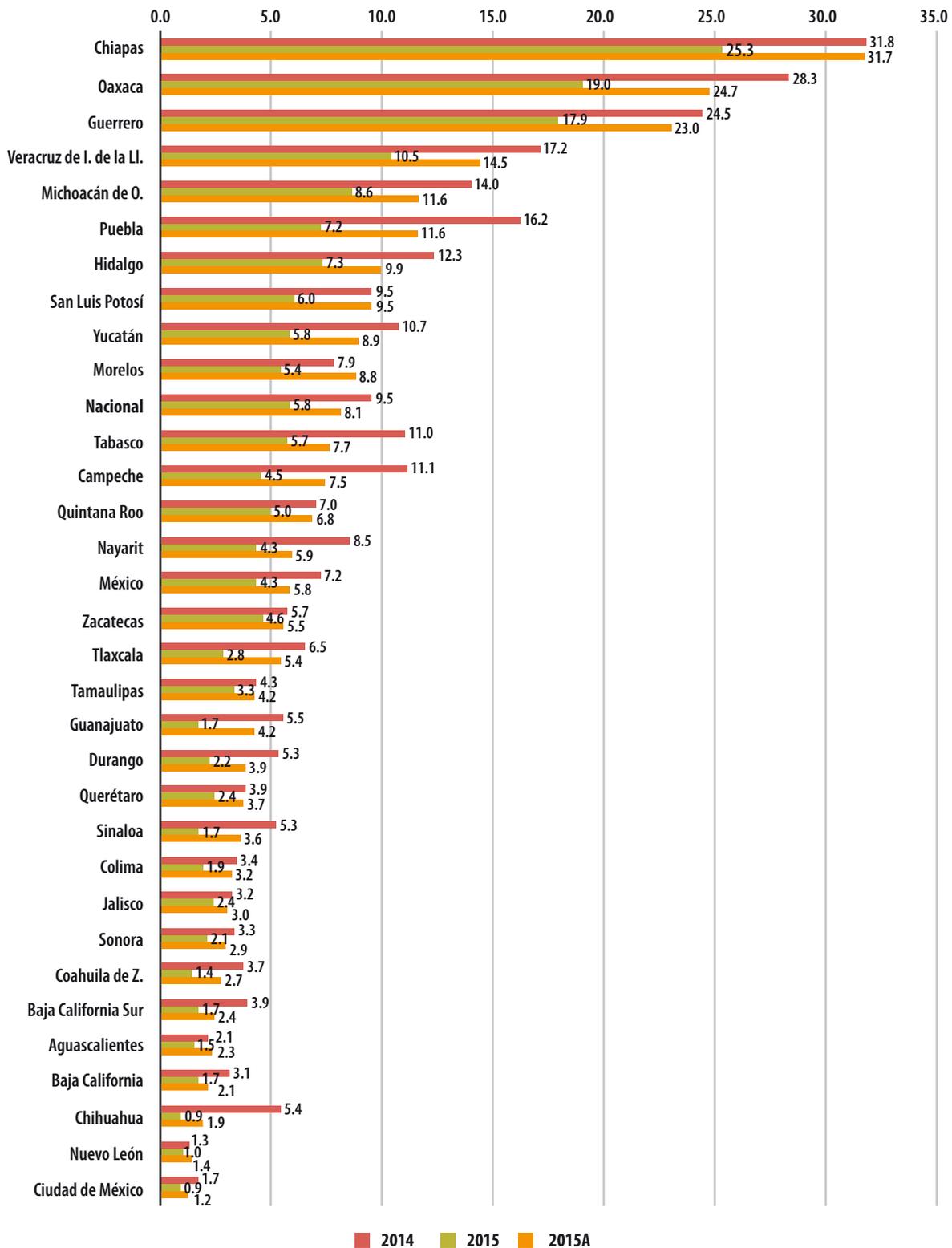


Por otro lado, con bajos porcentajes de pobreza extrema multidimensional se ubicaron Ciudad de México, Nuevo León y Chihuahua. En lo que concierne a Nuevo León, éste quedó prácticamente

igual en comparación con el ejercicio del 2014, con excepción de Ciudad de México, que bajó de forma ligera, y de Chihuahua, que tuvo una disminución importante (ver gráfica 16).

Gráfica 16

Porcentaje de la población en situación de pobreza extrema multidimensional 2014, 2015 y 2015A por entidad federativa (ordenado 2015A)



6. Conclusión y comentarios finales

6.1 Validación y evaluación de la metodología

En la metodología presentada se vierte la idea de aprovechar la tendencia histórica de los gastos y los ingresos, vistos en la ENIGH en su forma tradicional, para crear un indicador que explique los movimientos y la cuantía del crecimiento en relación con el ingreso corriente reportado por la ENIGH-MCS 2014 y el MCS 2015.

Se adoptan los gastos para obtener un factor de crecimiento que se debió observar entre la ENIGH-MCS 2014 y el MCS 2015. Lo que resta es ajustar, mediante proporciones, las cifras del 2015 que se obtienen del producto de la razón de ingresos de ambos ejercicios por la cifra a la que se desea ajustar.

Como parte de la evaluación, esta metodología se aplicará a los resultados de la ENIGH 2016 respecto a los del 2014 y se analizará la comparación de los ejercicios 2014, 2015A y 2016A. A su vez, la contraparte de la metodología que es aplicada a la ENIGH-MCS 2014 permite tener la comparación de los ejercicios 2014A, 2015 y 2016.

6.2 Conclusiones

- El modelo conceptual se sustenta en la idea de que es posible construir una regla de asociación de un parámetro de interés relacionado con el ingreso en los tiempos t_1 y t_0 , con $t_1 > t_0$, a través de un factor de crecimiento real $\Delta_{t_1}^e$ y el efecto de inflación, es decir, $I^{t_1} = \Delta_{t_1}^e I^{t_0C}$. De forma adicional, si la igualdad anterior se vulnera ante la presencia de factores exógenos, entonces existirá un nuevo factor $\Delta_{t_1}^c$ tal que $\hat{I}^{t_1} = \Delta_{t_1}^c \Delta_{t_1}^e \hat{I}^{t_0C}$, de manera que, para obtener un valor ajustado \tilde{I}^{t_1} del parámetro de interés que satisfaga la relación $\tilde{I}^{t_1} = \Delta_{t_1}^e \hat{I}^{t_0C}$ se debería de aplicar

el factor $\frac{1}{\Delta_{t_1}^c}$ a la estimación \hat{I}^{t_1} o, en caso inverso, aplicar el factor $\Delta_{t_1}^c$ a la estimación \hat{I}^{t_0C} .

- Con el objetivo de encontrar un factor de crecimiento real $\Delta_{t_1}^e$ para el ingreso de los hogares, se utiliza la variable ingreso corriente total (*IngCor*) como una variable *proxy* del ingreso teórico (*Y*), la variable de los gastos monetarios como una variable *proxy* del consumo teórico (*C*) y la de erogaciones financieras monetarias como una variable *proxy* del ahorro (*S*). Con éstas se estudió la relación lineal que hay entre los ingresos corrientes y el agregado de los gastos monetarios y erogaciones financieras monetarias de la ENIGH 1992-2014, contruidos de manera tradicional. Se hizo un análisis de cómo estas variables se han comportado con el paso del tiempo, que se desarrolló desde el ejercicio 1992 (nuevos pesos) hasta el del 2014, sin considerarse los correspondientes a 1984, 1989 y 2005 ya que se trata de ejercicios que no concuerdan con la aplicación bienal.
- Se observa que el ingreso y el gasto tienen una alta correlación (0.95) y se crea un modelo para predecir el ingreso 2016, una vez observado el gasto de dicho ejercicio, lo que nos ayuda también a predecir el ingreso 2015, aun cuando el proyecto no haya medido gastos por tratarse solo del MCS. De este modelo se estima que el ingreso 2016 debería ser 5.7% mayor al observado en el 2014 y, por consiguiente, es utilizado en su valor medio para el crecimiento del 2014 al 2015, resultando $\Delta_{t_1}^e$ en 2.81 por ciento.
- Posteriormente, se estima $\frac{1}{\Delta_{t_1}^c}$ y en consecuencia $\Delta_{t_1}^c$, para algunos dominios de interés, a partir de la información que hay en el MCS 2015 de manera independiente para ciertas fuentes de ingresos. Al final, se cuenta con una base de datos ajustada del MCS 2015 donde se elimina, en lo posible, el efecto de factores exógenos del ingreso. A decir, el factor se aplica a las fuentes: *Sueldo por trabajo subordinado*, *Remuneraciones en especie*, *Ingresos por trabajo independiente*, *Otros ingresos por trabajo sea o no subordinado*, *Ingresos provenientes por sociedades*, *Jubilaciones*,

Donativos, Remesas, Transferencias de instituciones y Transferencias de otros hogares.

- Conforme a los resultados del ajuste del MCS 2015 (2015A), el ingreso corriente promedio por hogar registró un incremento de 2.5% respecto al 2014, en comparación al 11.9% que presenta el incremento original 2014-2015. Por deciles de hogares, destaca el decil I con un valor de 2.1%, mientras que el incremento original fue de 33.6 por ciento.
- En cuanto el ingreso corriente por hogar, para el 2015A se registró un incremento de 5.9% respecto al 2014, en comparación con el 15.6% que presenta el incremento original 2014-2015. Por deciles de hogar, se observa que en todos hubo un incremento en favor del 2015A; en particular, destaca el decil I con un valor de 5.5%, mientras que el incremento original fue de 38%; en el 2015A, el coeficiente de Gini con las cifras ajustadas para medir la concentración del ingreso fue de 0.476.
- Al hacer un análisis por fuente de ingreso se observa que en todos los casos se registró un incremento en el periodo 2014-2015A. En particular, destaca el incremento del *Ingreso del trabajo* con un valor de 3.17%; cabe señalar que, con la cifra del 2015 sin el ajuste, tal incremento fue de 12.57 por ciento.
- Por entidad federativa, el ingreso promedio por hogar en el 2015A fue mayor en Nuevo León con 63 734 pesos, seguida de Ciudad de México, Baja California Sur y Aguascalientes con 57 941, 57 193 y 55 234 pesos, respectivamente. Por otro lado, en el 2015A los estados con menores ingresos promedio por hogar al trimestre son Guerrero (24 261 pesos), Oaxaca (25 525), Chiapas (27 707) y Veracruz de Ignacio de la Llave (28 361).
- En cuanto a la pobreza, el porcentaje de la población en situación de pobreza por ingresos es de 51.6 con los datos ajustados (en el 2014 fue de 53.2). En todos los grupos de pobreza se observa una caída notable del 2015 respecto al 2014; una situación similar se muestra con los datos ajustados, pero con menos pronunciamiento; en par-

ticular, la pobreza que experimentó menos diferencia es la moderada multidimensional al pasar de 36.6 a 36% en el periodo de referencia, mientras que la multidimensional pasó de 46.2 a 44.1% en el mismo periodo.

7. Anexos

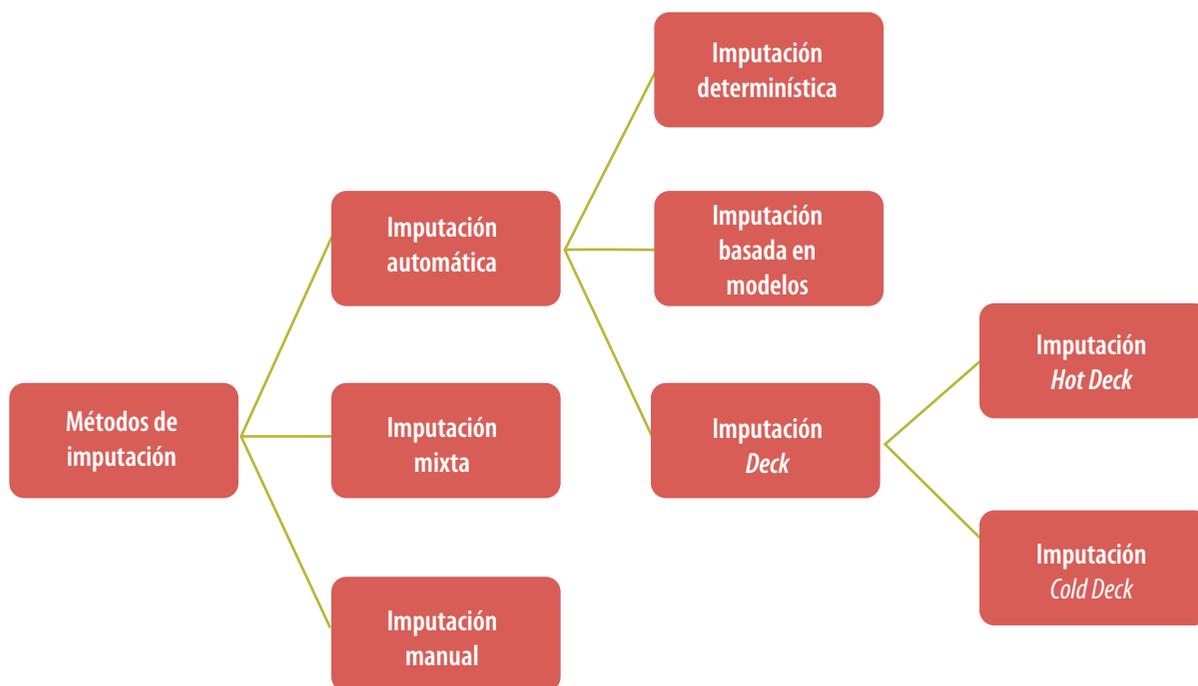
7.1 Clasificación de los métodos de imputación

En los registros de datos de todos los censos y encuestas hay lagunas debidas a información *no declarada, desconocida* u omitida por cualquier otro motivo. También, hay asientos inválidos a causa de equivocaciones de los declarantes, los empadronadores o los encargados de introducir los datos. Los métodos de corrección varían según el elemento de que se trate. En la mayoría de los casos es posible asignar a los elementos códigos válidos con una garantía razonable de que serán correctos, utilizando respuestas correspondientes a otros elementos de la misma persona o del mismo hogar o de los registros de otras personas u hogares (ver gráfico 1).¹³

- A La imputación automática es cuando ésta se realiza de manera automatizada y no se hace necesario revisar cada dato de manera manual, por lo que puede ser de tres maneras:
- Determinística. Cuando solo existe un valor correcto; por ejemplo, cuando falta la suma del total de unos datos, entonces éste es determinado a partir de otros valores en el mismo cuestionario.
 - Basada en modelos. Puede usar una media, mediana, un modelo de regresión, funciones, etc., para imputar el valor.
 - *Deck*. Se requiere de un cuestionario donador para suplir el valor faltante. Puede ser de dos maneras:
 - *Hot-Deck*. El donador proviene de la misma encuesta. La técnica del vecino más cercano es usada para encontrar un

¹³ ONU. *Manual de revisión de datos de los censos de población y vivienda*. Nueva York, ONU, 2011.

Clasificación de los métodos de imputación^a



^a ONU. *Glosario de términos de edición de datos estadísticos*. Ginebra, ONU, 2000.

cuestionario donador, el cual debe estar correlacionado con quien recibe por medio de otras variables.

- *Cold-Deck*. Parecido al *Hot-Deck*, salvo que el donador se encuentra en una encuesta previa.

- B) La imputación manual es, por el contrario, cuando se hace necesario que ésta se realice de manera personal, dato por dato.
- C) Imputación mixta. Utiliza una combinación de los métodos anteriores. Por ejemplo, es posible primero hacer una imputación determinística y si ésta falla, intentar una *Hot-Deck*; si aún sigue fallando, intentar con una imputación basada en modelos; y si aún no es satisfactoria, entonces se hace una imputación manual.

7.2 Pronóstico del ingreso corriente 2016

Dado que los resultados de la ENIGH 2016 no se encuentran publicados, es necesario estimar un gasto (pronóstico) que pueda ser implementado en el modelo (I).

Mediante un modelo lineal, se analizó la variable de gasto contra el tiempo (ver gráfica 17). Asumiendo la observación del 2006 como una posible observación atípica fue eliminada y reformulado el modelo (ver gráfica 18), donde se estimó el gasto 2016 tanto de manera puntual como por intervalo de confianza:

Estimación puntual

$$\text{Gasto 2016} = 1,110,703,910,951.68$$

Estimación por intervalo (90%)

$$\text{Límite inferior gasto 2016} = 972,109,252,564.17$$

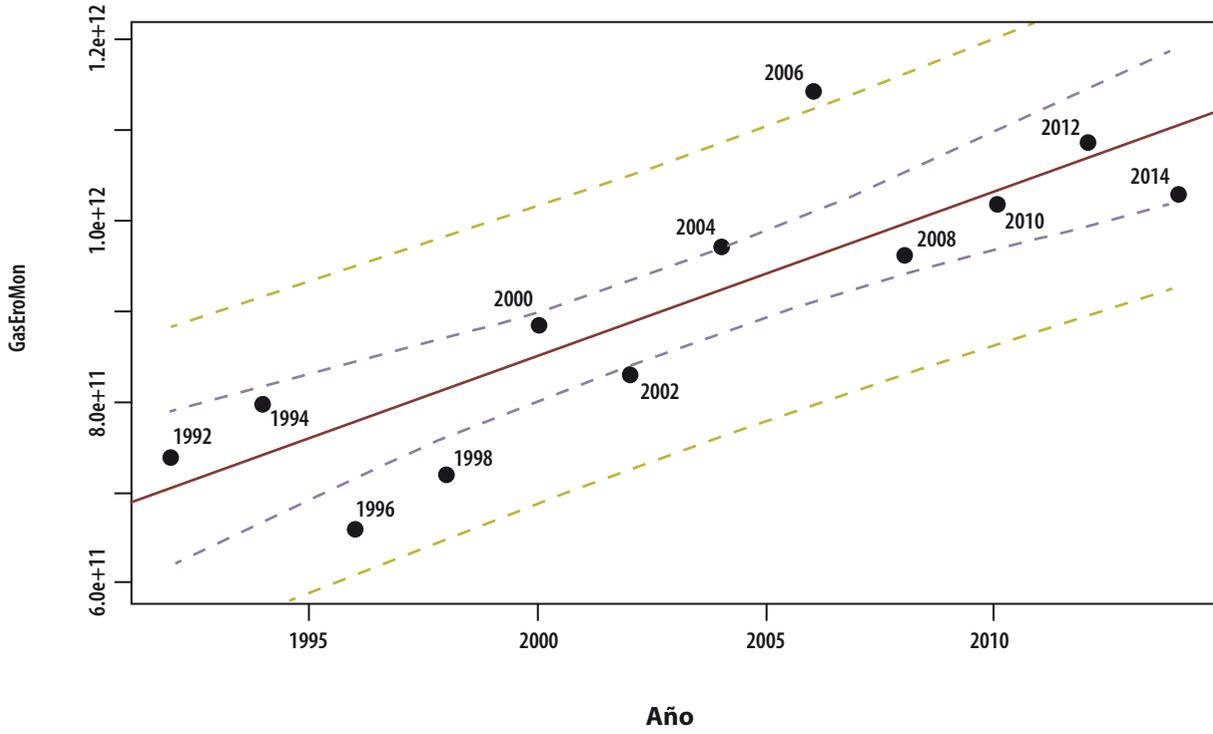
$$\text{Límite superior gasto 2016} = 1,249,298,569,339.20$$

Finalmente, la estimación del gasto 2016 se tomó de manera aleatoria dentro del intervalo de confianza, resultando en:

$$x_{\text{GasEroMon}}^{\widehat{\text{2016}}} = 1,117,011,119,366.3$$

Gráfica 17

Modelo del gasto vs. tiempo con atípico 2006



Gráfica 18

Modelo sin atípico y gasto 2016 estimado antes de la incorporación de un ruido aleatorio

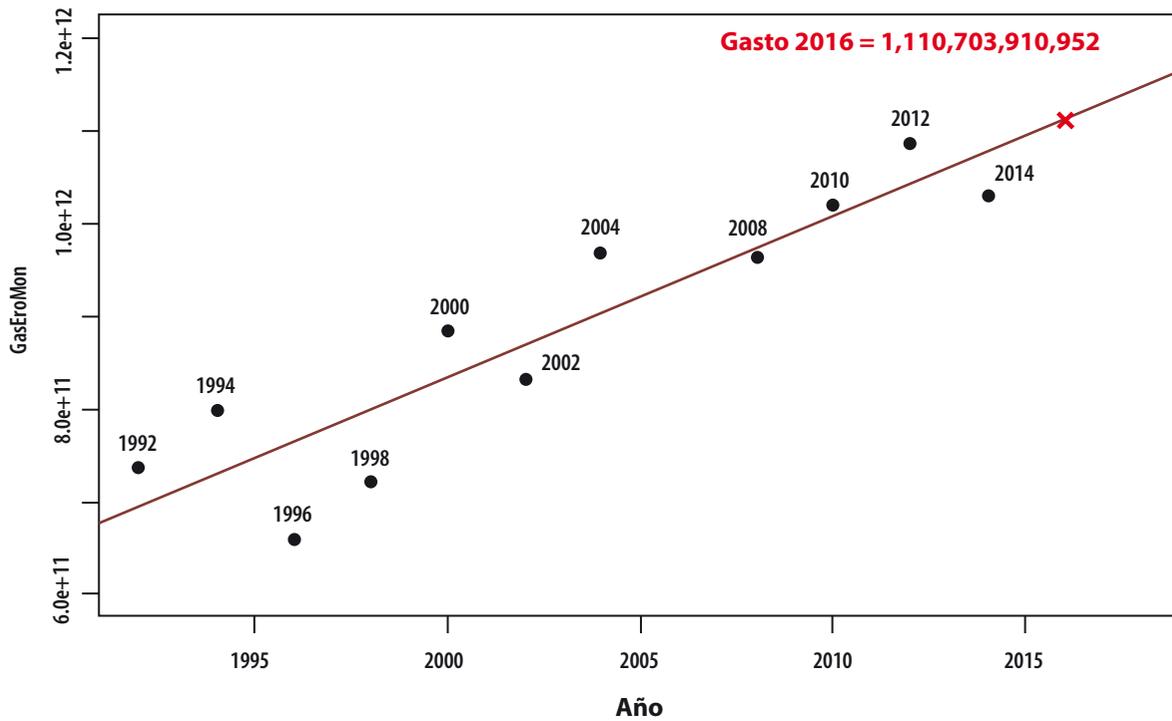


Tabla 10

Diferencias entre los deciles de la ENIGH-MCS 2014 y el MCS 2015

Deciles del ingreso corriente	MCS-ENIGH 2014	Error estándar	MCS 2015	Error estándar	Diferencia 2014-2015
Nacional	41 000	520	45 887	466	11.9%
I	6 116	70	8 169	86	33.6%
II	11 409	94	13 830	104	21.2%
III	15 622	105	18 387	122	17.7%
IV	19 778	121	23 147	146	17.0%
V	24 564	144	28 478	174	15.9%
VI	30 178	179	34 866	210	15.5%
VII	37 433	244	43 030	256	15.0%
VIII	48 105	347	54 139	325	12.5%
IX	66 510	550	73 252	541	10.1%
X	150 283	4 466	161 568	3 921	7.5%*

* A nivel de significancia de 0.05, no existe diferencia entre ejercicios.

Fuente: INEGI. ENIGH-MCS 2014 y MCS 2015.

7.3 Diferencias entre los deciles de la ENIGH-MCS 2014 y el MCS 2015

Éstas se muestran en la tabla 10.

7.4. Nota técnica

7.4.1 Software usado

Para la generación del modelo y la aplicación del algoritmo se utilizó el *software R* y *SPSS*.

8. Fuentes

C. Moore, Jeffrey. *Income Measurement Error in Surveys: A Review*. Suitland, Oficina del Censo de los Estados Unidos, 1997.

CONEVAL. *Metodología para la medición multidimensional de la pobreza en México*. México, DF; CONEVAL, 2014.

_____. *Programa de cálculo de la pobreza*. Ediciones 2010, 2012, 2014 y 2015. Greene, William H. *Econometric Analysis*. Nueva York, NYU-Prentice Hall, 2002.

INEGI. *ENIGH*. Ediciones 1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2005, 2006, 2008, 2010, 2012 y 2014. Documentación y bases de datos.

_____. *ENIGH y MCS*. Ediciones 2010, 2012, 2014. Documentación y bases de datos.

_____. *MCS 2015*. Base de datos.

_____. *Presentación sobre el Módulo de Condiciones Sociodemográficas 2015*. Ciudad de México, INEGI, 2016.

J. Barro, Robert. *Economic Growth*. Cambridge, MIT, 2003.

Obstfeld, Maurice. *Foundations of International Macroeconomics*. Cambridge, MIT, 1996.

Tarling, Roger. *Statistical Modelling for Social Researchers: Principles and Practice*. Nueva York, Routledge, 2009.

Propuesta de ajuste de la distribución *del ingreso corriente total per cápita mediante el gasto corriente total per cápita*

*Adjustment of per capita income distribution by per capita
expenditure: a proposal*

Alida M. Gutiérrez-Landeros, Marco A. Romero-Navarro, Mariana Galindo-Orozco, Carolina Chávez-Ruelas, David Rojas-Rosey, David López-Lira Bayod y Ricardo Aparicio-Jiménez*

El artículo presenta una propuesta de ajuste al ingreso mediante las variaciones del gasto en los hogares. Se muestran las ventajas que puede tener este método de ajuste respecto a otras metodologías y se indican cuáles son las implicaciones potenciales de este ejercicio en el análisis del espacio del bienestar económico y, en consecuencia, para la medición de la pobreza en México.

Palabras clave: ajuste al ingreso; gasto en los hogares; bienestar económico; pobreza.

This article presents a proposal of income adjustment through changes in household spending. The advantages of this method of adjustment with respect to other methodologies are pointed out, as well as some implications of this exercise on the analysis of the economic wellbeing space and, consequently, in the measurement of poverty in Mexico.

Key words: Income adjustment; Household spending; Economic wellbeing; Poverty.

* Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), amgutierrez@coneval.org.mx, mromero@coneval.org.mx, mrgalindo@coneval.org.mx, cchavez@coneval.org.mx, drojas@coneval.org.mx, dlopez@coneval.org.mx y rcaparicio@coneval.org.mx, respectivamente.

Nota: los autores agradecen las observaciones puntuales de las(os) consejeras(os) académicas(os) del Consejo Nacional de Evaluación de la Política de Desarrollo Social.



A businesswoman holding scales of justice/sorbetto/Getty Images

Introducción

De acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI), la información de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2016 incorporó acciones de mejora operativa en la capacitación, supervisión y control del levantamiento de los ingresos de las personas. Estas medidas estuvieron orientadas a mejorar la captación de las fuentes y montos de ingreso que perciben los informantes adecuados de los cuestionarios que integran la ENIGH. Su implementación provocó la ruptura de la continuidad temporal en serie de la información del ingreso de la ENIGH 2016.¹

1 INEGI. *Comunicado de prensa núm. 392/17*. "Presenta INEGI los resultados de una nueva serie de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2016...". Ciudad de México, 28 de agosto de 2017. Consultado el 10 de octubre de 2017.

La afectación a la medición histórica de la información del ingreso se tradujo en un cambio en uno de los indicadores que la *Ley General de Desarrollo Social* mandata utilizar en la definición, identificación y medición de la pobreza en México.² La exigencia de que las comparaciones de las estimaciones de pobreza en el tiempo sean de utilidad para los diversos usuarios de esta información condujo a la necesidad de generar ejercicios estadísticos que recuperen la consistencia de la información del ingreso con mediciones anteriores a través de modelos estadísticos.

2 El capítulo VI sobre la definición y medición de la pobreza, en el artículo 36 de la *Ley General de Desarrollo Social* del 2004, señala que "Los lineamientos y criterios que establezca el Consejo Nacional de Evaluación de la Política de Desarrollo Social para la definición, identificación y medición de la pobreza son de aplicación obligatoria para las entidades y dependencias públicas que participen en la ejecución de los programas de desarrollo social, y deberá utilizar la información que genere el Instituto Nacional de Estadística, Geografía e Informática [sic], independientemente de otros datos que se estime conveniente, al menos sobre los siguientes indicadores: I. Ingreso corriente total per cápita; II. Rezago educativo promedio en el hogar; III. Acceso a los servicios de salud; IV. Acceso a la seguridad social; V. Calidad y espacios de la vivienda; VI. Acceso a los servicios básicos en la vivienda; VII. Acceso a la alimentación y VIII. Grado de cohesión social."

El objetivo de este trabajo es describir una metodología que ajusta el ingreso corriente total per cápita (ICTPC) utilizado para la medición de la pobreza. El ajuste propuesto se realiza mediante las variaciones observadas en el gasto corriente total per cápita (GCTPC) con el propósito de obtener estimaciones de un nivel de ingreso que sea consistente con ejercicios estadísticos previos y que permita valorar la continuidad de la medición multidimensional de la pobreza. Este ajuste se hace empleando una variable proveniente de la propia ENIGH 2016 (el gasto) el cual, según lo señalado por el INEGI, no fue sometida a las modificaciones en campo que sí fueron aplicadas para la recolección de la información sobre ingreso. El hacer uso de la misma fuente de información (en este caso la ENIGH), contribuye a dar validez interna a las estimaciones que se generan de dicho proyecto estadístico y brinda un elemento de certidumbre respecto a las estimaciones obtenidas por el modelo estadístico para la continuidad (MEC) 2016 del Módulo de Condiciones Socioeconómicas (MCS)-ENIGH sobre los indicadores que sirven como insumo para la medición de la pobreza.

El artículo se encuentra organizado de la siguiente forma: la primera sección aborda de forma breve cómo la variable gasto aproxima el bienestar económico de las personas, cuál es su relación teórica y empírica con el ingreso y cuáles son algunas de las experiencias internacionales en medición de la pobreza a través del gasto; la segunda reseña la literatura sobre ajustes a encuestas en hogares que se han hecho para promover la comparabilidad; la tercera describe la metodología y el razonamiento que sirve de sustento para postular que las variaciones del gasto constituyen un insumo adecuado para realizar el ajuste al ingreso; y la última presenta los principales resultados y líneas de investigación derivadas de este ejercicio.

El gasto, el bienestar económico y su relación con el ingreso

El estudio del bienestar económico es una de las dimensiones cuyo análisis resulta relevante para la

comprensión de un concepto más amplio, como el bienestar de las personas y las sociedades. En la búsqueda de conocer el nivel de bienestar de las personas se han desarrollado tanto marcos teóricos comprensivos (donde se reconocen distintos niveles conceptuales relevantes del bienestar)³ como aproximaciones empíricas de la medición del fenómeno en encuestas donde se identifican diversas definiciones.⁴ En general, el punto de partida para evaluar el estado de bienestar de una persona, de una familia o de una sociedad parte de concepciones tradicionales sobre el acceso a bienes y servicios que sirven de satisfactores de las necesidades básicas. Esta concepción se genera bajo la premisa utilitarista que entre más acceso se tenga a recursos económicos o materiales, se estará en un nivel mayor de bienestar.

Una de las primeras decisiones metodológicas para hacer observable un concepto tan amplio como el bienestar económico requiere definir el indicador relevante a utilizar. Como lo menciona el Grupo de Canberra (ONU, 2011), con el fin de tener una perspectiva integral del bienestar económico de las personas que participan como informantes en encuestas en hogares es conveniente examinar el ingreso, el valor de la riqueza del cual disponen, los cambios en su posición de riqueza neta y su nivel de consumo. Estos indicadores cumplen con su objetivo, ya que son expresados en una misma unidad de medida (monetaria), resumen la información de un hogar al aproximar su bienestar económico y permiten no solo caracterizar el patrón de distribución entre la población, sino cada uno es útil para hacer comparaciones entre subgrupos y a través del tiempo.

Una vez que el ingreso y el gasto de los hogares se han contextualizado dentro del concepto de bienestar económico, es de utilidad, como punto de partida para esta discusión, revisar las posiciones teóricas principales en la ciencia económica acerca de la relación entre el ingreso y el consumo de los hogares.

³ Los cinco niveles conceptuales de Kahneman, Diener y Schwartz (1999).

⁴ La satisfacción de vida, la salud y la creación de índices compuestos que combinan tanto condiciones objetivas como percepciones subjetivas (Khan y Juster, 2002).

La primera teoría corresponde a la hipótesis de ingreso absoluto por parte de Keynes (2006). De acuerdo con ésta, las personas aumentan su consumo conforme se incrementa su ingreso; su magnitud está determinada por la propensión marginal a consumir. Además, el autor menciona que la relación no es proporcional entre los estratos del ingreso de una sociedad.

La segunda posición teórica corresponde a lo expuesto por Brady y Friedman (1947) y por Duesenberry (1948) donde, a diferencia de Keynes, afirman que el nivel y la proporción de consumo de un hogar se basan en la posición relativa que mantiene dicho hogar respecto a la distribución de hogares con características similares.

La tercera posición teórica relevante es la de Modigliani y Brumberg (1956), quienes señalan que los individuos efectúan sus decisiones de consumo basados en un criterio de optimización temporal asociado al ciclo vital.

En la teoría del ciclo vital se identifican periodos de ahorro y desahorro basados en las etapas de vida de la persona. De forma adicional, la hipótesis de ingreso permanente de Friedman (1957) busca explorar la dinámica del consumo en el corto plazo en relación con el ingreso promedio o esperado y no por el actual. Así, la proporción de consumo sobre el ingreso permanente es una función de la tasa de interés, la razón de activos al ingreso permanente y factores como la composición del hogar y la variabilidad del ingreso.

El trabajo de Deaton (1992) añade elementos para la comprensión microeconómica del consumo. Este autor no solo señala la importancia del nivel de análisis y la calidad de los datos utilizados, sino que aporta el uso de cohortes sintéticas para contrastar la teoría del ciclo vital y la hipótesis del ingreso permanente. Asimismo, subraya el papel de las preferencias de los individuos para definir cuál ha de ser su nivel de bienestar futuro a partir de las decisiones de consumo y ahorro individuales.

Plantear la restricción presupuestal que enfrentan los hogares sobre un periodo dado es una forma de caracterizar el ingreso y el consumo de los hogares. En este sentido, en el periodo de la vida, el valor presente del promedio del consumo debe igualar el valor presente del promedio del ingreso (no solo del actual, sino del ingreso esperado a lo largo de toda la vida). Si se generaliza esta noción, se puede obtener una restricción presupuestaria intertemporal que reflejará las decisiones de consumo (c), ahorro y deuda de los hogares en función de los ingresos esperados (y_1, y_2), de la tasa de interés (r_2) y de los activos iniciales (A_1). Así, la restricción presupuestaria toma la siguiente forma:

$$c_1 + \frac{c_2}{1+r_2} = A_1 + y_1 + \frac{y_2}{1+r_2}$$

Sobre el tema de medición de pobreza, Albert (2004) señala que existen ventajas y desventajas entre el consumo y el ingreso como indicador de bienestar económico. Una de las ventajas que reconoce en la utilización del ingreso como indicador de bienestar es el menor costo de recopilación de información de ingreso respecto al consumo. En contraste, entre las desventajas, señala que el ingreso tiende a ser más subreportado en comparación con el gasto, la sensibilidad del ingreso a fluctuaciones de corto plazo, el hecho de que ciertas fuentes de ingreso son difíciles de observar (como el ingreso para autoconsumo o el ingreso producto de negocios propios) y que el periodo de referencia no capte de forma fidedigna el ingreso promedio que perciben los hogares. Por otra parte, el consumo, por definición, suaviza las fluctuaciones a corto plazo y es más informativo sobre el bienestar económico a largo plazo de los hogares. En las desventajas del consumo se indica que las elecciones y preferencias de los integrantes de los hogares pueden no ser informativas necesariamente de su bienestar, además de la dificultad de medir componentes específicos del gasto (como el gasto en bienes durables).

Más aún, Meyer y Sullivan (2017) señalan que hay evidencia de que el consumo es una mejor herramienta que el ingreso para estimar el bienestar

económico de los hogares con menos recursos. Sin embargo, reconocen que se ha argumentado que el ingreso muestra ventajas conceptuales sobre el gasto, como sería el hecho de que un individuo puede declarar un bajo nivel de consumo, en tanto que el ingreso representa recursos que pueden ser utilizados para consumir sin incluir preferencias y decisiones de consumo de los individuos. Deaton (1992) señala, de forma adicional, que tanto el gasto como el ingreso cuentan con un error de medición. El análisis de la correlación entre los dos indicadores incluirá los errores de medición de ambas variables y, por lo tanto, podrían indicarse relaciones no precisas.

Así como el ingreso del cual disponen los informantes aproxima los recursos potenciales disponibles en un hogar, el gasto efectuado por las personas de dicho hogar aproxima su estándar de vida a través de las decisiones y preferencias que su consumo refleja. En cualquiera de las dos aproximaciones, tanto el ingreso como el gasto buscan identificar a aquellas personas que no cuenten con la suficiente disponibilidad de recursos económicos y que dicha restricción signifique una limitante no solo respecto a la adquisición de un estándar que le permita satisfacer a los integrantes de un hogar sus necesidades básicas alimentarias y no alimentarias, sino se traduzca en una condicionante de las libertades sociales de las personas y en las opciones vitales que puedan elegir de acuerdo con factores políticos, estructuras sociales y patrones culturales.

Medición de la pobreza a través del gasto

Un requisito para la evaluación del bienestar económico de los hogares es contar con un indicador del monto de recursos que permita analizar las adquisiciones de bienes y servicios, tal es el caso del ingreso. Sin embargo, cuando se analizan las definiciones y procedimientos empleados para capturar el ingreso, se presentan algunas limitaciones al momento de evaluar las características e incidencia de la pobreza. Esto lleva a preferir, en ocasiones, al gasto como indicador que los hoga-

res tienen para analizar la adquisición de bienes y servicios (Beccaria y Perelman, 1999).

Usar el consumo para poder realizar mediciones respecto a las condiciones de vida de los habitantes de los hogares implica realizar una elección de los modelos a ser empleados. Por una parte, se tiene la métrica de la utilidad monetaria de Samuelson (1974); por otro, existe la opción de emplear las razones de bienestar propuestas por Blackorby y Donaldson (1997) y revisadas por Ravallion (1998). Este último enfoque es el utilizado por instituciones como el Banco Mundial.⁵

Análisis de pobreza en países como Ghana, Vietnam, Nepal, Kirguistán, Ecuador, Sudáfrica, Panamá y Brasil han sido conducidos mediante el uso de medidas de consumo y la métrica de la utilidad monetaria (Deaton & Zaidi, 2002). Todos son países en vías de desarrollo, con altas tasas de autoempleo e informalidad y donde realizar una encuesta que capture el ingreso resultaría más costoso que aquellos instrumentos que revisan el consumo. Éstas son condiciones que favorecen el uso del gasto para dar cuenta del bienestar económico de la población de una forma más fidedigna.

No obstante, los esfuerzos de enfocarse en el consumo para realizar una medición de pobreza presentan algunas dificultades. Hay evidencia de que, si bien existe una relación entre el ingreso y el consumo, la misma no es constante a lo largo de la vida de las personas (Thurow, 1969). Los resultados también pueden ser modificados al incluir los componentes de incertidumbre y riesgo (Nagatani, 1972). Los efectos que el ingreso y el consumo puedan tener uno sobre otro muestran un retraso temporal y no hay consenso respecto al número de trimestres en que se puede esperar que los impactos sean reflejados (Campbell y Mankiw, 1990). Incluso, se presenta la exigencia de efectuar un análisis que haga precisiones sobre los tipos de bienes según sean de necesidad

⁵ Esta aproximación se encuentra en trabajos como el de Ravallion y Loshkin (2003) donde se discute la consistencia en términos de utilidad de las líneas de pobreza o el de Hentschel y Lanjouw (2000) que hace ajustes al gasto que reflejen el consumo de servicios básicos provistos por el sector público.

o de lujo, que divida a los grupos de ingreso por deciles para tomar en cuenta sus distintas elasticidades de gasto y que tome en cuenta la durabilidad de los bienes adquiridos (Aguiar y Blis, 2011).

Homogenización de mediciones

La literatura sobre la homogenización de encuestas no es amplia. En general, los esfuerzos de comparabilidad han sido propuestos por organismos internacionales y han estado encaminados a tener mediciones sujetas de ser comparables entre los países. En particular, Olson y Lanjouw (2001) demuestran que, bajo ciertas circunstancias, se pueden hacer comparaciones robustas entre medidas de incidencia de pobreza que utilizan definiciones alternativas de consumo. Este resultado es útil para confirmar la posibilidad de monitorear la pobreza usando encuestas menos extensas sobre los componentes del gasto y, por lo tanto, menos costosas. Son tres las condiciones bajo las cuales esto se cumple:

1. La relación entre dos definiciones de consumo sigue la ley de Engel.
2. Los patrones de consumo son los mismos entre grupos a comparar.
3. No hay error de medición o el grado de ruido entre datos es el mismo en los grupos a comparar.

Cabe destacar que la metodología seguida para su análisis es la comparación de niveles de incidencia de pobreza y el análisis de consistencia de los cambios, misma que se utiliza en la siguiente sección para determinar la relevancia del gasto como elemento para ajustar el ingreso.

Metodología de ajuste del ICTPC mediante el GCTPC

Para lograr el objetivo de corregir la evolución del ingreso mediante el gasto, es necesario recurrir a una definición del gasto comparable con el

ingreso utilizado en la medición de la pobreza.⁶ El ingreso relevante para la medición de la pobreza es el ingreso corriente total (ICT), que contabiliza los recursos disponibles del hogar. El gasto corriente total (GCT) da cuenta de los gastos realizados regularmente en el hogar para la adquisición de su canasta de consumo. Bajo este concepto no se considera el gasto que modifica el patrimonio del hogar o el de cualquiera de los integrantes de éste.⁷

Se determinó contabilizar el GCT como la suma del gasto monetario y no monetario reportado en la encuesta de hogares en los rubros (categorías de bienes y servicios) análogos a los que se incluyen en el ICT.⁸ Este punto es relevante ya que, al contrastarse las incidencias de población con ingreso inferior al valor de las líneas de bienestar, se busca que el concepto de gasto sea consistente con los rubros empleados en la construcción de las canastas alimentarias y no alimentarias, lo cual implica que la incidencia se puede calcular sin hacer ajuste alguno a las líneas de pobreza referidas.

Si se analiza el comportamiento del ICT y GCT se encuentra que, entre el 2008 y el 2016, la variación anual máxima del gasto corriente total real de los hogares fue de 8.5%, mientras que la mínima fue de -4.4 por ciento. Para el caso del ingreso corriente total de los hogares estimado a partir de la ENIGH, la variación fue de 15.9 y -1.9%, respectivamente. Como se observa en la gráfica 1, si se compara la variación de ambos indicadores, la brecha entre las variaciones porcentuales del gasto e ingreso alcanzó un nuevo máximo en el 2016.

La gráfica 2 compara la proporción de gasto e ingreso a partir de la suma de los totales. Se advierte que la proporción del GCT sobre el ICT rondaba en 78.3% entre el 2008 y el 2014 y, durante ese periodo, las proporciones anuales se desviaban de este

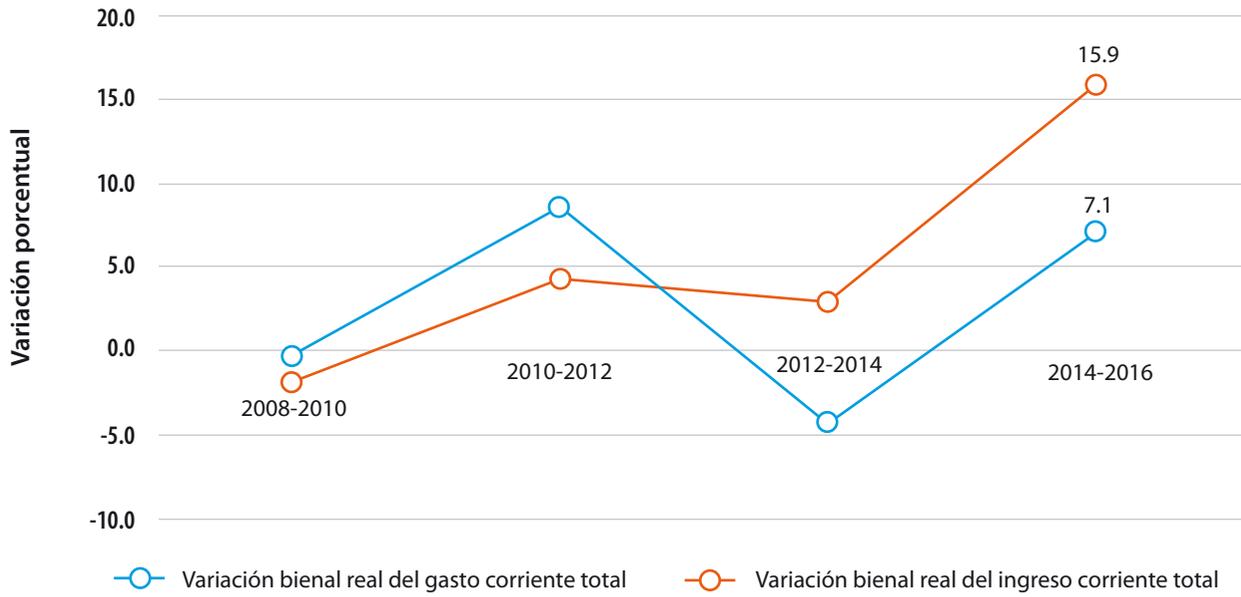
6 Para mayor detalle, consultar CONEVAL. *Metodología para la medición multidimensional de la pobreza en México*. 2014.

7 INEGI. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2014. Nueva construcción de ingresos y gastos. Diseño conceptual y definición de categorías y variables*. 2015.

8 El analizar el gasto incluyó expresarlo a precios reales de la misma forma que se hace con el ICT en la medición de la pobreza, es decir, deflactarlo a precios de agosto del 2016 de acuerdo con la decena de levantamiento.

Gráfica 1

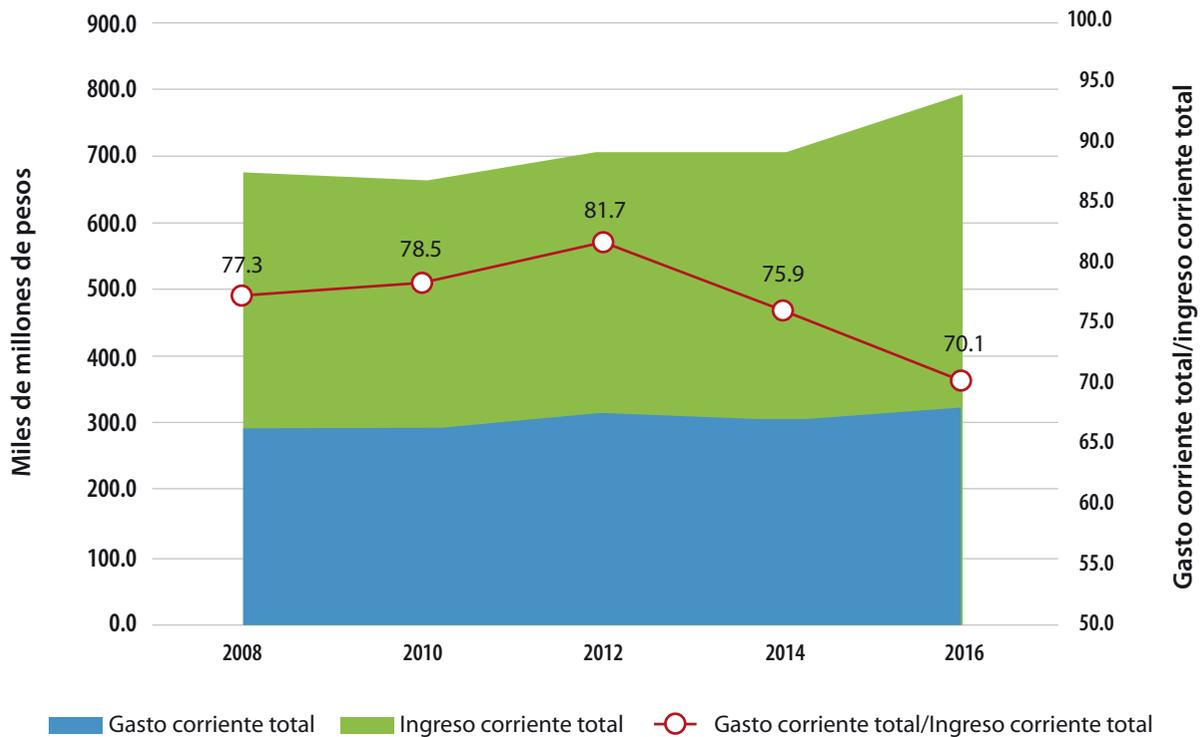
Variación bienal del gasto e ingreso corriente total



Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

Gráfica 2

Suma del GCT e ICT y su proporción, precios de agosto del 2016



Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

promedio no más de cuatro puntos porcentuales. En el 2016, en cambio, dicha proporción fue de 70.1%, desviándose de este promedio en 8.2 puntos porcentuales.

El análisis anterior se cumple de igual forma si se analizan estas dos variables en términos per cápita, es decir, considerando el número de integrantes del hogar y los efectos de economías de escala dentro de los hogares de acuerdo con la metodología del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). La tasa de crecimiento del ICTPC del 2014 al 2016 se encuentra fuera de la tendencia observada con anterioridad, situación que no es similar para el GCTPC. Al observar en el cuadro 1 el comportamiento de la proporción en cada decil de gasto sobre ingreso, la distribución se altera en el 2016 y se muestra una ruptura de la tendencia histórica previa.

A diferencia del ingreso, la tasa de crecimiento del gasto del 2014 al 2016 no mostró cambios que estuviesen atípicamente fuera del intervalo observado en el periodo 2008-2014. Por otro lado, una vez identificada una tendencia en la proporción del gasto entre el ingreso corriente del 2008

al 2014, el 2016 presenta la mayor desviación de dicha tendencia. Es así que podemos asumir que el gasto no muestra señales de un comportamiento atípico, por lo que es posible utilizarlo como variable para ajustar el ingreso y, con ello, validar de manera interna la tendencia que hubiese mostrado el ingreso de no haberse visto afectado su nivel por los cambios operativos en la recolección del ingreso implementados por el INEGI en la ENIGH 2016.

Evaluación del bienestar económico con el gasto

Hasta al momento se ha mostrado la tendencia constante entre el 2008 y el 2014 de la relación entre gasto e ingreso y que no se observa un cambio atípico entre el 2008 y el 2016 en la variación porcentual del gasto. Dado esto, es de interés realizar un ejercicio de medición de la pobreza en el cual la dimensión de bienestar económico se estime exclusivamente con el gasto.

La estimación mantiene la estructura de la construcción del ICTPC de la metodología del CONEVAL. Es decir, se cuenta con un componente monetario

Cuadro 1

Proporción del gasto corriente total per cápita sobre el ingreso corriente total per cápita, total y por deciles, a precios de agosto del 2016

Decil	Proporción				
	2008	2010	2012	2014	2016
1	1.29	1.33	1.31	1.15	0.96
2	1.04	1.03	1.04	0.98	0.85
3	0.96	0.95	0.97	0.91	0.82
4	0.92	0.91	0.92	0.87	0.80
5	0.88	0.87	0.89	0.84	0.79
6	0.85	0.84	0.85	0.81	0.77
7	0.82	0.83	0.84	0.78	0.76
8	0.79	0.81	0.82	0.75	0.74
9	0.76	0.78	0.80	0.72	0.72
10	0.68	0.70	0.76	0.71	0.61
Total	0.78	0.79	0.82	0.76	0.71

Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

y otro no monetario donde el primero será sustituido por el gasto monetario en bienes y servicios y permitirá estimar un GCTPC cuando se divida entre el tamaño de hogar escalado:

$$GCTPC = \frac{\text{gasto monetario} + \text{gasto no monetario}}{\text{tamaño de hogar escalado}}$$

Cabe precisar que el gasto monetario y no monetario consisten en erogaciones en los rubros de alimentos, vestido y calzado, viviendas y servicios de conservación, artículos de limpieza y otros rubros de gasto realizados por los integrantes de los hogares.

Como se puede observar en el cuadro 2 (y a diferencia de la caída de 8.4 puntos porcentuales en el porcentaje de la población que cuenta con un ICTPC por debajo de la línea de bienestar entre el 2014 y el 2016), este mismo indicador, utilizando el GCTPC, presenta una reducción de 5.5 puntos porcentuales. Asimismo, en la gráfica 3, usando el ICTPC, se observa una reducción mayor en el porcentaje de la población por debajo de la línea de bienestar mínimo que al emplear el GCTPC.

La gráfica 4 muestra que la diferencia en el porcentaje de la población con GCTPC por debajo de la

línea de bienestar mínimo disminuye a un ritmo de 4.0, 0.1, 3.5 y -4.4, respectivamente, entre el 2008 y 2010, el 2010 y 2012, el 2012 y 2014, y el 2014 y 2016. El comportamiento es similar si se utiliza el ICTPC, excepto en el periodo entre el 2014 y 2016, donde el cambio en el porcentaje de la población con ingreso inferior al valor de la línea de bienestar mínimo en este periodo es aproximadamente tres veces más, en valor absoluto, que el cambio bienal de mayor magnitud previo al 2014.

Ajuste del ICTPC por medio del GCTPC

La propuesta de ajuste al ICTPC se hace con los datos del GCTPC, el cual, de acuerdo con los niveles y variaciones expuestas, no mostró haber sido afectado por las adecuaciones llevadas por el INEGI en la captación de ingresos. El gasto, al conservar la forma en la cual fue indagado en levantamientos previos, se puede utilizar como herramienta para aproximar el comportamiento del ingreso que habría mostrado de no haber ocurrido los cambios en el proceso de captación de datos. Es decir, la evidencia empírica presentada con anterioridad permite suponer en esta metodología que los datos de gasto de los hogares no sufrieron modificación alguna en su captación, por lo que pueden ser comparados con el gasto de los hogares del levantamiento

Cuadro 2

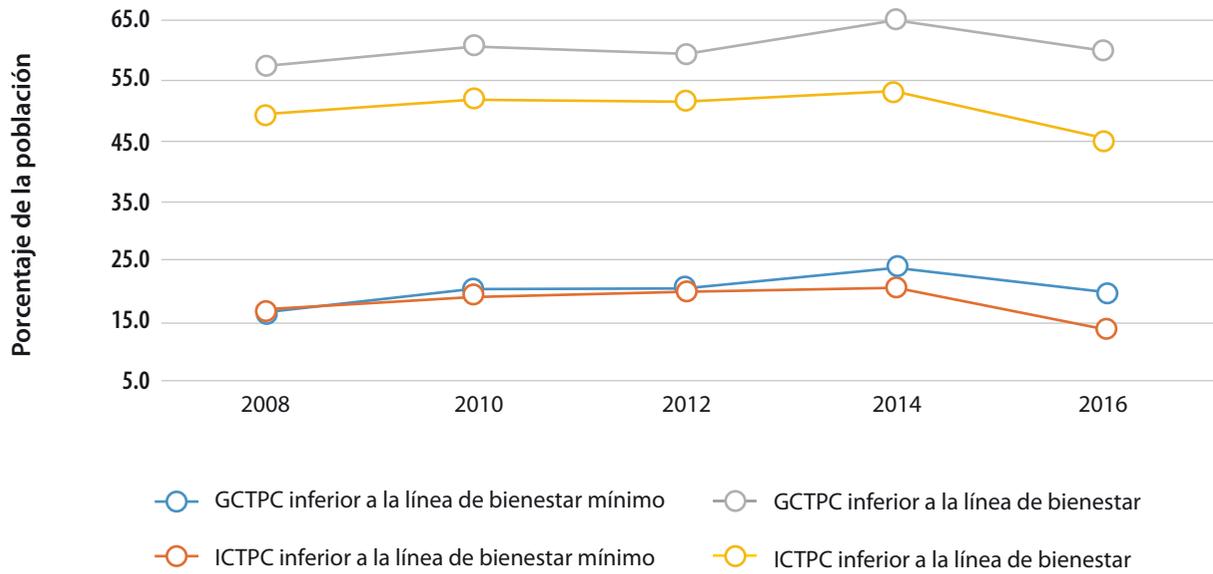
Porcentaje y número de personas por indicador de pobreza, 2010-2016

Indicador	Gasto					Ingreso				
	Porcentaje									
	2008	2010	2012	2014	2016	2008	2010	2012	2014	2016
Pobreza										
Población en situación de pobreza	49.9	51.3	51.3	53.5	48.3	44.3	46.1	45.5	46.2	39.8
Población en situación de pobreza extrema	10.5	11.8	10.4	10.8	7.8	10.6	11.3	9.8	9.5	6.3
Bienestar										
Población con gasto inferior a la línea de bienestar mínimo	16.4	20.5	20.6	24.1	19.7	16.8	19.4	20.0	20.6	13.5
Población con gasto inferior a la línea de bienestar	57.1	60.8	59.7	65.0	59.5	49.0	52.0	51.6	53.2	44.8

Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

Gráfica 3

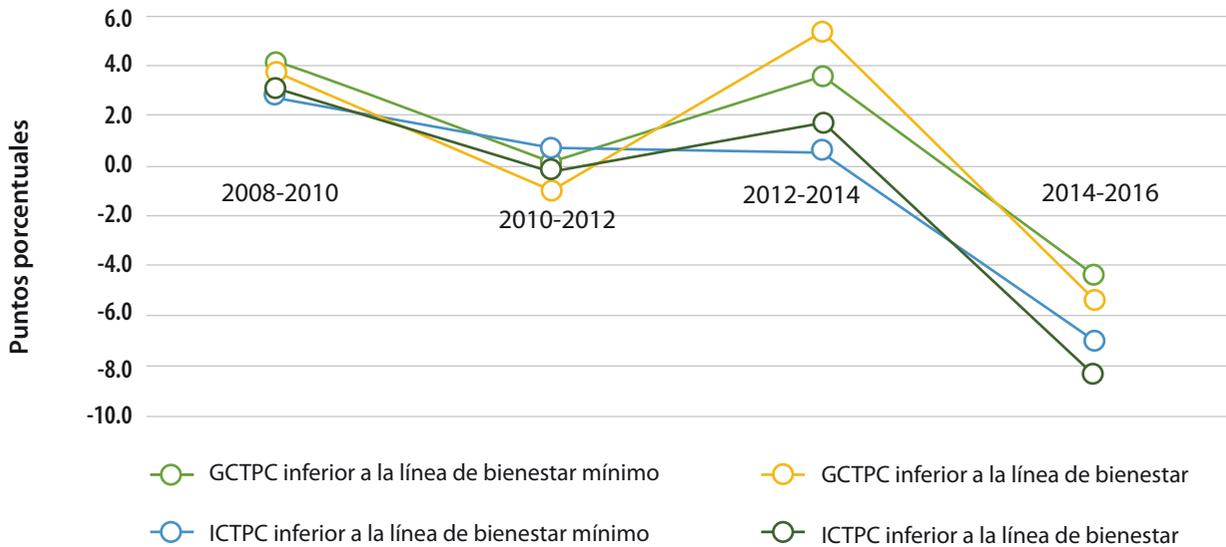
Incidencia de la población por debajo de la línea de bienestar y línea de bienestar mínimo, por indicador de bienestar económico



Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

Gráfica 4

Diferencias en la incidencia de la población por debajo de la línea de bienestar y línea de bienestar mínimo, por indicador de bienestar económico



Fuente: elaboración propia con base en las ENIGH 2008-2016 y MCS-ENIGH 2008-2014.

tamiento anterior (2014). Al ser consistentes los datos del gasto entre años, la tasa de crecimiento de éste se puede utilizar como una herramienta para aproximar el crecimiento que hubiera mostrado el ingreso.

La fuente de datos utilizada es las ENIGH años 2014 y 2016; se determinó utilizar la del 2014 en lugar del MCS de la ENIGH 2014 debido a que el Módulo no contiene toda la información de gasto necesaria para generar el gasto de los hogares.

El ajuste propuesto se realiza distinguiendo entre los ámbitos rural y urbano (las localidades urbanas son aquellas de 2 500 habitantes o más). Se escogió utilizar el corte rural y el urbano debido a que tanto la ENIGH 2014 como la edición 2016 son representativas a esos niveles de concentración poblacional. Aunque la del 2016 es representativa a nivel estatal, el tamaño de la muestra de la del 2014 no permite estimar indicadores de gasto con precisión estadística a nivel estatal.

El primer paso en esta metodología es generar un factor de ajuste, el cual se multiplicará por el ICT del hogar para determinar el nivel que se esperaría tuviera de no haberse presentado los cambios en la captación. Dado que los ingresos originales captados por la ENIGH 2016 muestran un nivel superior a lo esperado, el factor de ajuste debe ser menor a 1.

El factor de ajuste (fa) se construye a partir de las tasas de crecimiento del gasto promedio y del ingreso promedio por hogar para cada uno de los ámbitos (rural y urbano):

$$fa_i = \frac{1 + \Delta G_i}{1 + \Delta I_i}$$

donde fa representa el factor de ajuste; i indica el ámbito rural o urbano; $1 + \Delta G_i$ representa la tasa de crecimiento del gasto per cápita promedio en el ámbito i ; y $1 + \Delta I_i$, la tasa de crecimiento del ingreso per cápita promedio en el ámbito i . En otros términos, se trata de un ajuste proporcional en tanto el resultado de la división es constante para cada uno de los ámbitos de residencia.

Estimación del ICTPC ajustado

El segundo paso en el ejercicio es multiplicar el ingreso per cápita original de cada una de las personas representadas en la base de la ENIGH 2016 por el factor de ajuste del ámbito rural o urbano al que pertenece para obtener el nuevo ICTPC estimado. Es decir, el ingreso corriente total per cápita se ajusta de acuerdo con la razón del crecimiento del gasto y del ingreso promedio en su ámbito de residencia:

$$\hat{I}_{16,i,j} = I_{16,i,j} \cdot I fa_i,$$

donde $\hat{I}_{16,i,j}$ representa el ingreso per cápita estimado del 2016 de la persona j en el ámbito i , $I_{16,i,j}$ indica el ingreso per cápita del 2016 original de la persona j en el ámbito i y fa_i simboliza el factor de ajuste del ámbito i .

Para cada una de las personas que forman la muestra de la ENIGH 2016, se estimó su ICTPC con el factor de ajuste de su ámbito de residencia. El ingreso corriente total per cápita estimado fue empleado para generar las estimaciones de pobreza multidimensional de acuerdo con la metodología diseñada por el CONEVAL.

Interpretación del ajuste

El factor de ajuste se construyó con el objetivo de que, en promedio, el ICTPC estimado del 2016 fuera el resultado de aplicarle al ICTPC del 2014 el crecimiento del gasto per cápita del 2014 y 2016, en cada uno de los ámbitos de residencia. Al hacer la simplificación algebraica necesaria, el promedio del ICTPC estimado (\hat{I}_{16}) se puede expresar de la siguiente manera:

$$\hat{I}_{16,i} = I_{14,i} \cdot [1 + \Delta G_i],$$

donde $\hat{I}_{16,i}$ es el ICTPC estimado promedio en el ámbito i , $I_{14,i}$ es el ICTPC original promedio en el 2014 en el ámbito i y $1 + \Delta G_i$ es el crecimiento del gasto per cápita entre el 2014 y 2016.

En otras palabras, el ajuste tiene como propósito plantear un escenario en el que el ICTPC promedio crece a la misma tasa de cambio del gasto entre el 2014 y 2016. La bondad del ejercicio estriba en la restricción del crecimiento del ingreso a partir de una fuente interna de información que no se afectó por los cambios en los procesos de captación de información del INEGI, tal como sí sucedió con el gasto de la ENIGH.

Principales hallazgos y líneas de investigación

Con el ingreso ajustado por la metodología propuesta se generaron las estimaciones de pobreza⁹ y pobreza extrema, así como de población con un ingreso corriente total per cápita inferior al valor de la línea de bienestar mínimo e inferior a la línea de bienestar. Al utilizar el ajuste al ingreso por medio del crecimiento del gasto, la población en pobreza multidimensional en el 2016 llegó a 53.6 millones de personas, equivalente a 43.8% de la población, lo cual representa una disminución de

⁹ Se realizaron con el ajuste al ingreso por medio del gasto propuesto en el apartado anterior que utiliza la ENIGH 2016 de la nueva serie e incluye el ajuste a seguridad social de la nota técnica 2 publicada en la página del CONEVAL (<http://www.coneval.org.mx/Medicion/MP/Paginas/Notas-pobreza-16.aspx>). Se aclara que el ajuste del ingreso realizado en el MEC 2016 del MCS-ENIGH no fue utilizado en ningún momento en el ejercicio expuesto.

3.1% en el número de personas y 2.5 puntos en el porcentaje de la población.

La caída en la pobreza multidimensional —en el escenario en el cual el ingreso promedio creció a la misma tasa que el gasto promedio— resulta menor a la disminución mostrada con los datos originales de la ENIGH 2016, la cual se situaba en 48.8 millones de personas, que equivalen a 39.8% de la población. El cuadro 3 muestra que el ajuste al ingreso por el gasto evidencia una caída más consistente con el crecimiento económico moderado mostrado en el 2014 y con la tendencia observada en la misma serie histórica.

La elección de la metodología y variable externa de ajuste al ingreso tiene repercusiones sobre el análisis del bienestar económico de los hogares. En particular, se comprobó que la metodología propuesta de ajuste al ingreso por las variaciones del gasto es consistente y abona a la credibilidad de los resultados presentados por el MEC 2016. Además, ésta es una metodología de ajuste parsimoniosa que distingue entre los ámbitos urbano y rural, la cual reconoce las distintas dinámicas de gasto e ingreso entre dichos ámbitos. Como se mencionó en el artículo, la definición de gasto corriente adoptada para el ajuste emplea los mismos rubros utilizados para la construcción de las canastas alimentaria y no alimentaria, por lo que

Cuadro 3

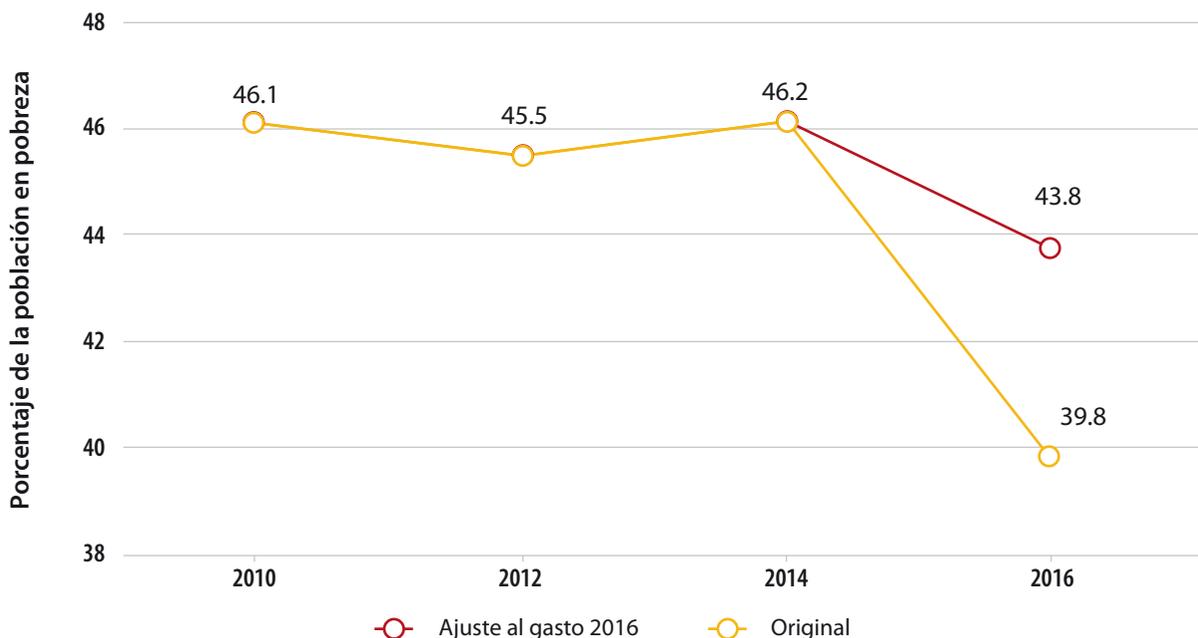
Porcentaje y número de personas por indicador de pobreza, 2010-2014 y 2016 ajustado

Indicadores	Estados Unidos Mexicanos							
	Porcentaje				Millones de personas			
	2010	2012	2014	2016	2010	2012	2014	2016
Pobreza								
Población en situación de pobreza	46.1	45.5	46.2	43.8	52.8	53.3	55.3	53.6
Población en situación de pobreza extrema	11.3	9.8	9.5	7.4	13.0	11.5	11.4	9.0
Población con ingreso inferior a la línea de bienestar mínimo	19.4	20.0	20.6	16.9	22.2	23.5	24.6	20.7
Población con ingreso inferior a la línea de bienestar	52.0	51.6	53.2	51.2	59.6	60.6	63.8	62.7

Fuente: elaboración propia con base en los MCS-ENIGH 2010, 2012 y 2014 y la ENIGH 2016.

Gráfica 5

Porcentaje de la población en pobreza, 2010-2016



Fuente: elaboración propia con base en los MCS-ENIGH 2010, 2012 y 2014 y la ENIGH 2016.

tiene consistencia conceptual el ajuste en tanto se utiliza la misma información que representa el patrón de consumo utilizado como referencia para establecer el umbral de pobreza monetaria.

La validez de la propuesta de ajuste recae sobre la evidencia de la consistencia del gasto con levantamientos anteriores y esto permite aproximar el comportamiento hipotético del ingreso con la relación de variaciones de estas dos variables. Con el fin de tener certeza sobre los procesos de captación de la información, se requiere un diseño experimental que confirme, a nivel hogar, que los cambios en las acciones operativas de recolección en la capacitación, supervisión y control del levantamiento de los ingresos no modificaron las respuestas de gasto de los informantes. Además, la relación entre el ingreso y el gasto debe ser examinada con mayor detalle con datos longitudinales, esto con el fin de corroborar que ambos indicadores son buenos predictores, en términos dinámicos, entre sí mismos.

Fuentes

- Aguar, M. y M. Bils. *Has consumption inequality mirrored income inequality?* NBER Working paper No. 16807. National Bureau of Economic Research, febrero del 2011.
- Albert, J. R. "Measuring poverty" (presentación en PowerPoint), en: curso *Introduction to poverty measurement and analysis*. Blantyre, República de Malawi, Ministerio de Planeación y Desarrollo Económico, 19-20 de enero de 2004.
- Beccaria L. y P. Perelman. *La utilización del gasto y del ingreso en la medición de pobreza*. 1999, pp. 515-527.
- Blackorby, C.; W. Bossert y D. Donaldson. "Critical-Level Utilitarianism and the Population-Ethics Dilemma", en: *Economics and Philosophy*. 13(02), octubre de 1997, pp. 197-230.
- Brady, D. y R. Friedman. *Savings and the Income Distribution*. Nueva York. National Bureau of Economic Research, 1947.
- Campbell, J. y G. Mankiw. "Permanent income, current income, and consumption", en: *Journal of Business & Economic Statistics*. 8 (3), julio de 1990, pp. 265-279.
- CONEVAL. *Metodología para la medición multidimensional de la pobreza en México*. Segunda edición. México, DF, CONEVAL, 2014.
- Diario Oficial de la Federación (20 de enero de 2004). *Ley General de Desarrollo Social*. Última reforma publicada el 1 de junio de 2016.

- Deaton, A. *Understanding consumption*. Oxford University Press, 1992.
- _____. *The analysis of household surveys: a microeconomic approach to development policy*. World Bank Publications, 1997.
- Deaton, A. y S. Zaidi. *Guidelines for constructing consumption aggregates for welfare analysis*. LSMS Working Paper 135. The World Bank, 2002.
- Duesenberry, J. S. "Income-Consumption Relations and Their Implications", en: Lloyd Metzler et al. *Income, Employment and Public Policy*. New York, W. W. Norton & Company, Inc.; 1948.
- Friedman, M. "The Permanent Income Hypothesis" (pp. 20-37), en: *A theory of the consumption function*. Cambridge, Massachusetts, National Bureau of Economic Research, 1957.
- Haughton, J. y S. Khandker. *Handbook on poverty and inequality*. Washington, DC, The World Bank, 2009.
- Hentschel, J. y P. Lanjouw. "Household welfare measurement and the pricing of basic services", en: *Journal of International Development*. Vol. 12, núm. 1, 2000.
- INEGI. *Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2014. Nueva construcción de ingresos y gastos. Diseño conceptual y definición de categorías y variables*. 2015.
- _____. *Comunicado de prensa núm. 392/17*. "Presenta INEGI los resultados de una nueva serie de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) 2016". Ciudad de México, 28 de agosto de 2017. Consultado el 10 de octubre de 2017.
- Kahn, R. y T. Juster. "Well-being: concepts and measures", en: *Journal of Social Issues*. Vol. 58, núm. 4, 2002, pp. 627-644.
- Kahneman, D.; E. Diener y N. Schwarz. *Well-being: The foundations of hedonic psychology*. New York, Russell Sage, 1999.
- Keynes, J. M. *Teoría general de la ocupación, el interés y el dinero*. Madrid, Fondo de Cultura Económica, 2006.
- Lanjow, J. O. y P. Lanjow. "How to compare apples and oranges: poverty measurement base on different definitions of consumption", en: *Review of Income and Wealth*. 47 (1), marzo de 2011, pp. 25-42.
- Meyer, B. D. y J. X. Sullivan. *Consumption and Income Inequality in the US Since the 1960s*. No. w23655. National Bureau of Economic Research, 2017.
- Modigliani, F. y R. Brumberg. "Utility Analysis and the Consumption Function: an Interpretation of Cross-section Data", en: Kurihara, K. (ed.). *Post Keynesian Economics*. New Brunswick, NJ, Rutgers University Press, 1954, pp. 388-436.
- Nagatani, K. "Life cycle saving: theory and fact", en: *American Economic Review*. (62), junio de 1972, 344-353.
- Organización Internacional del Trabajo. *Estadísticas de ingresos y gastos de los hogares. Informe II*. Decimoséptima Conferencia Internacional de Estadísticos del Trabajo, 2003.
- Organización de las Naciones Unidas. *Canberra Group Handbook on Household Income Statistics*. Segunda edición. Ginebra, Suiza, 2011.
- Ravallion, M. *Poverty lines in theory and practice*. LSMS Working Paper 113. The World Bank, 1998.
- Ravallion, M. y M. Lokshin. "On the utility consistency of poverty lines", en: *World Bank Policy Research Paper 3157*. 2003.
- Samuelson, P. "Complementarity – An essay on the 40th anniversary of the Hicks – Allen revolution in demand theory", en: *Journal of Economic Literature*. (15), 1974, pp. 24-55.
- Thurow, L. "The optimum lifetime distribution of consumption expenditures", en: *American Economic Review*. (59), junio de 1969, pp. 324-330.

Modelo estadístico 2016 para la continuidad del MCS-ENIGH

2016 Statistical Model for MCS-ENIGH's Continuity

José Alejandro Ruiz Sánchez* y **Ana Miriam Romo Anaya****

* Instituto Nacional de Estadística y Geografía (INEGI), jose.ruizs@inegi.org.mx.

** INEGI, miriam.romo@inegi.org.mx

Nota: los autores agradecen la colaboración de Lorenzo Cecilio Fernández en el proceso de integración del artículo.

Small figures whit money/Peter Muller/Getty Images.



En un proceso de mejora continua, el Instituto Nacional de Estadística y Geografía (INEGI) identificó áreas de oportunidad en el Módulo de Condiciones Socioeconómicas (MCS) de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), que se materializaron en la implementación de acciones operativas en la ENIGH 2016, con el objetivo de alcanzar un mayor apego al diseño de la encuesta. Con ello, se inicia una nueva serie de la ENIGH.

Para dar continuidad a los resultados bienales de la serie del MCS-ENIGH, se propone una metodología de edición de microdatos basada en el ajuste de una función de distribución bajo restricciones. En una primera etapa, solo se modifican los ingresos por trabajo principal reportados en la ENIGH 2016; el ajuste se realiza en función de las tasas de crecimiento observadas en la Encuesta Nacional de Ocupación y Empleo (ENOE) por entidad federativa. En la fase siguiente, se modifica el resto del ingreso total de los hogares; para ello, nos basamos en una relación constante que se ha observado desde el 2010 entre los ingresos totales y los ingresos por trabajo principal.

Como resultado de la modificación de los ingresos reportados, el crecimiento del ingreso total de los hogares entre el 2014 y los datos que arroja el modelo para el 2016 es de 2.1% (en comparación con 10.4%, si se utilizan datos de la ENIGH 2016). La variación para el primer decil es de 8.5% (mientras que con datos de la ENIGH 2016 fue de 29.9%).

Palabras clave: MCS-ENIGH; ENIGH 2016; imputación.

1. Introducción

En un proceso continuo de mejora, el INEGI identificó áreas de oportunidad alrededor de la capacitación, supervisión y control operativo del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH). A partir de ello, se llevaron a cabo acciones de fortalecimiento operativo en el levantamiento de la ENIGH 2016 en un esfuerzo por alcanzar un mayor apego al diseño de la encuesta.

As a way to keep improving, INEGI identified some opportunity areas in the MCS-ENIGH, which resulted in the implementation of certain operative actions in the ENIGH 2016. This was aimed to get as close as possible to the survey's design. This is the start of a new ENIGH series.

We propose a new micro-data editing methodology in order to keep the continuous biennial results of the MCS-ENIGH series. This methodology is based on the adjustment of restricted distribution functions. On a first stage, only incomes per main labor reported in ENIGH 2016 are modified, with the adjustment done according to the growth rates observed in ENOE by Federal entity. In the next stage, we modified the rest of households' total income based on a constant relation observed since 2010 between total incomes and incomes derived from a person's main labor.

As a result of the modifications of reported incomes, households' total income growth, between 2014 and the data we get from the model designed for 2016, is of 2.1% (unlike the 10.4% we get when compared to data from ENIGH 2016 without adjustments). The variation for the first decile is of 8.5% (it was of 29.9% with data from ENIGH 2016 without adjustments).

Key words: MCS-ENIGH; ENIGH 2016; imputation.

Con estas medidas implementadas, se inicia una nueva serie histórica para las encuestas de ingresos y gastos de los hogares. Sin embargo, y partiendo del reconocimiento de la necesidad de preservar la continuidad de sus ejercicios estadísticos, se presenta esta propuesta metodológica, cuyo objetivo es proveer de los insumos necesarios para la medición de la pobreza multidimensional, de tal manera que se mantenga la consistencia con los resultados de la serie bienal 2008-2014, tanto a nivel nacional como por entidad federativa.

En la medición de la pobreza multidimensional se incorporan variables no monetarias como la condición de acceso a salud, educación, seguridad social, etc., pero también otras que reflejan el ingreso de los hogares. Este documento está centrado solo en la modificación de este último grupo. En México, el encargado de medir la pobreza es el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), cuya metodología toma como insumo la base de datos del MCS-ENIGH. Por lo tanto, una parte fundamental del proceso de ajuste consistirá en el desarrollo de un mecanismo que permita transitar de relaciones agregadas a nuevos microdatos de ingresos que sustituyan a los originalmente reportados.

El ingreso total de los hogares proviene de múltiples fuentes, por ejemplo: salarios, ganancias de empresas, recursos que reciben de otros hogares, etcétera. La ENIGH captura esa heterogeneidad a través de una serie de reactivos que son recabados a nivel individuo. La gran cantidad de preguntas referidas a los distintos canales de ingreso, entre otras características, hacen que la ENIGH tenga elementos que no poseen otros operativos similares levantados en el país.

A pesar de la diversidad de fuentes, la mayor parte del ingreso de los mexicanos proviene de su ingreso laboral, es decir, de su trabajo como asalariado y/o independiente (principal y secundario) que, en conjunto, ha representado cerca de 70% del ingreso corriente total (ICT) desde el 2010. Este elevado porcentaje, junto con la existencia de un levantamiento regular sobre condiciones laborales, como lo es la Encuesta Nacional de Ocupación y Empleo (ENOE), crean una ventana de oportunidad para utilizar el ingreso laboral que reporta esta encuesta como ancla para ajustar los ingresos laborales.¹

Para la modificación de los ingresos laborales reportados por los hogares, aproximamos una función de densidad de probabilidad a los datos empíricos del ingreso de la ENIGH 2016; posteriormente,

se ajusta otra función de densidad pero ahora imponiéndole como condición que sea capaz de generar nuevos microdatos tales que arrojen un valor objetivo de un estadístico específico (por ejemplo, un determinado promedio muestral); en nuestro caso, ese valor proviene de la trayectoria para cada entidad federativa de los ingresos laborales de la ENOE.

Al modificar solo el ingreso laboral, la relación entre éste y el ingreso no laboral se ve alterada, rompiendo una trayectoria histórica que se venía observando desde el 2010. Para recomponer la relación, y en una segunda fase, se modifica el ingreso no laboral, de tal forma que se elimine el efecto colateral causado por el primer ajuste. Así, esta metodología modifica solo las variables de ingreso, dejando inalteradas las variables no-ingreso y los factores de expansión; esto último permite mantener el diseño muestral de la ENIGH.

El documento se encuentra dividido en siete secciones. En la segunda se presentan algunos estadísticos descriptivos que permiten conocer la magnitud y distribución de los efectos de las mejoras. En la tercera se desarrolla el marco teórico general de la metodología para modificar microdatos. En la siguiente se crean nuevos microdatos del ingreso por trabajo principal (ITP). En la quinta se presenta la segunda y última fase del ajuste a los ingresos de los hogares, relacionada con la modificación de los ingresos distintos al ITP (*ITP^c*). En la penúltima se muestran los principales resultados del ajuste y en la última, se concluye.

2. Análisis descriptivo

Como consecuencia de las acciones tomadas, el ICT que reportan los hogares en la ENIGH 2016 tiene incrementos sustanciales con respecto al 2014, en especial para los primeros deciles (ver cuadros 2.1 y 2.2).² Lo anterior se debe, en parte, a que ahora el problema de la subdeclaración que hacen los hogares es menor; ello se ve reflejado

¹ Salvo en el primer trimestre del año, el ingreso laboral de la ENOE hace referencia al correspondiente originado por el trabajo principal.

² A menos que se especifique lo contrario, las cifras reportadas para años anteriores al 2016 son referidas a los datos conjuntos del MCS-ENIGH.

Cuadro 2.1

Ingreso corriente promedio por decil (precios del 2016)

	2010	2012	2014	2016
Nacional	43 931	42 115	42 157	46 521
I	6 111	5 920	6 288	8 166
II	11 707	11 310	11 731	14 206
III	16 285	15 750	16 062	18 918
IV	20 876	20 326	20 336	23 556
V	26 098	25 357	25 258	28 813
VI	32 400	31 502	31 030	34 837
VII	40 352	39 343	38 489	42 431
VIII	52 028	50 494	49 463	53 383
IX	72 580	69 321	68 386	72 041
X	160 872	151 828	154 524	168 856

Cuadro 2.2

Variación porcentual del ICT por decil (promedio, precios del 2016)

	2010-2012	2012-2014	2014-2016
Nacional	-4.1%	0.1%	10.4%
I	-3.1%	6.2%	29.9%
II	-3.4%	3.7%	21.1%
III	-3.3%	2.0%	17.8%
IV	-2.6%	0.1%	15.8%
V	-2.8%	-0.4%	14.1%
VI	-2.8%	-1.5%	12.3%
VII	-2.5%	-2.2%	10.2%
VIII	-2.9%	-2.0%	7.9%
IX	-4.5%	-1.3%	5.3%
X	-5.6%	1.8%	9.3%

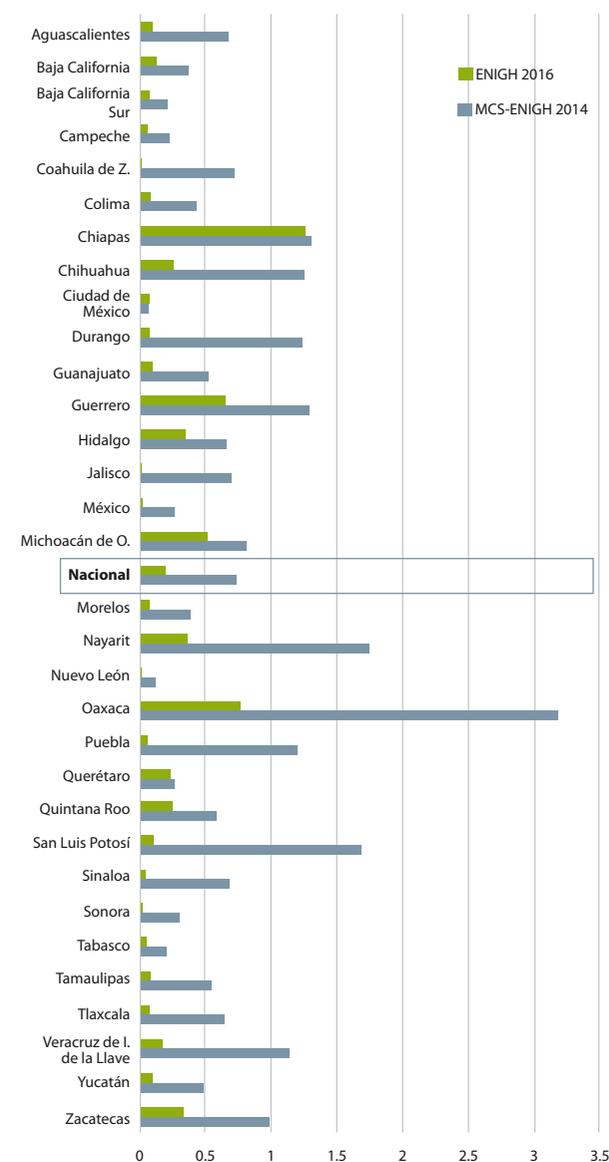
en el porcentaje que indicó recibir menos de 3 mil pesos trimestrales, el cual pasó de 0.7% en el 2014 a 0.2% en el 2016 a nivel nacional (ver figura 2.1).

A nivel entidad federativa, los porcentajes muestran la misma tendencia decreciente (excepto en la Ciudad de México); incluso, para algunas entidades como Coahuila de Zaragoza, Jalisco, México, Nuevo León y Sonora, el porcentaje es cercano a cero.

Los cambios en tendencia mostrados en los cuadros anteriores son el resultado de las mejoras implementadas en el 2016, por lo que la información de la ENIGH 2016 no es consistente con la de años anteriores.

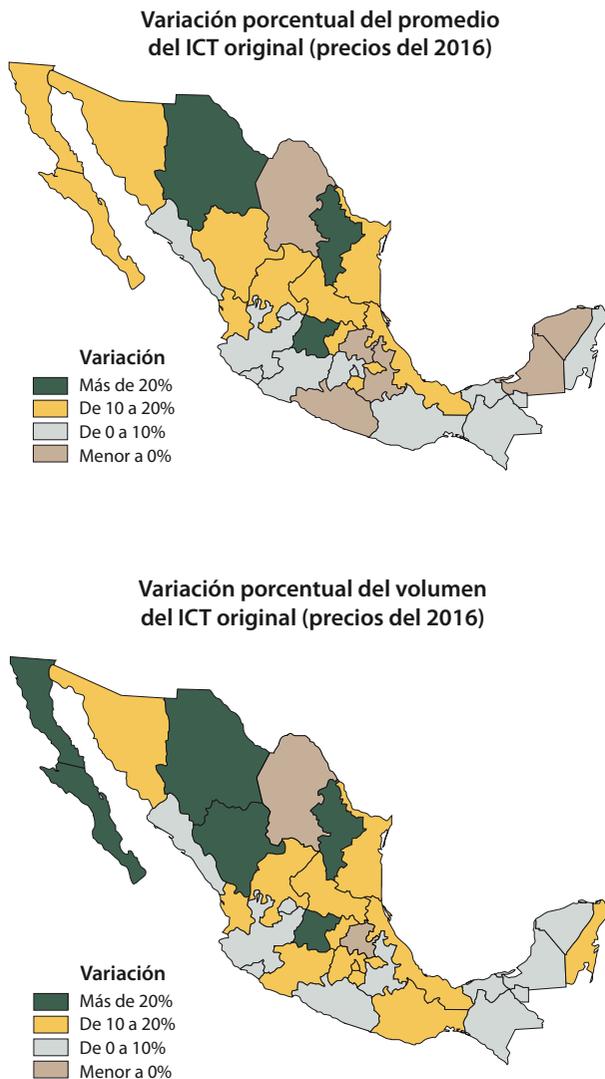
Figura 2.1

Porcentaje de hogares con ingresos inferiores a 3 mil pesos trimestrales (precios del 2016)



Por otra parte, y debido a que las instrucciones para una mejor recolección de la información se dieron a nivel nacional, la declaración de ingresos tuvo un efecto en las distintas entidades federativas, aunque de manera diferenciada. La figura 2.2 permite identificar regiones de acuerdo con su variación porcentual en el ingreso por hogar entre el 2014 y 2016. Observamos que, tanto en volumen como en promedio, la zona norte del país fue la que experimentó las mayores variaciones con respecto al 2014, llegando a niveles superiores a 20% en entidades como Nuevo León y Chihuahua.

Figura 2.2
Variación porcentual del ICT del 2014 al 2016
(precios del 2016)



Otro nivel de desagregación que se puede lograr con la información de la ENIGH es por fuentes de ingreso, es decir, aquellos rubros que componen el ingreso total de los hogares. Es deseable analizar el comportamiento con esta división ya que nos permite identificar y relativizar los cambios más sustanciales dependiendo de la procedencia de los ingresos. El cuadro 2.3 contiene el promedio nacional de cada una de las principales fuentes de ingreso, así como el peso que tienen con respecto al ingreso corriente total. Si bien el ingreso del trabajo no tuvo los incrementos de otras fuentes de ingreso, su peso dentro del ICT hace que sea un componente fundamental para recuperar la trayectoria histórica. Históricamente, la fuente de mayor importancia dentro del ICT son los ingresos del trabajo (principal y secundario), este componente ha representado poco menos de 70% del total.

Un aspecto a resaltar es que el comportamiento de estas dos fuentes de ingreso es distinto: los derivados por trabajo crecieron 2.9% entre el 2014 y 2016, mientras que el restante creció 26.9% en el mismo periodo (ver cuadro 2.4).

En términos del monto total de los ingresos por trabajo, la mayor variación porcentual con respecto al levantamiento previo se observó entre el 2014 y 2016 cuando aumentó 7.1% (a precios del 2016) (ver figura 2.3). Ello, en combinación con el cuadro 2.4, implica que el número de personas con ingresos laborales también tuvo un crecimiento sustancial.

Por otra parte, la gran diversidad de fuentes de ingresos captadas en la ENIGH no permite tener una contraparte para cada una de ellas, ya sea en encuestas o en agregados económicos; sin embargo, sí existe un levantamiento paralelo para el ingreso por trabajo: la ENOE; ésta, al igual que el MCS-ENIGH, tiene representatividad tanto a nivel nacional como por entidad federativa y ambas comparten el mismo marco muestral. La ENOE se levanta desde el 2005 con una periodicidad trimestral, cuya muestra rotatoria es de más de 120 mil viviendas; cada trimestre se sustituye 20% de las viviendas y cada vivienda es entrevistada durante cinco ocasiones

Cuadro 2.3

**Promedio por hogar para distintas fuentes
de ingreso y su peso relativo dentro del total (precios del 2016)**

	MCS-ENIGH 2010	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016
Promedio ingreso corriente	43 931 100%	42 115 100%	42 157 100%	46 521 100%
Ingreso del trabajo	30 109 69%	29 370 70%	29 059 69%	29 906 64%
Ingreso distinto del trabajo	13 822 31%	12 745 30%	13 098 31%	16 615 36%

Cuadro 2.4

Variación porcentual por fuente de ingreso (promedio de los hogares)

	2010 al 2012	2012 al 2014	2014 al 2016
Ingreso corriente total	-4.1%	0.1%	10.4%
Ingreso del trabajo	-2.5%	-1.1%	2.9%
Ingreso distinto del trabajo	-7.8%	2.8%	26.9%

Figura 2.3

**Variación porcentual del ingreso por trabajo
(volumen, precios del 2016)**



antes de ser reemplazada; esto permite que en cada trimestre se mantenga 80% de la muestra. Otra ventaja de la ENOE es que no fue objeto de las mejoras implementadas en la ENIGH 2016. El objetivo de la ENOE es obtener información sobre las características ocupacionales de la población, por lo que no obtiene información de todas las fuentes de ingresos de los hogares. Un problema bien identificado en la ENOE es el elevado número de personas que no reportan ingresos; el porcentaje es cercano a 60% del total de individuos a los cuales se les aplicó el *Cuestionario de ocupación y empleo*, en distintos levantamientos desde el 2010

a la fecha. Para aminorar esta situación, el cuestionario agrega un reactivo de rescate que pregunta por un rango salarial a aquellas personas que, aun cuando se les preguntó el monto de sus ingresos, no contestaron un ingreso puntual. Para incorporar las respuestas por rangos salariales, se tienen diferentes opciones, la que usamos en esta metodología es la misma que propone el CONEVAL en la construcción del índice de la tendencia laboral de la pobreza y que se encuentra detallada en el anexo del documento publicado en internet (http://www.coneval.org.mx/Informes/ITLP-IS/Nota%20ingresos%20e%20ITLP_IS.pdf).

La información que reporta el MCS-ENIGH sobre los ingresos de los hogares es de hasta seis meses anteriores a la fecha del levantamiento y, debido a que la obtención de la información es de agosto a noviembre, el periodo reportado en la encuesta es de febrero a octubre (de estos nueve meses, cada hogar solo reporta seis, dependiendo del mes en el que es encuestado). Por otra parte, y dado que la realización de la ENOE es trimestral, aquellos periodos donde

hay mayor coincidencia con la información del MCS-ENIGH son el segundo y el tercer trimestre; es por ello que, para los datos obtenidos de la ENOE, consideramos dichos periodos. La desventaja que esto conlleva es que en esos trimestres solo se recoge información del trabajo principal y se excluye lo correspondiente al secundario.

Como consecuencia de lo anterior, debemos identificar los ingresos obtenidos por el trabajo principal dentro del MCS-ENIGH y la ENIGH 2016 para hacerlos comparables con los de la ENOE. El MCS-ENIGH y la ENIGH 2016 generan información a nivel individuo y por tipo de ingreso recibido; este ingreso se encuentra clasificado por códigos y se pueden consultar en el documento metodológico de la encuesta. Los códigos que utilizaremos para generar la variable de ITP son los que se muestran en el cuadro 2.5 que, para fines prácticos, se pueden agrupar en ingresos por trabajo subordinado (columna a) y por trabajo independiente (columna b).

Para terminar de formar la variable ITP, tanto en el MCS-ENIGH como en la ENIGH 2016 y en la ENOE, sumamos los ingresos a nivel hogar. Para la ENOE se mantienen solo aquellos hogares en los que todos sus integrantes declaran algún ingreso (incluso si esa declaración es cero).

Como proporción del ICT, el ITP ha sido relativamente estable (61.5% en promedio) durante el periodo del 2010 al 2014, tal como se muestra en la figura 2.4. Dicha estabilidad será importante en nuestra metodología, como se explicará más adelante.

La figura 2.5 muestra el promedio nacional del ITP: en ella se observa una caída entre el 2010 y 2012 de más de 1 200 pesos; también se aprecia que el cambio más importante se dio entre el 2014 y 2016, con un crecimiento superior a los 2 000 pesos.

A grandes rasgos, la metodología usa la tasa de crecimiento del ITP_ENOE (ingreso por trabajo principal de la ENOE) dentro de cada entidad federativa, para ajustar el valor del ITP_ENIGH (ingreso por trabajo principal de la ENIGH) reportado por los hogares. Sin embargo, al ajustar solo el ITP, la

Cuadro 2.5

Apartados que forman la variable de ITP del MCS-ENIGH

a)

Código	Descripción
	Ingresos monetarios del trabajo principal para subordinados
P001	Sueldos, salarios o jornal
P002	Destajo
P003	Comisiones y propinas
P004	Horas extra
P005	Incentivos, gratificaciones o premios
P006	Bono, percepción adicional o sobresueldo
P007	Primas vacacionales y otras prestaciones en dinero

b)

Código	Descripción
	Ingresos por negocios del hogar, trabajo principal
P011	Sueldos o salarios
P012	Ganancias/utilidades
P013	Otros ingresos
	Ingresos por negocio propio, trabajo principal
P068	Por negocios con tipo de actividad industrial
P069	Por negocios con tipo de actividad comercial
P070	Por negocios prestadores de servicios
P071	Por negocios con actividades agrícolas
P072	Por negocios con actividades de cría y explotación de animales
P073	Por negocios con actividades de recolección, reforestación y tala de árboles
P074	Por negocios con actividades de pesca, caza y captura de animales

Figura 2.4

Proporción del ITP dentro del ICT en la ENIGH 2016

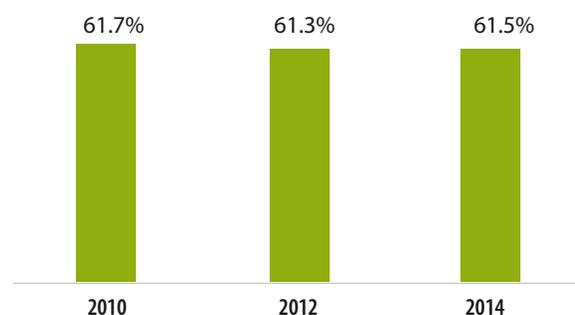
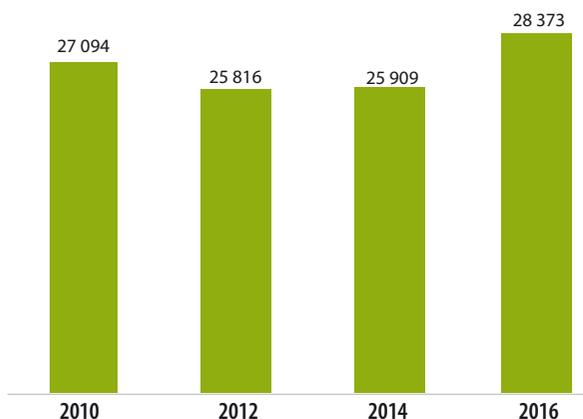


Figura 2.5

ITP promedio del MCS-ENIGH por hogar (montos trimestrales, precios del 2016)



proporción entre éste y el ICT se verá alterada. Si suponemos que esa relación debe ser similar a la observada históricamente (ver figura 2.4), ello nos permitiría tener un contrafactual para modificar el complemento del ITP (ITP^c). Así, el ajuste al ICT sería en dos fases: en la primera se modifica el ITP con base en el comportamiento de la ENOE, y en la segunda se modifica el ITP^c con base en el peso que a través del tiempo ha tenido el ITP dentro del ICT. En las siguientes secciones se detalla el proceso.

3. Metodología

La metodología toma como marco teórico el trabajo desarrollado por Bustos (2015). En su artículo, el autor propone que, además de aproximar una función de distribución a los datos que arrojan encuestas como la ENIGH, se tomen datos puntuales de fuentes como las cuentas nacionales para imponerlos como restricciones al ajuste de la distribución teórica y, así, corregir problemas como el de la subdeclaración. De esta forma, se tendrían estimaciones más cercanas a la realidad sobre la distribución del ingreso de los hogares.

Comencemos suponiendo que la variable de interés se puede modelar a través de una función de densidad de probabilidad paramétrica. Es decir, si dicha variable la denotamos como y , entonces $f(y; \theta)$ es la función de densidad asociada a

su función de distribución $F(y|\theta)$. Los valores del vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ son desconocidos y deben estimarse procurando errores mínimos de medición, de tal manera que se pueda reconstruir lo mejor posible la *distribución empírica* (aquella obtenida a partir de datos recopilados por medio de una encuesta, por ejemplo).

Recordemos que una *función de distribución acumulada teórica* representa las probabilidades acumuladas obtenidas de la función de densidad. Es denotada por $F_Y(y|\theta)$ y se interpreta como la probabilidad de que la variable Y asuma un valor menor o igual a y . Se define como:

$$F_Y(y|\theta) = P(Y \leq y) = \int_0^y f_Y(t|\theta) dt = p \quad (1)$$

Así, para un cierto intervalo de ingresos $[a, b]$, tenemos que:

$$P(a \leq y \leq b) = \int_a^b f_Y(y|\theta) dy = F_Y(b) - F_Y(a) = p_{a-b} \quad (2)$$

Debido a que la función $F(y|\theta)$ es estrictamente creciente, entonces existe su inversa, la cual es llamada *función cuantil*. Ésta es la que usaremos para obtener los microdatos estimados.

$$F^{-1}(p) = \min\{y : F_Y(y) \geq p\}; p \in (0,1) \quad (3)$$

Por otro lado, la *distribución acumulada empírica* es la frecuencia acumulada de los datos observados ordenados de forma ascendente; en nuestro caso, y como los datos provienen de un diseño de muestreo, usamos los factores de expansión para definir dicha frecuencia como la proporción de elementos (hogares) observados que tienen un valor menor o igual a un valor observado y . Se representan por la siguiente expresión:

$$p_{m_acum} = \frac{w_{m_acum}}{\sum_{h=1}^n w_h}; w_{m_acum} = \sum_{h=1}^m w_h$$

$$\text{para cada } m = 1, 2, \dots, n; \quad (4)$$

donde w_h es el factor de expansión asociado al hogar h y n es el número total de hogares en la muestra.

Ahora bien, una vez elegida la forma funcional paramétrica $f(y; \theta)$ lo siguiente es establecer el método de optimización que se usará para encontrar el estimador del vector de parámetros θ . Existen diversos métodos que cumplen con esta finalidad; entre los más comunes se encuentran el de momentos, el de mínimos cuadrados y el de máxima verosimilitud. Decidimos usar este último ya que, además de arrojar estimaciones con propiedades de eficiencia cuando se cumplen ciertas condiciones, hace máxima la probabilidad de que el modelo —dado por $f(y; \theta)$ — genere la muestra que se observó; también, ya hay un gran avance en la implementación computacional del algoritmo que resuelven de manera eficiente sistemas de ecuaciones no lineales.

El método consiste en tomar los n datos muestrales de la variable de interés, representada por el vector $Y = (y_1, y_2, \dots, y_n)$, y suponer que esta variable tiene una función de distribución asociada $f(y|\theta)$; el estimador de θ (el de máxima verosimilitud) se obtiene calculando el máximo³ de la función pseudo-log-verosimilitud para datos con pesos (Skinner *et al.*, 1989) denotada por $l(\theta)$ y definida como:

$$l(\theta) = \log L(\theta|Y) = \sum_{h=1}^n w_h \log f(y_h|\theta) \quad (5)$$

donde w_h corresponde al peso de la observación (hogar) h , con $h = 1, \dots, n$.

El valor de θ que maximiza la función anterior es llamado estimador de θ y lo denotaremos por $\hat{\theta}_{SR}$ (el subíndice hace referencia a que es el estimador de un ajuste sin restricciones).

³ La mayoría de usuarios que elige este método de estimación prefiere usar la siguiente regla de equivalencia: maximizar una función es matemáticamente equivalente a minimizar la expresión negativa de la función objetivo, es decir, $Max [l(\theta)] = Min [-l(\theta)]$; la preferencia radica en que, computacionalmente, los algoritmos están adaptados para la minimización de funciones.

Con respecto al modelo de ajuste con restricciones, éste puede ser expresado de la siguiente manera:

$$Max l(\theta) = \sum_{i=1}^n w_i \log f(y_i|\theta) \quad (6)$$

Sujeta a:

Restricciones de igualdad	y/o	Restricciones de desigualdad
$H_1(y, \theta) = c_1$		$I_1(y, \theta) \leq d_1$
$H_2(y, \theta) = c_2$		$I_2(y, \theta) \leq d_2$
...		...
$H_p(y, \theta) = c_p$		$I_p(y, \theta) \leq d_p$

siempre y cuando $p < \text{número de parámetros a estimar}$.⁴ c_i, d_i son constantes (valores objetivos).

De manera iterativa se encuentra el valor de θ que arroje un valor máximo de $l(\theta)$ y que cumpla con las restricciones impuestas. Es éste al que consideramos el mejor estimador y lo denotaremos como $\hat{\theta}_R$.

Las funciones $H_i(\cdot), I_i(\cdot)$, pueden estar en términos de la variable de interés y/o de los valores del vector-parámetro θ . Así, por ejemplo, si la intención es que la variable Y en un cierto rango $[0, c]$ tenga una probabilidad no mayor que c_1 , entonces, una de las restricciones tomaría la siguiente forma:

$$H_1(y, \theta) = \int_{y=0}^{y=c} f(y|\theta) dy = c_1$$

Si el objetivo es que la distribución teórica que ajusta a Y tenga una esperanza (promedio) igual a c_2 , entonces, una forma de expresar la función de restricción es:

$$H_2(y, \theta) = \int_{y=0}^{\infty} y f(y|\theta) dy = c_2$$

Por otro lado, es posible a través de la especificación del modelo con restricciones, modificar recursivamente los microdatos muestrales (es decir,

⁴ En caso contrario, el modelo expresado en (6) se convierte en la solución de un sistema de ecuaciones sin la dependencia de los valores observados de la variable aleatoria Y .

los datos de la encuesta) hasta obtener de ellos valores deseados de un estadístico muestral; así, por ejemplo, si deseamos que los nuevos microdatos (junto con sus factores de expansión w_h) tenga un promedio igual a c_3 , entonces la restricción de igualdad tendría la siguiente forma:

$$H_3(y, \theta) = \sum_{h=1}^n \frac{w_h * \hat{y}_{h|\theta}}{N} = c_3$$

donde $\hat{y}_{h|\theta}$ es el nuevo valor para la observación h , dados los valores del parámetro θ (en la sección 3.1 se propone una forma de obtener $\hat{y}_{h|\theta}$).

Éstos son solo algunos ejemplos de la forma que pueden tener las funciones que definen las restricciones; pueden variar según sea el propósito del modelo y, como veremos enseguida, es factible que a través del modelo restringido se puedan editar los valores de los microdatos.

3.1 Generación de nuevos microdatos

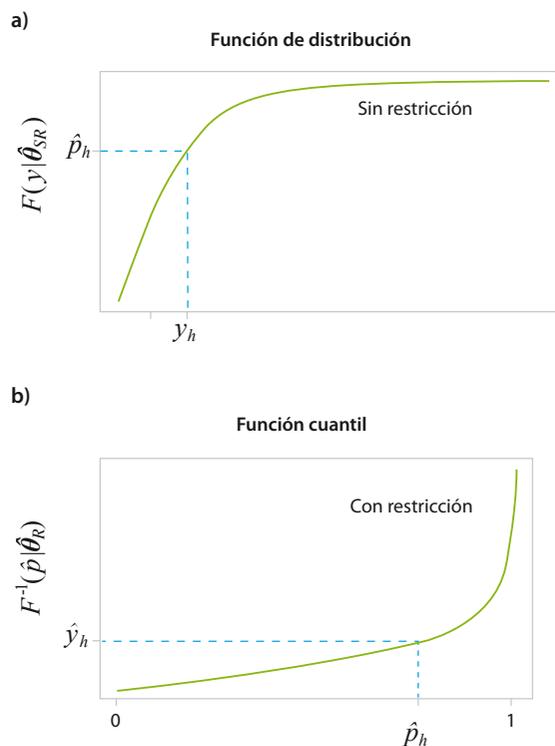
Bajo el supuesto de que $Y \sim f(y|\theta)$, y dados los estimadores $\hat{\theta}_{SR}$ y $\hat{\theta}_R$, la forma de obtener los nuevos valores empíricos es la siguiente: se calcula la probabilidad acumulada, $\hat{p}_h = F(y_h|\hat{\theta}_{SR})$, que le correspondería al hogar muestral h de acuerdo con el valor y_h reportado; para ello, utilizamos la función de distribución acumulada teórica definida en (1). Una vez obtenida \hat{p}_h para cada hogar de la muestra, nos olvidamos de la función de probabilidad sin restricciones y usamos la función inversa de la función acumulada con restricciones (determinada por el vector $\hat{\theta}_R$); es decir, la *función cuantil* especificada en (3). Esta función toma un valor \hat{p}_h y arroja un valor de \hat{y}_h , $F^{-1}(\hat{p}_h|\hat{\theta}_R) = \hat{y}_h$; de esta forma obtenemos, para cada hogar muestral, un nuevo valor de la variable y a través de una función de probabilidad con restricciones.

La figura 3.1 muestra de manera esquemática este proceso: la gráfica a) ejemplifica cómo a partir de un monto específico, y_h , se obtiene una probabilidad acumulada, \hat{p}_h , a través de la función acumulada *sin* restricciones; este proceso se repite

para cada uno de los hogares de la muestra. Una vez fijadas las probabilidades acumuladas para cada hogar, y con ayuda de la *función cuantil* obtenida mediante el proceso de optimización *con* restricciones, se genera un nuevo valor \hat{y}_h , tal como se muestra en la gráfica b).

Figura 3.1

Visualización del proceso para la obtención de microdatos



4. Ajuste al ingreso por trabajo principal

El ajuste al ingreso total de los hogares se realizará en dos etapas secuenciales, primero se modificarán los ingresos correspondientes al ITP y, después, los que forman el ICT, pero que son distintos al ITP (a los cuales llamamos complemento del ITP, ITP^c).

Con base en la metodología desarrollada en el apartado anterior, en esta sección se detalla la mo-

dificación de microdatos relacionados con el ITP. El modelo requiere de un valor objetivo que debe alcanzar un estadístico deseado que, en nuestro caso, será la mediana del ICT de la ENIGH 2016. Para generar el valor objetivo que se debe alcanzar, se siguen estos pasos: en cada entidad federativa por separado se toman las medianas del ingreso reportado dentro del segundo y tercer trimestres de la ENOE para el 2014 y 2016. Con ellas se obtienen los promedios entre los dos trimestres del mismo año. Posteriormente, se calculan las variaciones porcentuales del promedio de las medianas del 2016 con respecto al del 2014. De esta manera, tenemos 32 tasas de crecimiento de las medianas entre el 2014 y 2016. Enseguida, y para cada entidad federativa, se calculan las medianas del ICT del MCS-ENIGH 2014; a éstas les aplicamos la tasa de crecimiento obtenida con los datos de la ENOE para, finalmente, obtener una mediana objetivo para cada entidad en el 2016 que denotaremos como $Me = (Me_1, Me_2, \dots, Me_{32})$.

Una vez que hemos determinado tanto el estadístico deseado como su valor objetivo a alcanzar generamos nuevos valores del ICT para cada hogar, de acuerdo con la metodología de la sección anterior. Sin embargo, y para lograr ajustes diferenciados entre fuentes de ingreso, los valores originales del ICT y los nuevos valores son usados solo para modificar el ITP (como más adelante se explica), mientras que el ITP^c se modificará en función de la relación histórica entre el ITP y el ICT (ver figura 2.4).

Una de las primeras tareas a resolver es la elección de la forma funcional que mejor se ajusta a los ingresos observados. En el Anexo 1 se especifican las formas funcionales de cuatro funciones paramétricas⁵ elegidas para ser sometidas a pruebas de ajuste: *Gama Generalizada (GG)*, *Beta Generalizada del segundo tipo (GB2)*, *LogNormal (LN)* y *Dagum*. En el Anexo 2 se incluyen los resultados de los estadísticos de ajuste a partir de los cuales elegimos la función de distribución con la que trabajaremos.

5 Éstas fueron elegidas debido a que han sido muy usadas en la modelación del ingreso en distintos países con resultados satisfactorios.

Con base en dichos criterios, se determinó que la GB2 con cuatro parámetros (μ, σ, ν, τ) es la que mejor se ajusta a los ingresos reportados del ICT a nivel nacional.

Definida la función $f(y|\theta) = GB2(\theta) = GB2(\mu, \sigma, \nu, \tau)$, el proceso formal para ajustar el ingreso ICT en cada entidad es como se expresa a continuación.

Sea $Y^i = (y_1^i, y_2^i, \dots, y_n^i)$ el vector de ingreso corriente total para la entidad i ; sus pesos (factores de expansión) correspondientes están dados por $w^i = (w_1^i, w_2^i, \dots, w_n^i)$. Entonces, el ajuste de la distribución a los datos empíricos que corresponde a maximizar la *log*-verosimilitud de la $GB2(\theta)$ se expresa de la siguiente manera:

Para cada entidad $i = 1, 2, 3, \dots, 32$

$$\max l(\theta^i | Y^i) = \sum_{h=1}^{n_i} w_h^i \log f(y_h^i | \theta^i) \quad (7)$$

Resultado de estos ajustes obtendremos 32 estimadores de los parámetros de la distribución, los cuales representamos como $\hat{\theta}_{SR} = (\hat{\theta}_{SR}^1, \hat{\theta}_{SR}^2, \dots, \hat{\theta}_{SR}^{32})$, donde $\hat{\theta}_{SR}^i = (\hat{\mu}_{SR}^i, \hat{\sigma}_{SR}^i, \hat{\nu}_{SR}^i, \hat{\tau}_{SR}^i)^T$.

Posterior al ajuste obtenido, se realiza un nuevo ajuste en el que incluimos como restricción que los nuevos microdatos generen la *mediana objetivo* para la entidad federativa correspondiente.

Es decir, sea $Y_o^i = (y_{(1)}^i, y_{(2)}^i, \dots, y_{(n)}^i)$ el vector ordenado de ingresos para la entidad i y sus correspondientes pesos muestrales (factores de expansión) denotados por $w_o^i = (w_{(1)}^i, w_{(2)}^i, \dots, w_{(n)}^i)$; y sea el vector de medianas objetivo $Me = (Me_1, Me_2, \dots, Me_{32})$; para cada entidad $i = 1, 2, 3, \dots, 32$ se resuelve el siguiente problema de optimización restringida:

$$\max l(\theta^i) = \sum_{h=1}^{n_i} w_h^i \log f(y_h^i | \theta^i) \quad (8)$$

sujeto a:

a) Restricciones de igualdad:

$$1. F^{-1}(F(y_{(k)}^i | \hat{\theta}_{SR}^i) | \theta^i) = Me_i$$

donde:

$y^{(k)}$ es el ingreso asociado al subíndice k tal que $0.5 \leq \frac{W^{(k)}}{W^{(n)}}$ y $0.5 \leq 1 - \frac{W^{(k-1)}}{W^{(n)}}$ con $w_{(j)} = \sum_{h=1}^j w_{(h)}^i$

Me_i es la mediana objetivo de la entidad federativa i .

$\hat{\theta}_{SR}^i$ es el estimador de los parámetros de la función teórica de densidad sin restricciones para la entidad i .

$F(\cdot)$ es la función de probabilidad acumulada de $GB2$.

$F^{-1}(\cdot)$ es la función cuantil de $GB2$ y

$$2. \int_0^{max^i} f(y|\theta^i) dy = p^{max}(\hat{\theta}_{SR}^i)$$

donde:

$max^i = y_{(n)}^i$ es el valor máximo del ingreso de ICT en la entidad i .

$p^{max}(\hat{\theta}_{SR}^i) = F(y_{(n)}^i|\hat{\theta}_{SR}^i)$ es la probabilidad acumulada (bajo la función teórica sin restricciones) correspondiente al valor max en la entidad i .

- b) Restricciones de desigualdad: las propias del dominio de los valores de los parámetros de la función densidad, para $GB2$ (μ, σ, v, τ):

$$\mu, v, \tau > 0; \quad -\infty < \sigma < \infty; \quad -v < \frac{1}{\sigma} < \tau.$$

La segunda restricción de igualdad tiene como objetivo controlar el carácter no finito en el dominio de las funciones de distribución. Para lograrlo, tomamos el valor más grande del ICT que se obtuvo en la encuesta en cada entidad federativa y establecemos que los nuevos ingresos estimados estén en $[0, max^i]$.

Derivado de los modelos optimizados, obtenemos un vector que contiene los estimadores de los parámetros restringidos de la función $GB2$, que denotamos como $\hat{\theta}_R = (\hat{\theta}_R^1, \hat{\theta}_R^2, \dots, \hat{\theta}_R^{32})$, donde $\hat{\theta}_R^i = (\hat{\mu}_R^i, \hat{\sigma}_R^i, \hat{v}_R^i, \hat{\tau}_R^i)^T \forall i$.

Lo siguiente es calcular los nuevos microdatos del ICT; para ello, y como se explicó en la sección 3, utilizamos tanto la función de distribución acumulada como la cuantil. Primero, tomamos el valor reportado del ICT para cada hogar y , de acuerdo con la entidad federativa de pertenencia, se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_h^i = F(ICT_h^i|\hat{\theta}_{SR}^i) \forall h = 1, 2, \dots, n_i$; donde n_i es el número de hogares en la entidad i . Enseguida, se obtiene el valor estimado del ICT_{fase1} para cada hogar usando la función cuantil: $ICT_{fase1}_h^i = F^{-1}(\hat{p}_h^i|\hat{\theta}_{SR}^i) \forall h = 1, 2, \dots, n_i$.

De esta manera, obtenemos para cada hogar en toda la muestra un valor del ICT imputado que está determinado en función de los resultados del ajuste por entidad, $ICT_{fase1_h} = \hat{y}_h \forall h = 1, 2, \dots, n$.

Finalmente, el nuevo vector del ingreso por trabajo principal ($ITP_{ajustado}$) se obtiene como sigue:

- Para cada hogar de la muestra tenemos dos valores: $ICT_{original}$ y el ICT_{fase1} . Esto nos permite obtener un factor de corrección para cada hogar h ,

$$f = \frac{ICT_{fase1}}{ICT_{original}} \triangleq \left(\frac{\hat{y}_1}{y_1}, \frac{\hat{y}_2}{y_2}, \dots, \frac{\hat{y}_n}{y_n} \right) = (f_1, f_2, \dots, f_n)$$

- Calculamos un nuevo valor del ITP para cada hogar de la siguiente manera:

$$\begin{aligned} \widehat{ITP}_{ajustado} &= ITP_{original} * f \triangleq (y_{ITP_1} f_1, y_{ITP_2} f_2, \dots, y_{ITP_n} f_n) \\ \widehat{ITP}_{ajustado} &= (\hat{y}_{ITP_1}, \hat{y}_{ITP_2}, \dots, \hat{y}_{ITP_n}) \end{aligned} \tag{9}$$

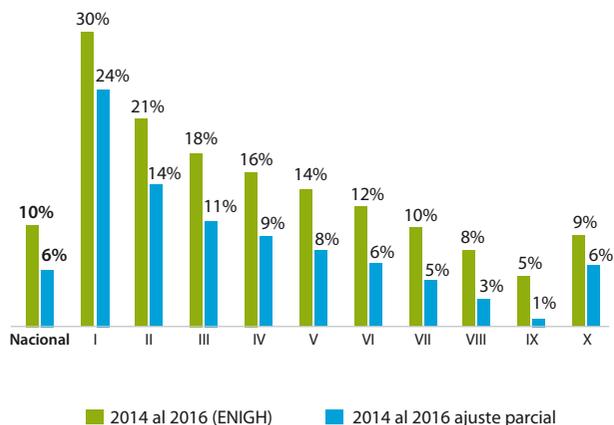
El factor de ajuste f solo es aplicado al ITP debido a que buscamos ajustes diferenciados (no en la misma proporción) para el ITP y para el ITP^c , ya que las mejoras operativas pudieron afectarlos de manera distinta (ver cuadro 2.4).

4.1 Resultados del ajuste al ITP de los hogares⁶

En este apartado presentamos los resultados del ajuste realizado al ingreso por trabajo principal (definido en el cuadro 2.5) de los hogares. El ITP promedio por hogar fue de 25 909 en el 2014, mientras que en la ENIGH 2016 fue de 28 373, ello representaría un incremento de 9.5% en términos reales. Con el ajuste descrito en el apartado anterior al ITP 2016, el promedio baja a 26 467, lo que equivale a un aumento de 2.1% con respecto al 2014.

Si solo ajustamos el ITP y mantenemos el ITP^c originalmente reportado, el promedio por hogar del ICT sería de 44 615 pesos, equivalente a un aumento de 5.8% con respecto al MCS-ENIGH 2014. La figura 4.1 muestra las variaciones del ICT por decil que se tendrían al ajustar solo el ITP, es decir, bajo un ajuste parcial.

Figura 4.1
Variación porcentual del ICT del 2014 al 2016



⁶ El software estadístico que se utilizó para implementar la metodología descrita fue R versión 3.4.0 (de uso libre) a través de su IDE RStudio versión 1.0.143. Las librerías usadas son:

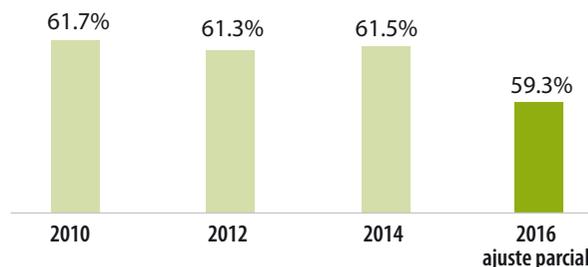
- foreign_0.8-67
- gamlss_5.0-2
- alabama_2015.3-1
- survey_3.32
- plyr_1.8.4
- dplyr_0.7.0
- GB2_2.1

Los resultados pueden variar dependiendo de las características del sistema de la computadora utilizada; el procesamiento del algoritmo fue realizado en un equipo con las siguientes características: Lanix con procesador AMD FX-8370 de ocho núcleos y 4.00 GHz, sistema operativo de 64-bits.

5. Ajuste al complemento del ITP

Un efecto colateral del ajuste realizado al ingreso por trabajo principal es que la proporción que históricamente guardaba (hasta antes del 2014) dentro del ICT se rompe con el ajuste de la sección anterior (ver figura 5.1). Este efecto es consecuencia de que solo se modifica una parte del ICT.

Figura 5.1
Proporción del ITP dentro del ICT



Si suponemos que en ausencia de las mejoras implementadas en la ENIGH 2016 la proporción entre el ITP y el ICT hubiese sido similar a la histórica (61.5% en promedio del 2010 al 2014),⁷ entonces podemos calcular el total del ingreso que no es ITP, tal que la proporción entre el ITP e ICT sea consistente con lo histórico. Es decir, llamemos ITP^c a la diferencia entre el ICT y el ITP. Después del ajuste al vector del ITP ($ITP^{ajustado}$) observamos que, a nivel nacional, los totales de cada ingreso cumplen con la siguiente relación:

$$\frac{T_{ITP^{ajustado}}}{T_{ICT^{ajuste\ parcial}}} \% = \frac{T_{ITP^{ajustado}}}{T_{ITP^{ajustado}} + T_{ITP^c}} \% = 59.3\%$$

Sin embargo, esa proporción debería ser cercana a 61.5% para hacerla consistente con lo que se venía observando hasta el 2014. De ser así, y dado que el valor del $T_{ITP^{ajustado}}$ es conocido, lo único que tendríamos que hacer para conocer el monto del T_{ITP^c} ajustado es el siguiente despeje:

$$\begin{aligned} \frac{T_{ITP^{ajustado}}}{T_{ICT^{ajuste\ total}}} \% &= \frac{T_{ITP^{ajustado}}}{T_{ITP^{ajustado}} + T_{ITP^c\ ajustado}} \% = 61.5\% \\ \Rightarrow T_{ITP^c\ ajustado} &= \frac{T_{ITP^{ajustado}}(1 - 0.615)}{0.615} \end{aligned}$$

⁷ La razón de tomar la proporción histórica (y no la observada en la ENIGH 2016) es que el modelo tiene como objetivo hacer los datos de la ENIGH 2016 comparables con la trayectoria que se venía observando, incluida la relación entre el ITP y su complemento.

Si además dividimos lo anterior entre el número total de hogares, tendríamos el promedio por hogar que garantiza obtener 61.5%:

$$prom_ITP^c\ ajustado = \frac{T_{ITP^c}\ ajustado}{número\ de\ hogares\ poblacionales} \quad (10)$$

Para distribuir el ajuste del ITP^c a los hogares, utilizamos de nuevo el marco metodológico desarrollado en la sección 3 y realizamos un procedimiento simétrico al de la sección anterior, aunque en este caso se hace un solo ajuste a nivel nacional; es decir, ajustamos una función de distribución $GB2$ al vector ITP^c para obtener un conjunto de parámetros $\hat{\theta}_{SR}^{ITP^c}$; después imponemos como restricción que el nuevo promedio de todos los microdatos del ITP^c sea igual al $prom_ITP^c\ ajustado$, de esta forma obtenemos el segundo conjunto de parámetros $\hat{\theta}_R^{ITP^c}$.

Formalmente, el modelo con restricciones a resolver es el siguiente: sea $H = \{h: 1, 2, \dots, n\}$ conjunto de hogares en la muestra, si suponemos que $ITP^c \sim GB2(z|\theta)$, entonces, el modelo con restricciones quedaría de la siguiente forma:

$$max_{\theta} l(\theta) = \sum_{i=h}^n w_h \log f(ITP_h^c|\theta) \quad (11)$$

sujeto a:

a) Dos restricciones de igualdad.

$$\frac{\sum_H w_h F^{-1}(F(ITP_h^c|\hat{\theta}_{SR}^{ITP^c})|\theta)}{\sum_H w_h} = prom_ITP^c\ ajustado$$

$$\int_0^{max} f(ITP_h^c|\theta) dITP_h^c = p^{max}(\hat{\theta}_{SR}^{ITP^c})$$

b) Restricciones de desigualdad: $\mu, v, \tau > 0$;
 $-\infty < \sigma < \infty$; $-v < -\frac{1}{\sigma} < \tau$.

Donde max es el máximo valor del ITP^c y $p^{max}(\hat{\theta}_{SR}^{ITP^c})$ es la probabilidad acumulada (bajo la función teórica sin restricciones de ITP^c) correspondiente al valor max . F y f son las funciones acumulada y de densidad de la $GB2$, respectivamente. De este proceso se obtienen los parámetros restringidos $\hat{\theta}_R^{ITP^c}$.

La generación de los nuevos microdatos del ITP^c se realizó de la siguiente manera (ver figura 3.1): se toma el valor reportado del ITP^c para cada hogar y se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_h^{ITP^c} = F(ITP_h^c|\hat{\theta}_{SR}^{ITP^c}) \forall h = 1, 2, \dots, n$, donde n es el número de hogares en la muestra. El nuevo valor estimado del ITP^c para cada hogar será entonces $\hat{ITP}_h^c = F^{-1}(\hat{p}_h^{ITP^c}|\hat{\theta}_R^{ITP^c}) \forall h = 1, 2, \dots, n$.

Así, el nuevo vector del ICT, \hat{ICT} , tendría un ajuste secuencial en dos etapas: 1) una donde se ajusta solo el vector de ITP y 2) otra donde solo se ajusta el vector de ITP^c . Los datos por hogar del nuevo vector del ICT, \hat{ICT}_h , están formados por cada uno de los dos procesos:

$$\hat{ICT}_h = \hat{ITP}_h + \hat{ITP}_h^c \quad (12)$$

Por otra parte, y dado que el objetivo del ajuste es poder medir pobreza, es necesario trasladar el ajuste del ICT a todas las fuentes de ingreso monetario y no monetario que lo componen. Para transferir el efecto del ajuste, se calculan dos factores por hogar: $\alpha_h = \frac{ITP_h}{ITP_h^c}$ y $\beta_h = \frac{ITP_h^c}{ITP_h}$. Cada fuente de ingreso que forma el ITP se multiplica por α_h , mientras que las fuentes de ingreso que forman el ITP^c se multiplican por β_h . Así se generan nuevas versiones de las tablas usadas por la metodología del CONEVAL para el cálculo de la pobreza.

6. Resultados

En los cuadros 6.1 y 6.2 se observa cómo con la ENIGH 2016 se obtienen ingresos considerablemente mayores a los históricamente reportados en los distintos deciles; de igual forma, podemos ver la

magnitud del ajuste con el modelo estadístico 2016 y cómo éste recupera la continuidad histórica del ingreso. Si bien el primer decil es el que mayor crecimen-

to presentó en términos porcentuales, el aumento es similar al registrado entre el 2012 y 2014 (6.2%), periodo donde también fue el de mayor crecimiento.

Cuadro 6.1

**Promedio del ICT por deciles de hogares
(precios constantes 2016)**

Decil	MCS-ENIGH 2008	MCS-ENIGH 2010	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016	Modelo estadístico 2016
Nacional	46 901	43 931	42 115	42 157	46 521	43 036
I	6 270	6 111	5 920	6 288	8 166	6 820
II	12 318	11 707	11 310	11 731	14 206	12 350
III	17 230	16 285	15 750	16 062	18 918	16 677
IV	22 162	20 876	20 326	20 336	23 556	21 029
V	27 599	26 098	25 357	25 258	28 813	25 920
VI	34 228	32 400	31 502	31 030	34 837	31 650
VII	43 037	40 352	39 343	38 489	42 431	38 841
VIII	55 119	52 028	50 494	49 463	53 383	49 218
IX	76 315	72 580	69 321	68 386	72 041	67 034
X	174 735	160 872	151 828	154 524	168 856	160 820

Cuadro 6.2

**Variación porcentual del ICT por deciles de hogares
(precios constantes 2016)**

Decil	2008-2010	2010-2012	2012-2014	2014-2016	2014-Modelo estadístico 2016
Nacional	-6.3%	-4.1%	0.1%	10.4%	2.1%
I	-2.5%	-3.1%	6.2%	29.9%	8.5%
II	-5.0%	-3.4%	3.7%	21.1%	5.3%
III	-5.5%	-3.3%	2.0%	17.8%	3.8%
IV	-5.8%	-2.6%	0.1%	15.8%	3.4%
V	-5.4%	-2.8%	-0.4%	14.1%	2.6%
VI	-5.3%	-2.8%	-1.5%	12.3%	2.0%
VII	-6.2%	-2.5%	-2.2%	10.2%	0.9%
VIII	-5.6%	-2.9%	-2.0%	7.9%	-0.5%
IX	-4.9%	-4.5%	-1.3%	5.3%	-2.0%
X	-7.9%	-5.6%	1.8%	9.3%	4.1%

Con respecto al comportamiento entre entidades, la ENIGH 2016 logra captar dos hogares cuyos ingresos son muy elevados, uno de ellos fue en Nuevo León y otro en Guanajuato. Esto es consecuencia del tamaño de muestra usado (por primera vez la ENIGH 2016 tiene representatividad a

nivel entidad federativa y para los ámbitos rural y urbano), lo cual aumenta la probabilidad de captar hogares ricos o con ingresos muy elevados. Tanto en el cuadro 6.3 como en la figura 6.1 se observa la corrección del modelo estadístico para las distintas entidades federativas.

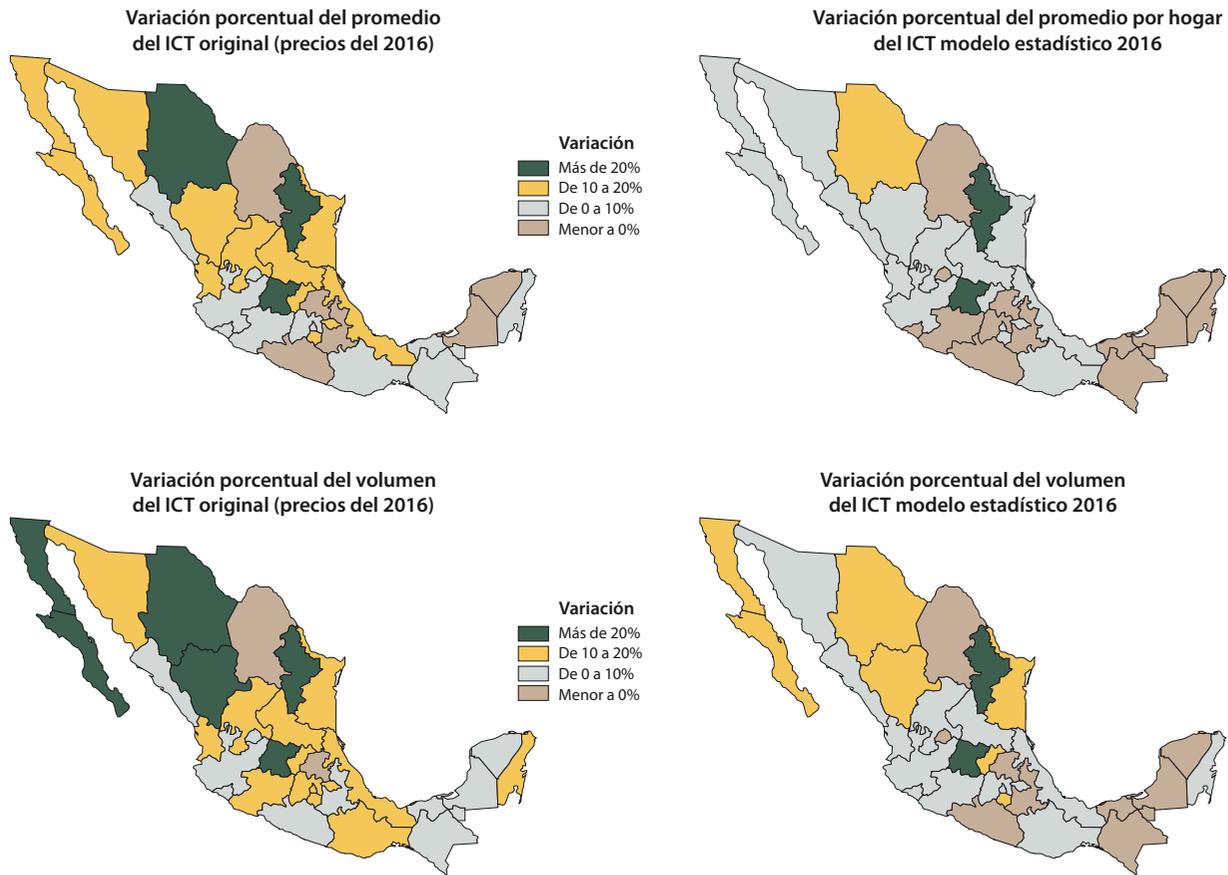
Cuadro 6.3

Promedio del ingreso corriente de hogares por entidad

Entidad federativa	MCS-ENIGH 2008	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016	Modelo estadístico 2016
1 Aguascalientes	52 572	45 394	48 870	49 247	45 938
2 Baja California	60 507	50 648	50 401	57 687	54 462
3 Baja California Sur	71 989	53 564	50 719	56 351	53 270
4 Campeche	46 521	44 145	45 641	44 211	42 055
5 Coahuila de Zaragoza	48 333	47 841	52 858	46 588	46 157
6 Colima	56 067	44 829	46 726	47 733	43 398
7 Chiapas	24 185	24 798	23 049	23 258	21 836
8 Chihuahua	55 842	44 097	39 872	48 111	45 228
9 Ciudad de México	73 834	58 490	66 738	70 834	65 241
10 Durango	37 355	34 499	34 305	40 384	37 998
11 Guanajuato	43 780	42 319	36 747	56 974	51 233
12 Guerrero	28 545	24 521	27 538	26 980	25 051
13 Hidalgo	33 689	33 173	33 717	32 535	31 418
14 Jalisco	54 407	47 333	48 705	52 367	47 864
15 México	45 204	44 732	40 992	43 319	39 217
16 Michoacán de Ocampo	35 279	33 787	31 475	33 788	30 113
17 Morelos	40 802	38 827	36 907	41 196	36 105
18 Nayarit	40 537	35 616	38 818	43 064	40 031
19 Nuevo León	71 316	63 974	63 108	87 653	90 394
20 Oaxaca	29 315	27 924	25 324	27 704	23 382
21 Puebla	32 458	30 915	37 759	35 369	31 094
22 Querétaro	57 072	49 385	49 759	56 605	51 163
23 Quintana Roo	55 184	44 497	49 387	50 568	49 623
24 San Luis Potosí	37 998	35 334	34 950	39 051	34 371
25 Sinaloa	52 145	43 785	45 085	48 591	44 361
26 Sonora	59 988	54 736	53 185	58 813	53 882
27 Tabasco	35 967	38 318	36 845	37 383	35 134
28 Tamaulipas	46 734	43 032	42 332	48 049	44 110
29 Tlaxcala	33 487	31 923	30 903	34 050	32 178
30 Veracruz de I. de la Llave	36 242	32 262	29 794	32 983	28 716
31 Yucatán	45 876	39 383	42 681	42 457	41 067
32 Zacatecas	36 712	32 469	31 938	36 246	33 903
Nacional	46 907	42 115	42 157	46 521	43 036

Figura 6.1

Variación porcentual del ICT ajustado del 2014 al 2106 (precios del 2016)



Con respecto a los cálculos de pobreza (ver cuadro 6.4), se consideró el ajuste a la carencia de seguridad social (adicional a los ajustes al ICT desarrollados en secciones anteriores) realizado por el CONEVAL: previo al levantamiento de la ENIGH 2016, el CONEVAL solicitó el cambio de fraseo para algunas preguntas relacionadas con el acceso a los servicios médicos y con otras dos prestaciones laborales que son consideradas para la identificación de la condición de carencia por seguridad social de la población trabajadora subordinada. Dicho cambio tuvo un efecto sobre la consistencia histórica de las respuestas que dan los informantes. La base de datos original de la ENIGH 2016 reporta tal cual los resultados del levantamiento en campo que resultan del nuevo fraseo. Sin embargo, buscando recuperar la continuidad de las series, se incorporan los resultados de un ejercicio estadístico que el CONEVAL diseñó para tal propósito. El ajuste se focaliza en la población trabajadora subordinada con

acceso a servicios médicos como prestación laboral para recuperar la consistencia de la serie histórica de la carencia por acceso a la seguridad social.⁸

6.1 Estimación de los errores del modelo

La metodología desarrollada en este documento no es parte de las metodologías tradicionales de imputación/edición de microdatos, por lo que hasta la fecha no se cuenta con procedimientos estadísticos para la cuantificación de los errores en las estimaciones que arroja. Sin embargo, y aprovechando procesos como el de *bootstrap*, podemos generar mediciones que permitan tener un acercamiento a errores del modelo. Para ello, los cuadros 6.5 y 6.6 muestran el resultado de 500 repeticiones

⁸ Para mayores detalles véase: http://www.coneval.org.mx/Medicion/MP/Documents/Pobreza_16/Notas_Pobreza_2016/Nota_tecnica_2_ajuste_seguridad_social.pdf

del algoritmo de ajuste del ingreso, con remuestras distintas obtenidas a través de *bootstrap*.

Como se observa en ambos cuadros, el promedio de las 500 repeticiones es muy cercano a los datos reportados en los cuadros 6.1 y 6.3. Las en-

tidades donde se observan las mayores distancias entre ambas estimaciones son Guanajuato, Nuevo León y Tamaulipas; de ellas, y como ya se ha mencionado, en las dos primeras se captaron hogares con ingresos muy elevados, lo que puede estar influyendo en las estimaciones.

Cuadro 6.4

Porcentaje de población en pobreza multidimensional

Entidad federativa	2010	2012	2014	Modelo estadístico 2016
Aguascalientes	38.1	37.8	34.8	28.2
Baja California	31.5	30.2	28.6	22.2
Baja California Sur	31.0	30.1	30.3	22.1
Campeche	50.5	44.7	43.6	43.8
Coahuila de Zaragoza	27.8	27.9	30.2	24.8
Colima	34.7	34.4	34.3	33.6
Chiapas	78.5	74.7	76.2	77.1
Chihuahua	38.8	35.3	34.4	30.6
Ciudad de México	28.5	28.9	28.4	27.6
Durango	51.6	50.1	43.5	36.0
Guanajuato	48.5	44.5	46.6	42.4
Guerrero	67.6	69.7	65.2	64.4
Hidalgo	54.7	52.8	54.3	50.6
Jalisco	37.0	39.8	35.4	31.8
México	42.9	45.3	49.6	47.9
Michoacán de Ocampo	54.7	54.4	59.2	55.3
Morelos	43.2	45.5	52.3	49.5
Nayarit	41.4	47.6	40.5	37.5
Nuevo León	21.0	23.2	20.4	14.2
Oaxaca	67.0	61.9	66.8	70.4
Puebla	61.5	64.5	64.5	59.4
Querétaro	41.4	36.9	34.2	31.1
Quintana Roo	34.6	38.8	35.9	28.8
San Luis Potosí	52.4	50.5	49.1	45.5
Sinaloa	36.7	36.3	39.4	30.8
Sonora	33.1	29.1	29.4	27.9
Tabasco	57.1	49.7	49.6	50.9
Tamaulipas	39.0	38.4	37.9	32.2
Tlaxcala	60.3	57.9	58.9	53.9
Veracruz de I. de la Llave	57.6	52.6	58.0	62.2
Yucatán	48.3	48.9	45.9	41.9
Zacatecas	60.2	54.2	52.3	49.0
Nacional	46.1	45.5	46.2	43.6

Cuadro 6.5

Resultados por decil del proceso *bootstrap*

Decil	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
I	6 820	6 802	27.58
II	12 350	12 367	29.08
III	16 677	16 719	30.27
IV	21 029	21 082	31.99
V	25 920	25 978	33.57
VI	31 650	31 717	35.55
VII	38 841	38 915	37.99
VIII	49 218	49 338	42.82
IX	67 034	67 237	54.91
X	160 820	160 927	374.41

Cuadro 6.6

Continúa

Resultados por entidad federativa del proceso *bootstrap*

Entidad	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
Aguascalientes	45 938	46 022	51.92
Baja California	54 462	54 484	62.41
Baja California Sur	53 270	53 265	60.29
Campeche	42 055	42 205	55.68
Coahuila de Zaragoza	46 157	46 258	52.13
Colima	43 398	43 694	56.25
Chiapas	21 836	21 928	36.06
Chihuahua	45 228	45 473	64.34
Ciudad de México	65 241	65 592	79.63
Durango	37 998	38 144	43.88
Guanajuato	51 233	50 718	419.04
Guerrero	25 051	25 192	41.21
Hidalgo	31 418	31 426	52.13
Jalisco	47 864	47 781	70.13
México	39 217	39 300	49.45

Resultados por entidad federativa del proceso *bootstrap*

Entidad	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
Michoacán de Ocampo	30 113	30 234	40.65
Morelos	36 105	36 233	52.32
Nayarit	40 031	40 136	69.32
Nuevo León	90 394	89 831	880.87
Oaxaca	23 382	23 594	47.56
Puebla	31 094	31 290	52.01
Querétaro	51 163	51 386	66.7
Quintana Roo	49 623	49 818	57.53
San Luis Potosí	34 371	34 543	45.52
Sinaloa	44 361	44 375	52.13
Sonora	53 882	54 106	138.13
Tabasco	35 134	35 054	48.55
Tamaulipas	44 110	44 540	99.72
Tlaxcala	32 178	32 272	40.82
Veracruz de I. de la Llave	28 716	28 888	48.67
Yucatán	41 067	41 181	73.82
Zacatecas	33 903	34 006	73.55

7. Conclusiones

En un proceso continuo de mejora, el INEGI identificó áreas de oportunidad alrededor de la capacitación, supervisión y control operativo del MCS-ENIGH. A partir de ello, se llevaron a cabo acciones de fortalecimiento operativo en el levantamiento de la ENIGH 2016, en un esfuerzo por alcanzar un mayor apego al diseño de la encuesta.

La mayoría de los componentes del ICT tuvo un crecimiento considerable en la ENIGH 2016 con respecto a la información del MCS-ENIGH 2014. Uno de ellos es el ingreso del trabajo, el cual creció

en este periodo 7.1% (para 2010-2012 y 2012-2014, las tasas de crecimiento fueron de 3.2 y 1.4%, respectivamente). Este rubro del ingreso ha representado más de 60% del ICT históricamente.

Los ingresos no laborales también crecieron de manera significativa; en términos del promedio por hogar, su crecimiento fue de 27% entre el 2014 y 2016, mientras que para el periodo 2012 al 2014 fue de solo 2.8 por ciento.

Debido al carácter único de la información que recopila la ENIGH, no existen otras encuestas con las cuales podamos comparar el comportamiento de sus

distintas fuentes de ingreso. Sin embargo, la ENOE nos permite hacerlo para la fuente principal del ingreso de los hogares: los ingresos por trabajo principal. Una característica fundamental de la ENOE es que la recolección de información ha sido similar en los distintos levantamientos y que, por lo tanto, no sufrió cambios importantes en la captación de ingresos.

La metodología que aquí se propone edita los microdatos reportados a través de funciones de distribución. Primero, se ajusta una función de distribución conocida (*GB2*, en nuestro caso) a los datos muestrales y, con ello, se obtienen estimaciones de los parámetros que determinan la forma funcional de la función de distribución. Después, se realiza un proceso similar, pero imponiendo un conjunto de restricciones que debe cumplir el ajuste; estas condiciones involucran la obtención de valores puntuales (valores objetivo) de estadísticos muestrales, lo cual implica la modificación de los datos reportados en la muestra.

El proceso de ajuste se realizó por separado para el ITP y para los ingresos distintos a estos últimos (*ITP^c*). Para obtener los valores objetivo de las restricciones impuestas, nos basamos en el crecimiento de la mediana de ingresos reportada en la ENOE; con ello modificamos el ITP. Los valores objetivo para las restricciones que entran en el ajuste al *ITP^c* se obtuvieron de una relación empírica que se ha venido observando desde el 2010 entre el *ITP^c* y el ICT, y la cual asumimos que se debe mantener.

En términos de resultados, y considerando ambos ajustes (tanto al ITP como al *ITP^c*), se logra que la variación entre el 2014 y 2016 del ICT promedio por hogar pase de 10.4 a 2.1 por ciento. El crecimiento para el primer decil es de 8.5% (en comparación con 29.9% al utilizar los datos de la ENIGH 2016). Cuando estos resultados se desagregan por entidad federativa, también se observan ajustes a la baja en las tasas de crecimiento para la mayoría de las entidades.

Referente a la medición de pobreza, una implicación de la disminución de los ingresos es el aumento en pobreza con respecto a los datos de la

ENIGH 2016. Con el ajuste propuesto, la población en situación de pobreza multidimensional es de 43.6% (para el 2014 fue de 46.2%). Cabe mencionar que esta cifra de pobreza incorpora un ajuste adicional a la carencia por seguridad social realizada por el CONEVAL.

Fuentes

- Bandourian, Ripsy; James McDonald and Robert Turley. "A Comparison of Parametric Models of Income Distribution Across Countries and Over Time Across Countries and Over Time", en: *Luxembourg Income Study Working Paper No. 305*. 2002.
- Bustos, Alfredo. "Estimation of the distribution of income from survey data, adjusting for compatibility with other sources", en: *Statistical Journal of the IAOS. Journal of the International Association for Official Statistics*. Vol. 31, no. 4, noviembre del 2015, pp. 565-577.
- García, Carmelo; Mercedes Prieto e Hipólito Simón. *La modelización paramétrica de las distribuciones salariales. Un estudio aplicado al caso español*. Universidad de Alicante, Departamento de Análisis Económico Aplicado, junio del 2012, 31 pp.
- Graf, Monique; Desislava Nedyalkova; Ralf Münnich; Jan Seger and Stefan Zins. *Parametric estimation of income distributions and indicators of poverty and social exclusion. Advanced Methodology for European Laeken Indicators*. Sevent Framework Programme, 2011, 74 pp.
- McDonald, James B. "Some generalized functions for the size distribution of income", en: *Modeling Income Distributions and Lorenz Curves*. Capítulo 3. Mayo de 1984, pp. 37-55.
- McDonald, James B. and Michael Ransom. "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality", en: *Modeling Income Distributions and Lorenz Curves*. Cap. 8. Economic Studies in Inequality Social Exclusion and Well-Being. Springer, 2008, pp. 147-164.
- McDonald, James B. and Yexiao J. X. "A generalization of the beta distribution with applications", en: *Journal of Econometrics*. Vol. 66, Issues 1-2, marzo-abril de 1995, pp. 133-152.
- Rigby, R. A. and D. M. Stasinopoulos. "Generalized additive models for location, scale and shape", en: *Appl. Statist.* 54, Part 3. 2005, pp. 507-554.
- Singh, S. K. and G. S. Maddala. "A function for size distribution or incomes", en: *Modeling Income Distributions and Lorenz Curves*. Vol. 44. Septiembre de 1976, pp. 963-970.
- Skinner, C.; D. Holt and T. Smith (editors). *Analysis of Complex Surveys*. New York, USA, John Wiley and Sons, 1989.
- Stasinopoulos, Mikis; Bob Rigby and Calliope Akantziliotou. *Instructions on how to use the gamlss package in R*. Second edition. Enero del 2008, 206 pp.

Anexo 1

En estudios previos, y para distintos países, se han probado familias de distribuciones para ajustar variables, como ingreso y salario, con resultados satisfactorios. Algunas de las funciones más utilizadas se muestran en la figura A.1 obtenida de Bandourian *et al.* (2002).

La siguiente parametrización de las formas funcionales elegidas para el estudio puede ser encontrada en el documento *Instructions on how to use the gamlss package in R*.

Beta Generalizada tipo 2: GB2 (μ, σ, v, τ)

$$f_Y(y|\mu, \sigma, v, \tau) = |\sigma|y^{\sigma v - 1} \{ \mu^{\sigma v} B(v, \tau) [1 + (\frac{y}{\mu})^\sigma]^{v + \tau} \}^{-1}$$

Para $y > 0$, donde $\mu, v, \tau > 0, -\infty < \sigma < \infty$ y $-v < \frac{1}{\sigma} < \tau$.

Dagum

$$Dagum(y|\mu, \sigma, v) = GB2(y|\mu, \sigma, v, \tau=1)$$

Gamma Generalizada: GG (μ, σ, v)

$$f_Y(y|\mu, \sigma, v) = \frac{|v|\theta^\theta z^\theta \exp\{-\theta z\}}{\Gamma(\theta)y}$$

Para $y > 0$, donde $\mu > 0, \sigma > 0$ y $-\infty < v < \infty$. Y donde $z = (\frac{y}{\mu})^v$ y $\theta = 1/(\sigma^2 v^2)$.

LogNormal distribución: LogNor (μ, σ)

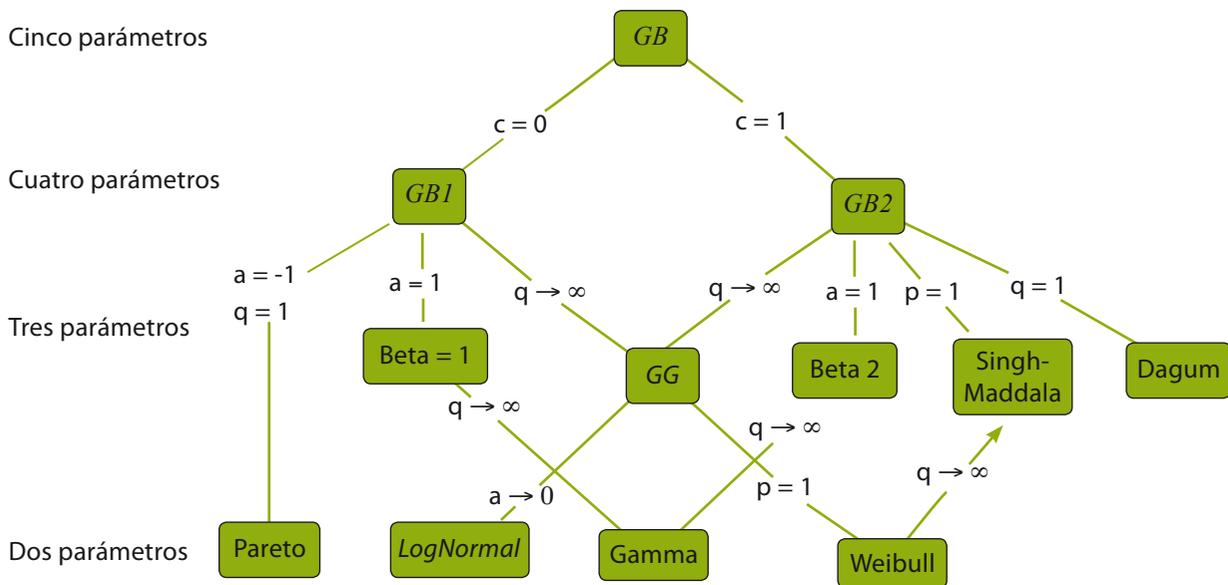
$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left\{-\frac{[\log(y) - \mu]^2}{2\sigma^2}\right\}$$

Para $y > 0$, donde $\mu > 0$ y $\sigma > 0$.

Aquí $E(Y) = \omega^{1/2} e^\mu$ y $Var(Y) = \omega(\omega - 1)e^{2\mu}$, donde $\omega = \exp(\sigma^2)$.

Figura A.1

Esquema de las funciones que se usan en la modelación del ingreso



Anexo 2

Para determinar la función paramétrica a utilizar, realizamos algunas pruebas de bondad de ajuste

Estadísticos para pruebas de bondad de ajuste

Definimos:

$$p_i(\hat{\theta}) = F(y_i|\hat{\theta}) - F(y_{i-1}|\hat{\theta})$$

$$N = \sum_{i=1}^n w_i$$

w_i = número de hogares que representa el ingreso y_i .

SSE suma del cuadrado de errores

$$SSE = \sum_{i=1}^n \left(\frac{w_i}{N} - p_i(\hat{\theta}) \right)^2$$

$$SSE_{acum} = \sum_{i=1}^n \left(\frac{\sum_{m=1}^i w_m}{N} - F(y_i|\hat{\theta}) \right)^2$$

Estadístico χ^2

$$\chi^2 = N \sum_{i=1}^n \left[\frac{\left(\frac{w_i}{N} - p_i(\hat{\theta}) \right)^2}{p_i(\hat{\theta})} \right]$$

Log-verosimilitud

$$l(\theta_k)$$

Aikaike

$$AIC(k) = 2k - 2l(\theta_k)$$

BIC

$$BIC(k) = \ln(n)k - 2l(\theta_k)$$

K = número de parámetros.

y elementos de visualización que nos permitieran elegir la función que mejor se ajusta a los datos empíricos. El resultado fue la selección de la función *GB2* para nuestro modelo.

Visualización

Función de densidad

En primera instancia se puede comprobar si la función elegida se adecua al histograma de los datos observados en la encuesta (ver grupo 2 de gráficas).

Función de distribución acumulada

En el grupo 1 se muestran las probabilidades acumuladas que asigna la función teórica a los ingresos observados $F(Y|\hat{\theta}) = P(Y \leq y); y > 0$.

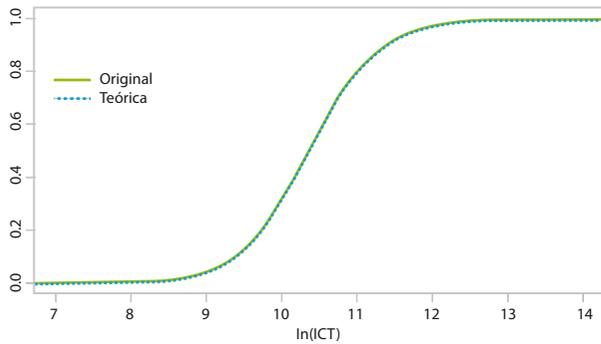
pp-plot

En el grupo de gráficas 3 refleja la discrepancia para cada observación de la muestra con respecto a los valores que la función del modelo teórico pronostica; cuanto más se aproximan a la diagonal del cuadrante la calidad del ajuste es mejor.

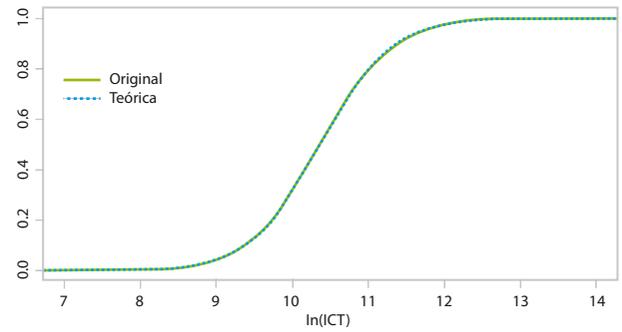
Grupo 1

Distribución acumulada, teórica (ajustada) vs. ENIGH 2016 (empírica)

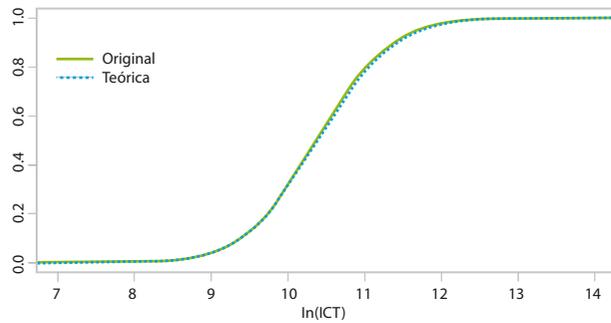
Ajuste *Dagum*: muestra expandida
Distribución acumulada, original vs. teórica



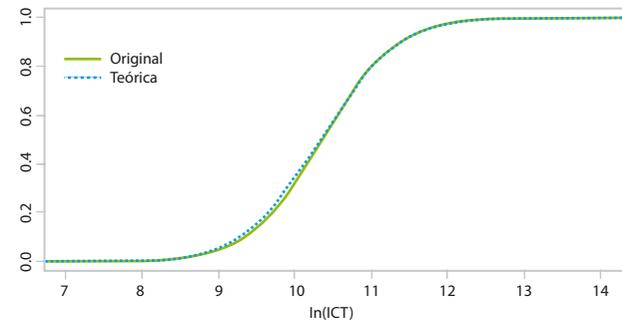
Ajuste *GB2*: muestra expandida
Distribución acumulada, original vs. teórica



Ajuste *GG*: muestra expandida
Distribución acumulada, original vs. teórica



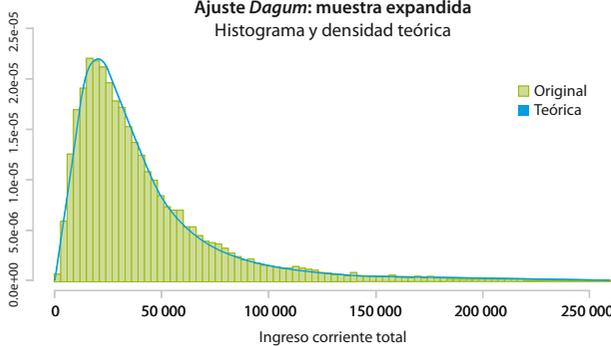
Ajuste *LogNor*: muestra expandida
Distribución acumulada, original vs. teórica



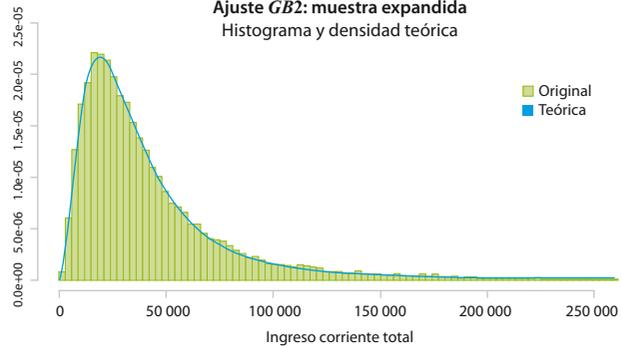
Grupo 2

Densidad, teórica (ajustada) vs. ENIGH 2016 (empírica)

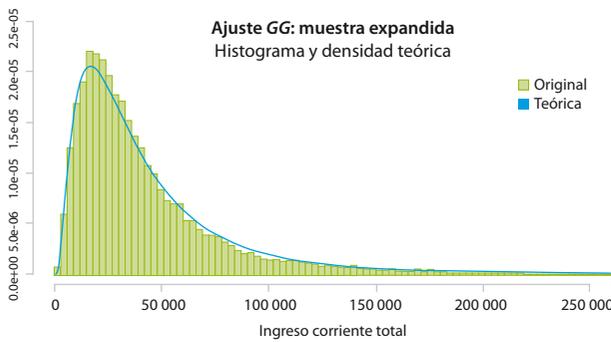
Ajuste *Dagum*: muestra expandida
Histograma y densidad teórica



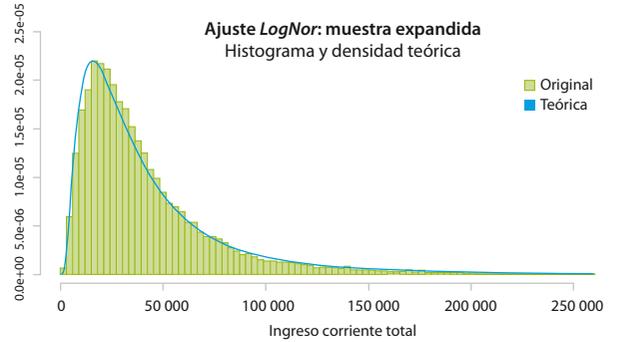
Ajuste *GB2*: muestra expandida
Histograma y densidad teórica



Ajuste *GG*: muestra expandida
Histograma y densidad teórica

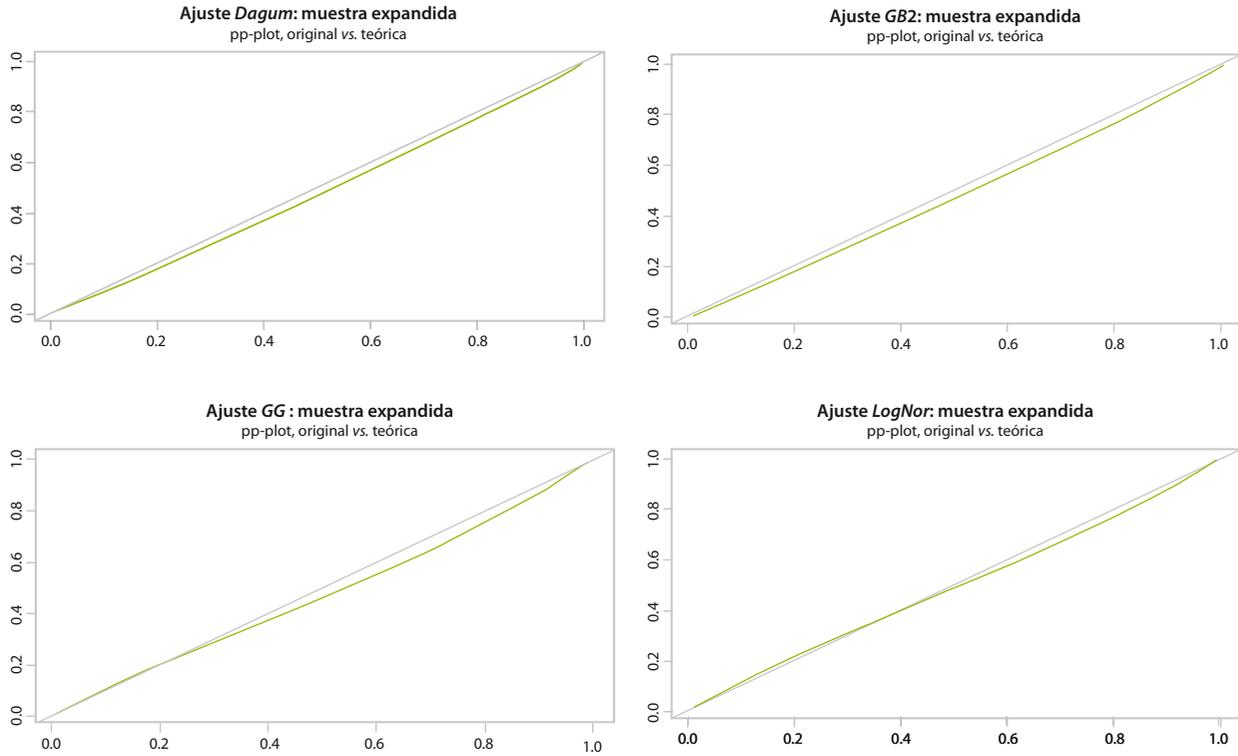


Ajuste *LogNor*: muestra expandida
Histograma y densidad teórica



Grupo 3

pp-plots, teórica (ajustada) vs. ENIGH 2016 (empírica)



Estadísticos

Suma cuadrática de errores

Decil	GB2	GG	LogNor	Dagum
I	7.54	199.33	366.29	23.35
II	2.65	581.95	1 974.89	62.28
III	6.05	291.24	2 365.92	98.51
IV	26.30	19.68	1 550.56	80.60
V	3.20	156.51	904.04	15.55
VI	0.40	731.64	219.12	63.93
VII	0.81	1 397.93	20.16	111.35
VIII	8.07	1 295.13	92.24	147.25
IX	8.41	546.10	103.47	52.57
X	4.18	48.22	15.85	22.66

Log-verosimilitud y BIC

	GB2	GG	LogNor	Dagum
Log-verosimilitud	387 168 080	387 702 106	387 956 925	387 212 144
BIC	774 336 167	775 404 219	775 913 854	774 423 900

Promedio del ICT por decil

Decil	Empíricos	<i>GB2</i>	<i>GG</i>	<i>LogNor</i>	<i>Dagum</i>
I	8 166.343	8 088.938	7 658.307	7 611.810	8 090.211
II	14 206.059	14 229.802	13 548.606	13 090.490	14 424.132
III	18 918.086	18 977.140	18 470.364	17 684.250	19 166.112
IV	23 555.613	23 690.323	23 552.246	22 467.080	23 780.692
V	28 812.506	28 827.502	29 216.253	27 851.860	28 753.483
VI	34 836.654	34 845.130	35 912.906	34 294.500	34 555.896
VII	42 431.256	42 464.582	44 358.019	42 536.250	41 932.233
VIII	53 383.482	53 215.095	56 017.062	54 122.790	52 488.715
IX	72 040.802	71 708.933	75 008.577	73 490.730	71 238.109
X	168 855.531	153 052.109	136 041.401	140 650.320	169 229.232

Cuantiles

Cuantiles	Empíricos	<i>GB2</i>	<i>GG</i>	<i>LogNor</i>	<i>Dagum</i>
1	4 776.380	4 396.711	4 341.613	4 516.718	4 208.322
2	6 017.870	5 798.776	5 556.088	5 658.439	5 684.630
5	8 524.280	8 511.483	7 995.246	7 932.096	8 541.460
10	11 587.230	11 661.153	10 984.335	10 707.107	11 811.650
20	16 616.400	16 657.933	16 025.386	15 396.725	16 868.953
30	21 184.660	21 296.024	20 945.848	20 006.730	21 448.162
40	26 072.200	26 153.967	26 253.846	25 025.791	26 171.796
50	31 657.340	31 643.406	32 346.301	30 849.512	31 468.989
60	38 285.170	38 302.901	39 760.637	38 028.336	37 893.551
70	47 059.050	47 112.358	49 459.267	47 567.238	46 462.043
80	60 687.960	60 424.185	63 645.591	61 808.619	59 673.107
90	87 032.500	86 942.300	89 793.806	88 886.264	87 158.768
95	120 797.390	120 120.403	118 730.288	119 978.997	123 730.982
98	175 332.940	178 492.345	161 774.033	168 167.788	193 120.251
99	226 429.530	237 723.877	198 248.318	210 624.507	268 848.551

Incentivos, error de medición y estimación de la pobreza en México

Incentives, measurement error, and poverty estimation in Mexico

Alfonso Miranda*, a,c,d y **Jaime Sainz Santamaría****, b,c

^a División de Economía, Centro de Investigación y Docencia Económicas (CIDE); ^b División de Administración Pública, CIDE; ^c Programa de Estudios Longitudinales, Experimentos y Encuestas (PANEL); ^d Instituto para el Estudio del Trabajo (IZA), 53113 Bonn, Alemania.

* alfonso.miranda@cide.edu

** jaime.sainz@cide.edu

Nota: agradecemos a Cristina Álvarez Venzor por su excelente apoyo como asistente de investigación en la elaboración del presente documento; aquí se presenta la opinión de los autores y no necesariamente la del CIDE o del GTA; cualquier error u omisión es responsabilidad exclusiva de los autores.



Rear View Of Boy Looking At Fountain/Petro Salvador/EyeEm/Getty Images

En el 2015, el Instituto Nacional de Estadística y Geografía (INEGI) modificó la estrategia de recolección de datos del Módulo de Condiciones Socioeconómicas (MCS), la encuesta con la que se estiman los cambios de pobreza en México. La publicación de los resultados del MCS 2015 generó un intenso debate público debido a la discrepancia sustantiva en las tasas de pobreza 2014 y 2015, la cual no puede atribuirse exclusivamente a cambios verdaderos de la pobreza.

El presente artículo argumenta que los cambios en el trabajo de campo 2015 modificaron los incentivos de los encuestadores durante la recolección de datos, con lo que incrementaron los esfuerzos en campo para verificar la validez y consistencia de los datos sobre ingreso en los hogares. Con ello, posiblemente mejoró la estimación del ingreso y la pobreza en México. No obstante, debido a que el INEGI no implementó un estudio —previo al despliegue de campo principal del MCS 2015— para estimar los posibles impactos del cambio metodológico, no existe evidencia empírica que muestre efectivamente que se logró una mejora en la medición del ingreso. Un resultado imprevisto es que la nueva estrategia rompió la comparabilidad de la serie MCS 2015 con las estadísticas basadas en versiones anteriores del Módulo.

Con el fin de dar continuidad a la serie, el INEGI elaboró un conjunto de ejercicios de ajuste del MCS 2015 por diversas vías, enfoques y métodos (*ejercicios sintéticos*). Aquí se analizan las ventajas y limitaciones de éstos y se propone como alternativa un enfoque experimental. Por último, se presentan algunas recomendaciones y reflexiones sobre cuál es el camino a seguir para el MCS y los cálculos de pobreza en México.

Palabras clave: medición de la pobreza; error de medición; sistema de incentivos en encuestas; análisis longitudinal.

In 2015 INEGI introduced changes to the fieldwork of the Socioeconomic Conditions Module (MCS by its Spanish acronym), which is the main survey that is used in Mexico to measure income and track changes in poverty. The publication of the results of the MCS 2015 sparked an intense public debate due to the stark discrepancy in the poverty rates 2014 and 2015, which cannot be exclusively attributed to true changes in poverty.

The present paper argues that INEGI's fieldwork changes in 2015 modified enumerators' incentives to purposely verify the validity and consistency of the household income information collected by the MCS. Possibly, the changes improved income measurement and poverty estimation. However, because INEGI did not carry out a study —before MCS 2015's main field deployment— to estimate the potential impact of the methodological changes, there is no empirical evidence to show that effectively an improvement on income measurement was achieved. An unintended outcome was that the new fieldwork strategy broke the comparability of the MCS 2015 with statistics based on previous versions of the MCS.

To recover the continuity of the series, INEGI performed a set of adjustment exercises for the MCS 2015 using different approaches and methods (“synthetic exercises”). We comment on the advantages and limitations of each synthetic exercise and, as an alternative, we put forward an experimental approach that in our opinion can satisfactorily deal with today's empirical and theoretical challenges for building income and poverty series that are comparable over time. Finally, we make the case for the use of longitudinal surveys for income and poverty measurement in Mexico.

Key words: poverty measurement; measurement error; incentive structure in surveys; longitudinal analysis.

1. Introducción

El Módulo de Condiciones Socioeconómicas (MCS) de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) es la fuente primaria de información que el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) utiliza para la medición de la pobreza multidimensional en México. En el 2015, el Instituto Nacional de Estadística y Geografía (INEGI) acordó con el CONEVAL que el levantamiento del MCS se realizara por primera vez como una encuesta independiente.

En julio del 2016, el INEGI dio a conocer los resultados del MCS 2015, y los niveles de ingreso que se registraron fueron mayores de lo esperado. De acuerdo con los datos publicados, el ingreso promedio de los hogares aumentó en 11.9% entre el 2014 y 2015. De forma importante, se detectó que el ingreso trimestral se incrementó considerablemente en cada decil de la distribución, que fue de 33% en el primer decil, 21.2% en el segundo, 17.7% en el tercero, 17% en el cuarto, 15.9% en el quinto, 15.5% en el sexto, 15% en el séptimo, 12.5% en el octavo, 10.1% en el noveno y 7.5% en el décimo.

Un análisis descriptivo revela que el tamaño promedio del hogar en el 2015 fue más pequeño de lo esperado por las proyecciones de población del Consejo Nacional de Población (CONAPO) y el ejercicio 2014 del MCS. Asimismo, el número de hogares aumentó 3.3%, muy por arriba de la proyección de 1.3% del CONAPO. En consecuencia, se sospecha que los hogares más pequeños están sobrerrepresentados. Además, entre el 2014 y 2015, el número de perceptores por hogar aumentó en 3.9% y el ingreso promedio por perceptor, en 7.7 por ciento. Claramente, el componente que más contribuye al incremento del ingreso es el aumento en el ingreso por perceptor.

El incremento del ingreso registrado en apenas un año resulta inverosímil y provoca cambios dramáticos en los cálculos de pobreza que elabora el CONEVAL. Desde un principio, el INEGI anunció que los cálculos de ingreso y pobreza basados en

el MCS 2015 no eran comparables con las estadísticas generadas con el MCS de años anteriores. La falta de comparabilidad entre las series dificulta la tarea de medición de pobreza del CONEVAL, que se usa en las fórmulas para estipular el monto de recursos que recibe cada estado y municipio de la República Mexicana, con base en la *Ley General de Desarrollo Social* (que define las *zonas de interés prioritarias* para el gobierno federal) y la *Ley de Coordinación Fiscal* (que utiliza la medición para distribuir el Fondo de Aportaciones para la Infraestructura Social), entre otras leyes. En consecuencia, los cambios en la distribución del ingreso entre el MCS 2014 y MCS 2015 —y la pérdida de comparabilidad de la serie de índices de pobreza— tienen potencialmente consecuencias muy serias en la vida de millones de mexicanos.

Ante la relevancia del problema, el INEGI y el CONEVAL acordaron que se elaborara un ejercicio *sintético* que permita restablecer la comparabilidad de los datos del MCS 2014 y MCS 2015 de forma que el CONEVAL esté en condiciones de calcular un índice de pobreza 2015 que mantenga la continuidad en la serie de los índices de pobreza.

De acuerdo con el INEGI, el diseño del MCS es el mismo que en años anteriores. Se mantuvo la misma metodología en el 2015 que en los levantamientos anteriores: a) mismo diseño conceptual; b) mismo diseño estadístico; c) mismo operativo de campo (con cambios sutiles en apariencia, pero cuyos posibles impactos se discuten en este documento); d) mismo procesamiento y generación de datos; e) control muy similar; f) capacitación muy similar; y g) validación muy similar. Sin embargo, el INEGI reconoce que sí cambiaron algunos aspectos con el fin de mejorar la medición del ingreso. El ejercicio 2015 puso énfasis en captar los ingresos en forma adecuada y se detectaron, a partir del ejercicio 2014, áreas de oportunidad en la capacitación del personal que impactaban el registro de la información en temas como los siguientes: a) los ingresos de los negocios del hogar, b) las justificaciones de los hogares sin ingreso, c) las descripciones de los buscadores de trabajo y de personas no económicamente activas, d) las descripciones de

la clave *otros ingresos*, e) las descripciones de *otras prestaciones* y e) el reporte del tipo de informante para los negocios del hogar.

El operativo de campo cambió, porque se agregó un nuevo apartado al *Manual del entrevistador*. En este artículo argumentamos que tal cambio potencialmente modificó la manera en que se conduce la entrevista. Entre otros aspectos, el apartado 10.2 instruye a los encuestadores como sigue:

- “Ten en mente la información previamente registrada en los diferentes cuestionarios, y si consideras que dicha información se contradice o requiere algún sondeo adicional, realiza el mismo y registra las observaciones necesarias que permitan avalar las diferentes situaciones”.
- “Cualquier información que te cause duda es mejor aclararla con el informante y realizar una observación al respecto”.
- “Cuando el informante te confirme que no tuvo ingreso alguno, realiza preguntas adicionales que te permitan identificar cómo se sostiene o cómo le hace para vivir, y en caso de que sea necesario, corrige la información previamente registrada o justifica la situación en las observaciones al final del cuestionario. De lo contrario, procederá un retorno a campo”.
- “Para los integrantes del hogar no ocupados, verifica que se cuente con información en alguno de los apartados de la Sección VII de dicho cuestionario; de no ser así, verifica esta situación con el informante y realiza las correcciones pertinentes”.

Por otro lado, se definió el criterio C14, que establece condiciones bajo las cuales se debe volver a campo para verificar casos que generaran dudas sobre su veracidad. Su implementación involucró verificar en las oficinas centrales del INEGI la justificación de los entrevistadores para los hogares sin ingreso o con ingresos bajos. En ese ejercicio se evaluó, con base en la información del hogar, si ameritaba un retorno a campo. El criterio C14 se aplicó a 647 folios (0.98% de la muestra), de los cuales 215 (33.2%) derivaron en cambios de ingreso. De acuerdo con el INEGI, en ningún momento

el personal operativo tuvo conocimiento del umbral para detonar el criterio.

Este artículo tiene una serie de notas conceptuales que pueden ser útiles para la discusión acerca de la medición de la pobreza en México, específicamente sobre la importancia del sistema de incentivos que enfrentan los encuestadores en los operativos de campo del INEGI y sobre la relevancia de expandir la estrategia de medición usando un enfoque longitudinal. La sección 2 contiene un diagnóstico de cuál es el posible origen de las diferencias entre el MCS 2014 y MCS 2015. Enseguida, la 3 muestra comentarios acerca de algunos de los ejercicios sintéticos que hasta el día de hoy ha realizado el INEGI. En la 4 se presentan algunos puntos e ideas de los cálculos de pobreza en México y reflexiona sobre cuál es el camino hacia adelante en el MCS y los cálculos de pobreza en México. Por último, se dan las conclusiones.

2. ¿Cuál es el origen de las diferencias entre el MCS 2014 y 2015?

En esta sección se presenta un modelo conceptual que muestra cuáles podrían ser las causas de las diferencias observadas entre los levantamientos 2014 y 2015 del Módulo.

Considere como punto de partida la ecuación de ingreso de Mincer (1958) para el 2014 y 2015:

$$y_{14} = \mathbf{x}_{14}\beta_{14} + u_{14} \quad (1)$$

$$y_{15} = \mathbf{x}_{15}\beta_{15} + u_{15} \quad (2)$$

donde y representa el ingreso, \mathbf{x} es un vector $1 \times K$ de características socioeconómicas (incluyendo la constante) y u es un error aleatorio. El vector (y, \mathbf{x}) tiene una función de densidad conjunta $f(y, \mathbf{x})$, la cual potencialmente cambia en el tiempo. En otras palabras, $f(y_t, \mathbf{x}_t)$ define un proceso estocástico a través del tiempo con $t = \{2014, 2015\}$. En general,

permitimos correlación del error con una o más variables explicativas $E(u|x) \neq 0$, de forma tal que el vector x es potencialmente endógeno. Suponga que el ingreso del 2014 fue medido por medio del MCS 2014 y el ingreso en el 2015, por el MCS 2015.

Para guiar la discusión, la gráfica 1 representa, de manera hipotética, la función de densidad conjunta del ingreso y las variables socioeconómicas x para los años 2014 y 2015. Note que (a) y (b) de esta gráfica no corresponden a los datos reales del MCS 2014 y 2015, son simplemente artefactos teóricos que nos ayudarán a discutir las aristas del problema que nos concierne. No se interprete, sin embargo, que las formas ilustradas de $f(y_{14}, x_{14})$ y $f(y_{15}, x_{15})$ son caprichosas. Por el contrario, el diseño tiene por objetivo resaltar características relevantes que exhiben los datos reales.

Sin pérdida de generalidad, se considera el caso donde el vector x contiene una sola variable x (además de la constante). En la gráfica se representan dos ejemplos posibles de la distribución marginal del ingreso: $f(y)$ en la *pared del fondo* de la gráfica y la distribución marginal de la variable control $f(x)$ en la *pared del lado izquierdo*. En el centro se representan ejemplos de la distribución condicional $f(y|x)$. La función de densidad conjunta $f(y, x)$ es la *montaña* de tres dimensiones que se forma con la unión de todas las funciones de den-

sidad condicional a lo largo del rango de la x . Note que la densidad conjunta $f(y, x)$ puede tener una forma muy irregular (gráfica 1.a). En consecuencia, como se muestra, la forma de la densidad condicional $f(y|x)$ puede ser muy diferente para distintos valores de x . Eso involucra no solo la media condicional $E(y|x)$, sino en general a todos los momentos condicionales $m(y|x)$.

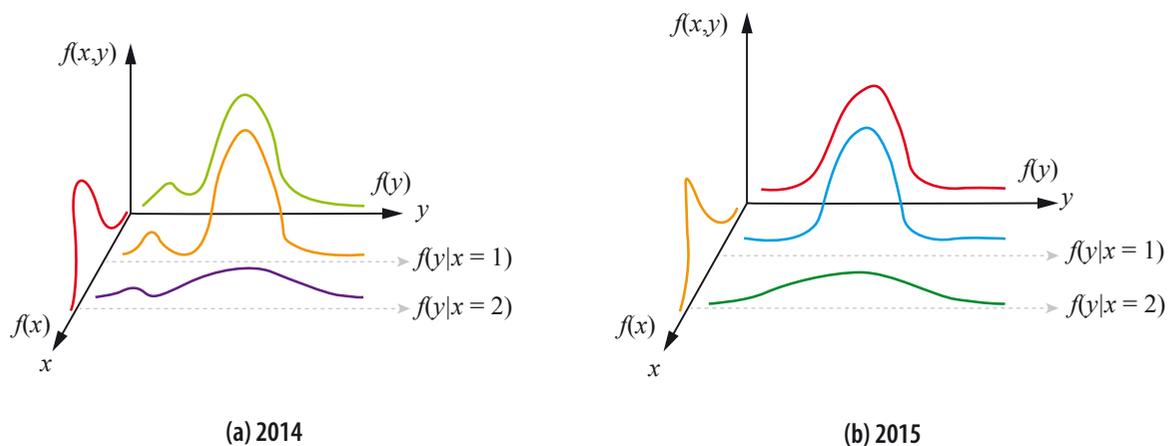
La pregunta central que nos ocupa es: ¿qué pudo cambiar entre el MCS 2014 y el MCS 2015?

En general, pudo cambiar:

- 1 La densidad (distribución) marginal de las variables socioeconómicas $f(x)$. Esto es, los hogares (personas) de México modificaron su perfil socioeconómico de forma importante entre el 2014 y 2015, promediando sobre la distribución de ingresos.
- 2 La densidad (distribución) marginal del ingreso $f(y)$. Esto es, los hogares (personas) de México recibieron un ingreso monetario significativamente diferente entre el 2014 y 2015, promediando sobre la distribución de sus características socioeconómicas.
- 3 La densidad (distribución) conjunta del ingreso y las características socioeconómicas de la población $f(y, x)$. Esto es, el ingreso que recibió el conjunto de personas definido por

Gráfica 1

Distribución del ingreso 2014 y 2015



un perfil sociodemográfico cambió de manera importante entre el 2014 y 2015.¹ Esto incluye:

- a) Un cambio en los retornos monetarios a las características socioeconómicas en el mercado de trabajo. Esto es, un desplazamiento de la media condicional $E(y|x) = x\beta$ debido a un cambio en los coeficientes β .
- b) Un cambio en el perfil sociodemográfico de la población (ver numeral 1) $f(x)$.
- c) Un cambio en la densidad (distribución) marginal del error $f(u)$.
- d) Un cambio en la densidad (distribución) condicional del error dadas las características sociodemográficas $f(u|x)$.

Es difícil —sino imposible— pensar que el perfil sociodemográfico de los hogares en México cambió de forma importante entre el MCS 2014 y MCS 2015, ya que solo hay un año de diferencia entre los dos levantamientos. En un año, simplemente, no cambia la estructura de la población de manera importante y $f(x)$ debería permanecer fija (a menos que el diseño de la muestra 2014 ó 2015 esté mal). Un argumento similar descarta que, en un periodo de 12 meses, existan cambios importantes en el retorno a las características demográficas de la población en el mercado de trabajo. En otras palabras, esperamos $\beta_{14} = \beta_{15}$. Entonces, si hay cambios en $f(y, x)$ entre el MCS 2014 y MCS 2015, éstos seguro se deben a modificaciones en la distribución (densidad) condicional del error dadas las características socioeconómicas $f(u|x)$.

Para aclarar el punto, la gráfica 2 superpone las dos densidades conjuntas $f(u_{14}, x_{14})$ y $f(u_{15}, x_{15})$, indicando en verde (líneas diagonales) la diferencia entre ambas gráficas. Aquí se dibuja la distribución conjunta $f(u, x)$ tomando en cuenta que, si fijamos los coeficientes β , la distribución de y está enteramente determinada por x y u . Note que los cambios entre el 2014 y 2015 están centrados en la parte izquierda de la distribución. En particular, en el 2014 (y en versiones previas del MCS) se ob-

serva una *joroba* en la cola izquierda que se forma porque una proporción importante de la población se niega a declarar ingresos, declara ingresos cero o declara ingresos muy pequeños, en un intento por *ocultar* información sensible o confidencial. Este comportamiento genera un error de medición que es particularmente acentuado en la parte izquierda de la distribución. Cabe aclarar que ésta no es una característica especial de los datos mexicanos, sino algo que se observa en la distribución del ingreso de todos los países del mundo. Existe una literatura amplia en el tema (ver, por ejemplo, Duncan y Hill, 1985; Rodgers y otros, 1993; Bound y Krueger, 1991; Bound y otros, 1994 y 2001; Hausman, 2001; Hyslop y Imbens, 2001; Biemer y otros, 2014; Kasprzyk, 2005; Gottschalk y Huynh, 2010; Nicoletti y otros, 2011; ChangHwan y Tamborini, 2014). Sabemos, también, que el deseo que las personas sienten por ocultar su ingreso es función del ingreso mismo y no solo de las características sociodemográficas. Así es, precisamente, cuando una persona tiene ingresos muy grandes o muy pequeños que tendrá mayores incentivos para ocultar o reportar de manera errónea su ingreso al encuestador del MCS (Duncan y Hill, 1985; Rodgers y otros, 1993; Bound y otros, 1994; Biemer y otros, 2014). Luego entonces, podemos concluir que el error de medición del ingreso es más grande en las colas de la distribución y que es función de las características socioeconómicas del hogar (persona), incluyendo el ingreso mismo.

Un ingrediente que complica aún más las cosas es que, potencialmente, los hogares (personas) que se niegan a declarar ingreso tienen características observables e inobservables de forma sistemática diferentes a los hogares que sí lo declaran. De hecho, con gran probabilidad, un dato faltante de ingreso no se genera como un acontecimiento casual, determinado completamente por el azar.

Rubin (1976) y Little y Rubin (2002) nos explican que los datos se pierden al azar, o son MAR (por sus siglas en inglés), cuando la probabilidad de tener un dato faltante depende de variables observadas, pero no de inobservadas. Por el contrario, cuando la probabilidad de perder un dato depende de va-

¹ Por ejemplo, los hombres de 35 años de edad con estudios de secundaria que son empleados en la industria de la construcción de la Ciudad de México.

riables inobservadas, se dice que los datos no se pierden al azar (o que son NMAR). Y si los datos que faltan son NMAR, la literatura muestra que surge un problema de sesgo de selección cuando el análisis se basa en la muestra observada y métodos convencionales de estimación (ver, por ejemplo, Heckman, 1979; Rubin, 1976; Little y Rubin, 2002). En el caso que nos ocupa, tenemos muy buenas razones para sospechar que los hogares más pobres y más ricos tienen mayor probabilidad de no declarar de manera correcta su ingreso (unos para evitar el estigma o por la dificultad de cuantificarlos, otros por razones de seguridad). Luego, seguramente, hay problemas de selección en la cola izquierda de la distribución del ingreso en el MCS debido a datos faltantes NMAR. El exceso de ingresos muy pequeños en el MCS se debe a un fenómeno muy similar al de datos faltantes. En la literatura esto se conoce como *clumping* y es causado por un reporte erróneo (*missreporting*) del ingreso, en especial entre los autoempleados (Jenkins, 1996; Parmeter, 2008; Hurst y otros, 2014).

Volvamos a nuestra pregunta: ¿qué cambió entre el 2014 y 2015? El INEGI ha explicado que, a diferencia de ejercicios anteriores, el MCS 2015 introdujo cambios en el operativo de campo con el objetivo de captar los ingresos de forma adecuada. En particular, se puso énfasis en validar con mucho más cuidado que en años anteriores la consistencia de la información de ingreso. El apartado 10.2

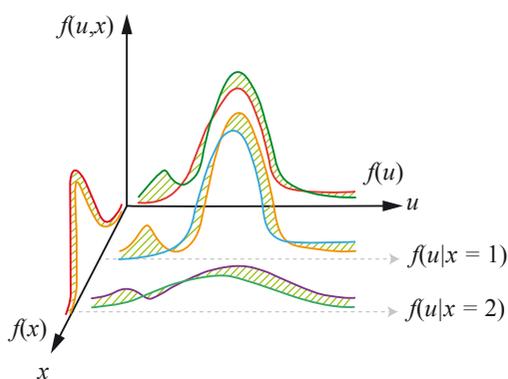
del *Manual del entrevistador* lo instruye a ser crítico e inquisitivo al momento de solicitar la información sobre ingreso. Ante cualquier duda sobre la consistencia de los datos, se pide al encuestador realizar preguntas adicionales para aclarar cómo el entrevistado se sostiene o qué hace para vivir. Si en función de las preguntas adicionales es necesario realizar modificaciones, debe corregir toda la información previamente registrada o justificar la situación en las observaciones al final del cuestionario. Más aún, ese apartado dice que si el entrevistador no justifica la situación en las observaciones de manera suficiente habría la posibilidad de un “retorno a campo”.

El criterio C14 estableció el umbral que se usó en la práctica para decidir qué folios requerían volver a campo para verificar la consistencia de la información sobre el ingreso. Además de dicho criterio, de acuerdo con el INEGI, se revisó en sus oficinas centrales la justificación de los entrevistadores para los hogares sin ingreso o con bajos ingresos y evaluó si ameritaba un retorno a campo. En la práctica, el retorno a campo se aplicó únicamente a 647 folios (0.98% de la muestra). Del total de folios verificados, solo en 215 casos (33%) se modificó el ingreso del hogar. De acuerdo con INEGI, en ningún momento el personal operativo en campo tuvo conocimiento del umbral para detonar el criterio.

Claramente, los cambios implementados al *Manual del entrevistador* y al entrenamiento tienen por efecto un cambio sustancial al operativo de campo del MCS 2015 respecto a lo que se hacía en el pasado. En años anteriores, el encuestador preguntaba ingreso y anotaba exactamente lo que la persona entrevistada le respondía. Si el informante se negaba a proporcionar datos de ingreso, el *campo* se dejaba sin llenar; si decía cero, se registraba un ingreso de cero. Si bien es cierto que en años anteriores el supervisor de campo verificaba la calidad y veracidad de la información, y que siempre había casos en los que se determinaba volver, en ejercicios anteriores el énfasis en el proceso de verificación de consistencia fue diferente, al menos en la capacitación del encuestador.

Gráfica 2

Comparación de las distribuciones 2014 y 2015



El MCS 2015 no solo introdujo un protocolo más estricto para verificar la consistencia y veracidad de la información del ingreso, sino que también dio, explícitamente, incentivos a los encuestadores para que hicieran bien su trabajo. En particular, se estableció en el *Manual* que el entrevistador tendría que “volver a campo” en caso de ser necesario.

Una de las lecciones más importantes de la teoría económica es que los incentivos importan. Esto es, modifican el comportamiento de los actores, y es posible que el apartado 10.2 modificara el *contrato* implícito entre el INEGI y el encuestador. En consecuencia, los incentivos y, probablemente, el comportamiento del entrevistador fueron diferentes para toda la muestra, no solo en los 647 casos que fueron finalmente sujetos a fiscalización. Para conocer la estructura de incentivos que enfrentó el encuestador, deben de conocerse —además de las indicaciones del *Manual del entrevistador* que se describen en este reporte— elementos clave del contrato implícito entre INEGI y quien recolectó los datos. En específico, es relevante saber si el gasto y esfuerzo que involucra regresar a campo son compartidos en alguna medida entre el INEGI y el encuestador: ¿el INEGI le paga por volver a campo, incluyendo viáticos?, ¿figura de alguna forma volver a campo en los indicadores de productividad del encuestador?

Los cambios en el operativo de campo del MCS 2015 (respecto a los ejercicios anteriores) afectaron de manera fundamental la distribución del error. El efecto se resintió en especial en la cola izquierda de la distribución, en los primeros deciles. La gráfica 2 ilustra las diferencias entre $f(u_{14})$ y $f(u_{15})$ en color amarillo. En la misma línea de investigación, la gráfica 3 muestra la diferencia del error $du = u_{15} - u_{14}$ a lo largo del rango de la x .

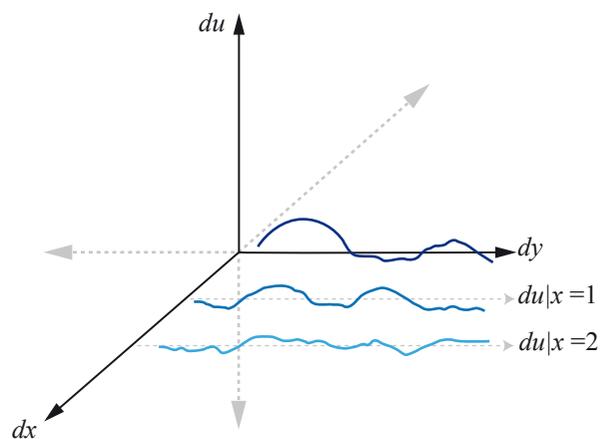
Cabe aclarar que el problema no sería grave si los cambios en el protocolo de campo del 2015 solo afectan la distribución (densidad) marginal del error $f(u)$, pero no la distribución condicional $f(u|x)$. Esto ocurriría, por ejemplo, si las personas con baja educación reportaran su ingreso de forma errónea —o en el caso en que simplemente se ne-

garan a reportar ingreso— con la misma frecuencia y en el mismo sentido y magnitud que aquellas con alta educación. Bajo esas condiciones, el cambio en el error entre el 2014 y 2015 no es función de las variables socioeconómicas y , por lo tanto, es cero cuando se promedia sobre la distribución de las x . En consecuencia, una estrategia que calibra la distribución de las x del 2015 para que sea igual a la distribución de las x del 2014 debería restablecer de manera satisfactoria la distribución del ingreso. El problema es que, por razones que ya expusimos antes, lo más probable es que los cambios en el protocolo de campo afectaran la distribución del error dadas las variables control $f(u|x)$. Eso hace que los cambios en el error de medición sean sistemáticos en diferentes áreas de la distribución de las x , es decir, que sean endógenos. En ese contexto, no es posible recuperar una distribución del ingreso en el 2015 que sea comparable con la del 2014, ya que el error es una variable inobservada.

Dado que el error u no es observable, ¿qué podemos hacer para recuperar una distribución del ingreso con base en el MCS 2015 que sea comparable con la distribución del ingreso del MCS 2014? Lo mejor sería implementar un experimento. Básicamente, la idea sería pedir a una muestra representativa de personas de la población que respondiera el cuestionario del MCS bajo las condiciones del protocolo de campo 2014 y bajo

Gráfica 3

Diferencia en la distribución del error 2014 y 2015



las condiciones del protocolo de campo 2015. El mismo conjunto de personas debe responder el cuestionario en ambas condiciones. Así sería posible inferir cuál es el cambio en el comportamiento que generan las diferencias en el protocolo de campo e investigar qué cambios se presentan en la distribución del error, dadas las variables socioeconómicas. Se deberán definir cuidadosamente el grupo control y el grupo de tratamiento. El experimento también debería de investigar en qué medida los cambios en el protocolo de campo afectan el comportamiento del entrevistador. Para ello, se deberán diseñar varios protocolos y definir de manera adecuada el grupo control y los grupos de tratamiento.

Existen otras alternativas que pueden lograr un ajuste, por naturaleza imperfectos, de la distribución del ingreso 2015. Aquí la clave es tratar de usar la densidad conjunta del ingreso y las variables socioeconómicas $f(y, x)$ para generar un ejercicio sintético. Ajustar la distribución de las x no es suficiente, ya que fundamentalmente lo que cambió es la distribución del error dadas las x . De alguna forma se trata de *reestablecer* la distribución del error 2014 en la distribución del ingreso 2015. La clave es modelar, para grupos definidos por el vector de variables explicativas x , el ingreso en el 2015 dadas las x y el ingreso 2014. En otras palabras, requerimos modelar:

$$y_{g,15} | \mathbf{x}_g, y_{g,14}; g = 1, \dots, G \quad (3)$$

donde g representa el índice de grupo. Esta estrategia explotaría la densidad conjunta $f(y_{g,15}, y_{g,14}, \mathbf{x}_g)$, la cual contiene información sobre el error en el 2014. La información proviene de la probabilidad condicional de $y_{g,15}$ dado $(y_{g,14}, \mathbf{x}_g)$

$$P(y_{g,15} | \mathbf{x}_g, y_{g,14}); g = 1, \dots, G \quad (4)$$

Se puede implementar una estrategia de regresión, una de apareamiento (*matching*) en x_{14} y x_{15} con donación de y_{14} , o bien, una de múltiple imputación de las x_{15} en base x_{14} con donación de la y_{14} .

El truco es usar la distribución conjunta de $(y_{g,15}, y_{g,14}, \mathbf{x}_g)$ para restablecer la distribución del error en el 2015. Las estrategias sugeridas pueden ser empleadas para construir ponderadores de la muestra del 2015 que una vez aplicados a los datos —usando ponderación inversamente proporcional a la probabilidad (IPW)— recuperen una distribución del ingreso 2015 que es comparable con la distribución del ingreso 2014. Por supuesto, cualquiera de estos ejercicios serían imperfectos. La desventaja principal es que si se quiere medir los cambios en la pobreza entre el 2014 y 2015, el ejercicio sintético de alguna forma eliminará cambios reales en la distribución del ingreso (y el error) entre ambos años. Como mencionamos antes, es prácticamente imposible recuperar una distribución del ingreso 2015 que sea comparable con la del 2014 y que refleje cambios reales en la distribución del error no asociados a las modificaciones del protocolo de campo.

Una alternativa a la construcción de ponderadores para los datos del 2015 es crear, con base en las estrategias ya delineadas, una o varias bases de datos sintéticas que sustituyan los datos del 2015. Aquí la idea es muestrear repetidamente de la distribución conjunta de $(y_{g,15}, y_{g,14}, \mathbf{x}_g)$.

3. Comentarios a los ejercicios de ajuste (sintéticos) presentados por el INEGI

3.1. Ajuste demográfico por posestratificación

En el MCS 2015, el tamaño promedio del hogar es más pequeño de lo esperado de acuerdo con las proyecciones de población del CONAPO y el ejercicio 2014 del MCS. En consecuencia, se sospecha que los hogares más pequeños están sobrerrepresentados en la muestra. Además, el análisis descriptivo de la muestra 2015 indica que el número de perceptores por hogar y el ingreso promedio por perceptor aumentaron de forma importante. El componente que más contribuye al incremento total del ingreso es el aumento en el ingreso por perceptor.

La posestratificación corrige la sobrerrepresentación de los hogares pequeños haciendo uso de la Encuesta Intercensal 2015 para ajustar los ponderadores. La idea es que, una vez ponderada la muestra, se recupere una distribución de las variables socioeconómicas que sea comparable con los ejercicios anteriores.

Después de la posestratificación, el tamaño promedio de los hogares coincide con el del levantamiento del 2014. Antes de la ponderación, el MCS encontraba un crecimiento de 3.3% en el número de hogares entre el 2014 y 2015. Dicho crecimiento estaba en discordancia con las proyecciones de población del CONAPO. Después de la ponderación se tiene un crecimiento de solo 1.3%, muy cerca de la proyección del CONAPO de 1.29 por ciento.

El ejercicio de posestratificación logra reducir el ingreso corriente total (ICT) entre el 2014 y 2015. Resultados similares se reportan para el ingreso promedio por perceptor (IPP) y el número de perceptores por hogar (PH). Sin embargo, la corrección es pequeña.

3.1.1 Comentarios

Este sintético está basado en corregir la distribución de las variables socioeconómicas x con el fin de que $f(x_{15})$ sea lo más similar posible a $f(x_{14})$.

Dado que el objetivo principal del ejercicio es reparar la distribución de las variables socioeconómicas en el 2015 $f(x_{15})$, y solo de forma indirecta el ingreso, es de esperar que el ejercicio tenga un impacto menor en la distribución marginal del ingreso. En consecuencia, el ejercicio hace poco por restablecer la comparabilidad entre el MCS 2014 y el MCS 2015.

Con el fin de evitar complicaciones en el uso de ponderadores, en la elaboración de la posestratificación se decidió construir un solo ponderador para toda la muestra. Desde el punto de vista de los autores, hay una ventana de oportunidad para mejorar la posestratificación elaborando un

ponderador a nivel individual y uno a nivel hogar.² Muchas encuestas en otras partes del mundo proveen ponderadores diferentes para hogares y para personas. Los usuarios pueden hacer uso correcto de los diferentes ponderadores si la base de datos contiene suficiente detalle sobre cómo emplearlos.

3.2. Ajuste demográfico por imputación

En la ENIGH, los hogares son visitados por siete días consecutivos, mientras que en el MCS 2015 la visita es de solo tres. En este contexto, la caída inesperada en el tamaño del hogar se puede deber a la omisión de personas, ya que en campo algunas veces el informante olvida reportar miembros del hogar en la primera visita. En la ENIGH, esos errores se corrigen en visitas posteriores, pero en el MCS 2015 hay menos tiempo para que el informante los corrija. La falta de muestra ENIGH exacerba el problema.

El ejercicio de imputación tiene como objetivo ajustar los factores de expansión del MCS 2015, como medio para restablecer la distribución correcta del tamaño del hogar en la muestra. Dicho objetivo se logra imputando los miembros del hogar perdidos en la muestra, para así dar menos peso a los hogares pequeños y más peso a los grandes.

Se propone corregir las características de los hogares del MCS por vía de un método de imputación *Cold-Deck*, que tiene por objetivo restituir a los miembros del hogar faltantes en el 2015; esta técnica define dos conjuntos de hogares: donadores y receptores. Los receptores son aquellos hogares donde se supone que desapareció uno o más miembros en el MCS 2015. Los donadores son

2 Por ejemplo, el *Estudio Longitudinal de los Hogares en el Reino Unido* proporciona un conjunto de ponderadores para hacer inferencia a nivel hogar y otro para hacer inferencia a nivel individuo (ver <https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/mainstage-waves-1-7-user-guide.pdf>). De forma similar, el *Estudio Panel de la Dinámica del Ingreso de los EE.UU.* proporciona un conjunto de ponderadores para hacer análisis de corte transversal y otro conjunto de ponderadores para hacer análisis longitudinal (ver https://psidonline.isr.umich.edu/data/weights/long_weight_15.pdf).

aquellos que se parecen a los hogares receptores en todas sus características, excepto en el número de miembros. Esto es, se busca que la distribución de las variables socioeconómicas $f(x)$ de donadores y receptores sea igual. Una vez apareados hogares donadores y receptores, se procede a transferir/imputar uno o más miembros del hogar donador al receptor. La técnica *Hot-Deck* usa la misma encuesta para buscar hogares donadores. Por el contrario, la *Cold-Deck* utiliza otras encuestas para buscar donadores. El proceso de imputación se realiza entidad por entidad. Para hacer la imputación, se usaron varias bases de datos: a) los MCS 2012, 2013, 2014 y 2015; b) la Encuesta Intercensal 2015; c) la Encuesta Nacional de Ocupación y Empleo (ENOE), tercer trimestre del 2015; y d) las proyecciones de los hogares en México y por entidades federativas 2010-2030.

Los resultados del ejercicio logran el objetivo de incrementar el tamaño del hogar de acuerdo con las tasas de variación implícitas calculadas a partir de las proyecciones de hogares y población del CONAPO. Los efectos del ejercicio sobre el ingreso y la pobreza son marginales. Se logra disminuir muy poco el monto del ingreso corriente y aumentar, también muy poco, al promedio de ingreso por hogar (debido a la disminución de los hogares) y la pobreza.

3.2.1 Comentarios

Este sintético está basado en corregir la distribución de las variables socioeconómicas x vía un ejercicio de imputación con el fin de que $f(x_{15})$ sea lo más similar posible a $f(x_{14})$.

En el ejercicio de imputación, el hogar donador es apareado al hogar receptor en función de características socioeconómicas x , pero no en función del ingreso. Éste queda fuera del vector de características de apareamiento.

Si el objetivo del ejercicio es corregir $f(x)$, y solo indirectamente $f(y)$, la corrección que se observa en la distribución del ingreso es pequeña y hace

poco por restablecer la comparabilidad entre el MCS 2014 y el MCS 2015. Sin embargo, si los objetivos del ejercicio se extienden a lograr una buena corrección de $f(y)$, entonces es posible mejorar el ejercicio explotando la distribución conjunta del ingreso y las características socioeconómicas $f(y, x)$. Esto puede ayudar a *restituir* las características de la distribución del error 2014 en el 2015, dado los socioeconómicos. La discusión de la sección 2 abunda en este último punto. Para lograr dicho objetivo, es posible incluir el ingreso en el vector de características que se usan para implementar la imputación.

3.3. Propuesta para recuperar la comparabilidad del MCS 2015 usando máquinas de soporte vectorial

La técnica está basada en el uso de dos conjuntos de datos diferentes: 1) una base de datos de entrenamiento y 2) una base de datos de predicción. La primera contiene una serie de casos, en este caso hogares, con características x y clasificadas *correctamente* por deciles de ingreso, i.e. de acuerdo con la distribución $f(y, x)$ del MCS de años anteriores. La de predicción son los datos del MCS 2015. Se busca clasificar a los hogares del MCS 2015 en deciles de ingreso con base en sus características socioeconómicas x y la experiencia de clasificación que la máquina de soporte vectorial (SVM, por sus siglas en inglés) obtiene de la base de datos de entrenamiento. Ésta incluyó los deciles (que tienen el rol de clases o categorías) y las covariables a nivel hogar de los MCS 2010, 2012 y 2014. SVM resuelve un problema de optimización que busca hiperplanos —que son función de las características socioeconómicas x — que identifican observaciones que pertenecen a una misma categoría, en este caso un decil. El problema de optimización minimiza el error de clasificación y asigna cada hogar del MCS 2015 a un decil pronosticado. Note que este procedimiento remueve, en un sentido limitado, el supuesto de que $f(u|x) = f(u)$.

Con base en la clasificación pronosticada, se hace un ajuste en los factores de expansión para

otorgarle mayor o menor peso a cada hogar, de acuerdo con el decil al que haya sido asignado. De esta forma, hogares con ingresos bajos tendrán un mayor peso.

El ajuste por *Machine Learning* logra que la variación del ICT entre el 2014 y 2015 a nivel nacional pase de 15.6 a 4.5 por ciento. En el primer decil, el ICT pasa de 38 a 11.2%, mientras que en el X pasa de 11.1 a 4.1 por ciento. En términos de pobreza, la población en pobreza pasa de ser 38.1 a 42.5%; la pobreza extrema, de 6.3 a 8.2%; y la moderada, de 31.8 a 34.3 por ciento.

3.3.1 Comentarios

El ejercicio de Machine Learning es el más prometedor. ¿Por qué SVM logra mejores resultados? Básicamente, porque se usa la distribución conjunta de las características socioeconómicas x y el ingreso y , $f(y, x)$ para entrenar a la máquina de clasificación. En la sección 2 se recomienda, precisamente, hacer uso de la distribución conjunta para mejorar la predicción.

Abundemos un poco. La base de datos de entrenamiento tiene información de hogares con una serie de demográficos x que están clasificados en deciles de ingreso. Claramente, aquí hay información de la distribución conjunta $f(y, x)$. De hecho, la base de datos de entrenamiento contiene una *malla* de $10 \times K$ celdas, donde K es el número de celdas que genera la combinación de valores que puede tomar x ; esa malla es una versión gruesa de la distribución conjunta $f(y, x)$ —que no usa toda la información disponible sobre $f(y, x)$, ya que los datos están clasificados únicamente en 10 celdas a pesar de que y es continua—; cuando SVM usa esa malla para entrenarse, el procedimiento explota, de forma limitada, la distribución conjunta para predecir el decil al que un hogar del MCS 2015 pertenece con base en su vector de características x .

Machine Learning está en el camino correcto. El sintético se puede mejorar si en vez de usar una malla con 10 celdas se usan 50 ó 100.

3.4. Armonización de los ingresos del MCS 2015 usando la ENOE como contrafactual

En ejercicios anteriores al 2015, el ingreso laboral promedio por hogar en el sexto decil del MCS y la ENOE es muy similar. En el 2015, esa coincidencia histórica se rompió debido a los cambios en el operativo de campo del MCS.

La continuidad de la ENOE y su similitud con el MCS para los levantamientos previos al 2015 permiten realizar un ejercicio sintético para restablecer la tendencia histórica que se rompió con el MCS 2015. El ejercicio consiste en un ajuste a nivel agregado que desplaza hacia atrás la distribución marginal del ingreso $f(y)$ del MCS 2015 hasta lograr que el decil VI de la distribución marginal ajustada $f^*(y)$ coincida con el sexto decil de la ENOE 2015. Este método de ajuste, basado en el de método Máxima Verosimilitud, fue inicialmente sugerido por Bustos (2015). La desventaja del ejercicio sintético es que es necesario especificar una forma funcional para la distribución del ingreso. Después de considerar varias opciones, Ruiz y Romo (2017) eligieron una función de distribución Beta Generalizada (GB2). El procedimiento logra un ajuste *macro*, en el sentido de que modela la distribución del ingreso de los hogares a nivel agregado y no el ingreso de cada hogar de forma separada, para después agregar y estudiar las características de la distribución predicha. Nótese que aquí se modela la distribución marginal del ingreso $f(y)$ y se ignora la información extra que contiene la distribución conjunta $f(y, x)$.

Una vez que se tiene el ajuste macro de la distribución marginal $f^*(y)$, se procede a generar una base de microdatos consistente. En particular, para cada hogar h en la muestra del MCS se calcula la probabilidad de observar un ingreso menor o igual al reportado y_h bajo la distribución marginal del ingreso en el MCS 2015 sin ajustar $\hat{p}_h = F(y_h)$, donde $F(\bullet)$ es la función acumulada de probabilidad. Con \hat{p}_h a la mano, se procede a usar la función inversa $F^{-1}(\bullet)$ para calcular el ingreso del hogar bajo la distribución ajustada $\hat{y}_h^* = F^{-1}(\hat{p}_h)$.

En términos de resultados, el ejercicio logra que la variación entre el 2014 y 2015 del ICT pase de ser 15.6% a ser solo de 4.7 por ciento. Por otro lado, el ICT promedio por hogar, que presentaba una diferencia de 11.9% antes del ajuste, varía en 1.4% después del ejercicio sintético.

Los resultados agregados del ejercicio sintético son razonables, bajo el criterio de que parecen reconstruir la tendencia histórica. Sin embargo, cuando se analizan los deciles de ingreso, el sintético sigue obteniendo cifras insatisfactorias. Por ejemplo, el crecimiento del ICT entre el 2014 y 2015 en el decil I después del ajuste es aún de 13.4% (antes del ajuste era 38%). Y con excepción de los deciles IX y X, el crecimiento del ingreso entre el 2014 y 2015 es superior a 5% en todos los deciles. Claramente, un aumento del ingreso de los hogares en un solo año que resulta inverosímil dadas las condiciones macroeconómicas de México entre el 2014 y 2015.

3.4.1 Comentarios

Una versión modificada, pero basada en los mismos principios, de este ejercicio es lo que finalmente eligió el INEGI para construir el modelo estadístico 2015 —y 2016— para la compatibilidad del MCS-ENIGH. Por un lado, tiene la ventaja de que hace uso de la ENOE, la cual tiene registro del ingreso laboral, como fuente de información independiente. La ENOE es, probablemente, la segunda encuesta más importante levantada por el INEGI, donde se registra el ingreso laboral de los hogares y que tiene representatividad para los 32 estados de la República. Luego entonces, es intuitivo usar información de esta encuesta para ajustar el MCS, en especial porque en la ENOE 2015 no se observaron quiebres en la distribución del ingreso en el 2015 respecto a la tendencia histórica. Por otro lado, el sintético tiene la relativa ventaja de ser intuitivo y fácil de explicar. Sin embargo, desde nuestro punto de vista, tiene serias desventajas.

En primer lugar, el ejercicio no hace uso de toda la información disponible ya que, en lugar de ex-

plotar la riqueza de los microdatos del MCS y de la ENOE, se decide hacer un ajuste donde solo se modelan agregados macroeconómicos y se desestima la información que contiene la distribución conjunta $f(y, x)$. No se explota ningún aspecto de la distribución conjunta $f(y, x)$. Simplemente, se usa el sexto decil —o en el ejercicio final, la mediana del ingreso a nivel estatal reportado entre el segundo y el tercer trimestre de la encuesta— de la ENOE 2015 como *pivote* para desplazar y *anclar* la distribución marginal del ingreso $f(y)$ del MCS 2015. En consecuencia: 1) la aproximación no aprovecha información relevante y 2) no tiene una clara justificación teórica, más allá de apelar al uso de una regularidad en las series de ingreso: históricamente, el sexto decil de la ENOE coincide con el sexto decil del Módulo de Condiciones Socioeconómicas.³

Hay una serie de preguntas obligadas: ¿por qué la ENOE y el MCS coinciden justo en el sexto decil?; si las dos encuestas miden ingreso laboral en la misma población, ¿no deberían coincidir a lo largo de toda la distribución?; ¿hay una razón teórica que explique por qué el punto de contacto es el sexto decil? Estas preguntas son difíciles de responder. Aquí lanzamos una hipótesis: posiblemente, la mayoría de las personas en el sexto decil son trabajadores del sector formal que reciben sus salarios por nómina, no tienen negocios ni actividades empresariales y pagan impuestos vía el sistema de retención automática del impuesto sobre la renta (ISR). En otras palabras, es precisamente en el sexto decil que se concentran los hogares con ingresos fáciles de medir y que tienen muy pocos incentivos a mentir sobre sus ingresos. Es allí, en el sexto decil, donde el error de medición es mínimo. Conforme nos movemos hacia las colas de la distribución del ingreso, el error de medición crece. En los extremos, la distribución del ingreso en el MCS y la ENOE se separan y no es fácil decir

³ En el modelo estadístico para la compatibilidad —2015 y 2016— que finalmente publicó el INEGI se decidió no usar el sexto decil como ancla. En su lugar se usa la mediana del ingreso como ancla (en parte, en respuesta a los puntos débiles que aquí identificamos). Para agregar flexibilidad, se estimó el modelo para cada entidad federativa. Claramente, estimar el modelo entidad por entidad agrega flexibilidad al modelo. Sin embargo, dentro de cada entidad, se sigue usando un solo punto de anclaje. Por tanto, la ganancia en flexibilidad es pequeña.

desde un punto de vista teórico cuál de los dos instrumentos es mejor.

Otro problema con el sintético por ajuste del ingreso laboral es que el ejercicio *macro* usa un solo pivote (o punto de anclaje). Claramente, el método es poco flexible, y sería mucho mejor hacer uso de varios puntos de anclaje para permitir un mejor ajuste. En apariencia, la única razón que impide usar varios puntos de anclaje es más práctica que teórica: la ENOE y el MCS solo coinciden en el sexto decil. A la luz de este punto, la ventaja inicial del sintético —es intuitivo y fácil de explicar— disminuye, ya que la simpleza se logra a partir de introducir supuestos difíciles de justificar y con la desventaja de quitar flexibilidad al modelo para adaptarse a los datos.⁴

Por último, el sintético tiene el problema de ser un método que modela agregados *macro*, que al ajustar solo la distribución marginal del ingreso $f(y)$ —y no la distribución conjunta $f(y, x)$ — rompe de manera implícita la relación entre el ingreso y las variables socioeconómicas x . Después, sin hacer explícitos los supuestos que justifican la validez del ejercicio, usa la función inversa $F^{*-1}(\bullet)$ y crea una base de microdatos sintética. Esto, efectivamente, equivale a crear una distribución conjunta artificial $f^*(y, x)$ que no tiene relación, a menos que se introduzcan supuestos muy restrictivos, con la distribución conjunta original $f(y, x)$.⁵ En la práctica, eso quiere decir que si el investigador usa la base de datos original del MCS 2015 para hacer análisis multivariado encontrará un conjunto de relaciones y coeficientes entre y y x muy diferentes a que si usa la base de microdatos sintética. En conclusión, con el ejercicio sintético la relación entre el ingreso y las variables socioeconómicas ha sido artificialmente modificada. El usuario debiera ser advertido explícitamente de evitar hacer análisis multivariado con la base de datos sintética.

4 En el ejercicio estadístico final se usa la mediana del ingreso en lugar de la media en el sexto decil. El punto débil persiste: ¿por qué usar solo un punto de anclaje? ¿Por qué usar como anclaje la mediana y no cualquier otro percentil?

5 En particular, se requiere que: 1) los datos sean MAR, 2) que los errores de medición no sean función del ingreso, 3) que al agregar los errores de medición promedien cero y 4) que exista suficiente continuidad y suavidad en $f(y, x)$ para justificar el uso de la función inversa.

4. Comentarios sobre los cálculos de pobreza en México y el camino hacia adelante

La problemática expuesta en el presente documento sugiere que es necesario reflexionar sobre la metodología y los datos que se usan hoy en día para la medición de pobreza en México, en especial si se quiere medir cómo cambia la pobreza a través del tiempo. Dicha reflexión permitirá identificar limitaciones de la actual metodología y trazar una ruta para mejorar los cálculos de la pobreza en el futuro.

Sabemos que, en general, la información proveniente de encuestas tiende a subestimar el ingreso (Nicoletti y otros, 2011). Hyslop y Townsend (2016) encontraron que los ingresos reportados, a través de una encuesta, por trabajadores de Nueva Zelanda son entre 2 y 4% menores que los reportados por datos administrativos. Los problemas de medición de la pobreza son aún más agudos en los países en desarrollo, principalmente en el caso de la población en los primeros deciles. Haughton y Khandker (2009) identificaron numerosas razones por las que los cálculos tienden a subestimar el ingreso de la población: la gente tiene problemas en recordar su ingreso, hay reticencias a declarar todas las fuentes de ingreso —en especial algunas que pueden provenir de fuentes ilegales o socialmente punibles—, es difícil calcular el valor de enseres que cambian de valor en el tiempo (ganado, por ejemplo). Haughton y Khandker (2009) señalan que la alternativa (medir consumo) puede complementar la información, pero no está exenta de errores: la gente subreporta consumo de alcohol o no reporta lujos o bienes ilícitos. En suma, las dificultades se traducen en tasas de pobreza sobreestimadas (Haughton y Khandker, 2009), a pesar de los esfuerzos del analista para tomar en cuenta el hecho de que la pobreza tiene múltiples dimensiones y que se manifiestan en una falta de acceso al consumo de bienes y servicios elementales (v.g. el enfoque de pobreza multidimensional del CONEVAL).

El problema de sobreestimación es más grave cuando se usan encuestas de corte transversal

Tabla 1

Cambios netos en la pobreza medida con cortes transversales repetidas

	t	$t + 1$	Cambio
Por arriba de la línea de pobreza	0.56	0.65	0.09
Por abajo de la línea de pobreza	0.44	0.35	-0.09
Total	1.00	1.00	0.00

para estimar los cambios en los niveles de pobreza a través del tiempo (ver, por ejemplo, Gottschalk y Huynh, 2010; ChangHwan y Tamborini, 2014; Chesher y Schluter, 2002; Bound y otros, 1994; Bound y Krueger, 1991; Duncan y Hill, 1985). Veamos por qué haciendo uso de un ejemplo hipotético.

Al tomar una fotografía representativa de una población, v.g. la población del estado de Oaxaca, en un punto del tiempo t , una encuesta de corte transversal del estilo del Módulo de Condiciones Socioeconómicas permite determinar qué proporción de la población se encuentra por arriba y por abajo de la línea de pobreza en el punto t . Pensemos ahora en la dinámica temporal. Suponga, por ejemplo, que en $t + 1$ se repite el ejercicio y se levanta una segunda encuesta de corte transversal en Oaxaca. Considere que ambas encuestas son representativas de ese estado, pero que el conjunto de personas que entran a la muestra en el tiempo t es diferente del conjunto de personas que entran a la muestra en el tiempo $t + 1$, como sucede en las encuestas de corte transversal. En $t + 1$ podemos también medir la proporción de la población de Oaxaca que se encuentra por arriba y por abajo de la línea de pobreza. La tabla 1 presenta datos hipotéticos.

¿Podemos concluir a partir de la tabla 1 que disminuyó la pobreza en la población entre t y $t + 1$? Sí, pero eso no garantiza que las personas que eran pobres en t siguen siendo pobres en $t + 1$ ni que las que no eran pobres en t sigan sin ser pobres en $t + 1$. Todo lo que podemos saber son los flujos netos dentro y fuera de la pobreza. Para saber bien qué pasó entre t y $t + 1$, es necesario hacer uso de datos longitudinales (conocidos también como de panel) para calcular las tasas verdaderas de transición dentro y fuera de la pobreza (también conocidos como flujos brutos). A diferencia de un corte transversal, una encuesta longitudinal sigue al mismo conjunto de personas —que son una muestra representativa de la población— a través del tiempo.

Con la dimensión longitudinal se puede establecer con precisión si una determinada persona permanece o no bajo la línea de pobreza en los dos puntos en el tiempo. Con dicha información es posible calcular la matriz de transiciones.

Continuemos con nuestro caso hipotético. De la tabla 2 descubrimos que gran parte de las personas que caen en condición de pobreza no permanecen pobres por mucho tiempo. Ésos son los llamados pobres transicionales, compuestos por

Tabla 2

Probabilidad de transición dentro y fuera de la pobreza $t + 1$

	Pobre	No pobre	Total
Pobre	0.15	0.29	0.44
t No pobre	0.20	0.36	0.56
Total	0.35	0.65	1.00

personas que en el momento de la entrevista no estaban empleados o que aceptaron temporalmente salarios bajos, pero que rápidamente se movieron hacia arriba. En el ejemplo, los pobres estructurales representan 15% de la población. Éste es el conjunto de personas que están en condición de pobreza de manera permanente. De forma similar, 36% de la población está permanentemente fuera de la pobreza. Claramente, como el ejemplo ilustra, la dimensión longitudinal de los datos es sustantiva para entender de forma profunda el fenómeno de la pobreza.

Note que la probabilidad de transición fuera de la pobreza es diferente de la probabilidad de transición dentro de la pobreza. De hecho, una persona pobre en el tiempo t tiene una probabilidad de 0.29 de salir de la pobreza en $t + 1$, mientras que una que no es pobre en t enfrenta una probabilidad de caer en pobreza de solo 0.20. La diferencia en las tasas de transición dentro y fuera de la pobreza explican la disminución neta en 9% en la proporción de pobres en la población entre t y $t + 1$.

En este punto cabe preguntar: ¿qué parte de la pobreza transicional es explicada por cambios en la distribución del error de medición entre t y $t + 1$?, ¿qué parte es generada por accidentes aleatorios que verdaderamente afectan el ingreso? Por ejemplo, desempleo temporal, enfermedad, eventos catastróficos (desastres naturales) o cambios en las condiciones macroeconómicas.

Si el análisis se hace exclusivamente con base en datos repetidos de corte transversal es imposible dar respuesta, ya que el analista no tiene idea de cómo cambia el error de medición entre el tiempo t y el tiempo $t + 1$. Con datos de diseño longitudinal es posible estimar modelos para la probabilidad de transición dentro y fuera de la pobreza y, con ello, *limpiar* (controlar) la influencia de la variación de los errores de medición del ingreso en los cálculos de pobreza. Es posible, además, comparar las diferentes dinámicas de transición entre estados (en este caso, pobre vs. no pobre) y cuáles son los efectos de las políticas públicas para combatir la pobreza y la inequidad.

En conclusión: la dimensión longitudinal no solo promete mayor riqueza de análisis (teórica y empírica) sino también permite controlar de mejor manera la influencia de las variaciones en la distribución del error de medición del ingreso. Es, desde nuestro punto de vista, la ruta hacia el futuro en el cálculo de la pobreza en México.

5. Conclusiones

En este documento se exploran las razones que explican la diferencias en la distribución del ingreso en la versión 2014 (y anteriores) y la del 2015 del Módulo de Condiciones Socioeconómicas.

La literatura sobre distribución de ingreso y su medición en encuestas nos proporciona elementos para sugerir que la versión 2014 (y anteriores) subestimaba los ingresos, tanto en la parte derecha como en la izquierda de la distribución. Se sabe que existe un error de medición que es generado por una combinación de problemas de datos truncados, datos censurados y datos faltantes. La literatura indica que el problema es más grave en la parte izquierda de la distribución. Para complicar las cosas, la probabilidad de no reporte, truncamiento y/o censura es función de las características de los hogares y las personas, incluyendo su ingreso. En otras palabras, la selección a un estado de observación, truncamiento y censura es informativa.

Dado que el error de medición depende de variables inobservables, es muy probable que se tenga un problema de sesgo de selección, el cual se manifiesta tanto en el número de casos con ingresos no declarados como en la frecuencia de valores pequeños en la cola izquierda y los montos parciales declarados en la cola derecha de la distribución. Éste es un problema común en todos los países del mundo y en todas las encuestas que miden ingreso. Este error de medición estaba presente en las versiones del MCS 2014 y anteriores. A partir de aquí se realizaban las mediciones de pobreza en México y, fundamentalmente, las estimaciones de la variación en la pobreza.

El protocolo de campo cambió en el 2015 con una clara intención de mejorar. Posiblemente lo logró, pues es probable que se mida mejor con la nueva estructura de incentivos. Sin embargo, al no prever que la modificación de incentivos podría alterar las estrategias y el nivel de esfuerzo de los encuestadores al recabar información, el INEGI no implementó un estudio —previo al despliegue de campo principal del MCS 2015— para estimar los posibles impactos del cambio metodológico. En consecuencia, no existe evidencia empírica que muestre efectivamente que se logró una mejora en la medición del ingreso. Un resultado imprevisto es que la nueva estrategia rompió la comparabilidad de la serie MCS 2015 con las estadísticas basadas en versiones anteriores del MCS.

Hoy por hoy, no sabemos cómo se compara la distribución del error en el 2014 con la distribución del error en el 2015. Conocer esas distribuciones y sus diferencias son la clave para lograr el mejor ajuste posible de la base de datos del 2015 y hacerla comparable con la del 2014.

Entender cómo cambió la distribución del error de medición es una tarea complicada. Es muy probable que no existen los datos necesarios para resolver el problema. Básicamente, se requiere saber cómo una persona (la misma persona) cambia su comportamiento cuando se enfrenta al protocolo 2015 en relación con el protocolo 2014. No solo hay que investigar el comportamiento del entrevistado, también hace falta investigar si el entrevistador modificó su comportamiento con el cambio de protocolo, en especial porque es posible que se hayan modificado sus incentivos para verificar la veracidad de la información.

La solución elegida para reconstruir la serie histórica fue la elaboración de un conjunto de ejercicios sintéticos, entre los cuales se escogió el enfoque basado en el ajuste del ingreso laboral con la ENOE. Cada uno de los ejercicios (incluido el que fue seleccionado para dar continuidad a la serie) tiene algunas ventajas (las cifras son más verosímiles, al menos en el agregado), pero presentan también problemas teó-

ricos y empíricos que no pueden resolverse con los métodos propuestos.

En el presente documento se sugiere que la mejor forma de proceder es diseñar un experimento que permita esclarecer todas las incógnitas que hoy existen. Sin un experimento será difícil saber bien a bien qué pasó durante el levantamiento. En consecuencia, será difícil lograr un sintético bien fundamentado que satisfaga a todas las partes. El experimento supone diseñar una encuesta con diseño longitudinal que siga por algún tiempo al mismo conjunto de personas. Solo así se podrán definir bien grupos de control y de tratamiento, además de garantizar que exista soporte común. En principio se requiere someter un grupo de individuos solo al protocolo 2014, otro grupo únicamente al protocolo 2015 y un tercero a los dos protocolos. Algo similar se necesita diseñar del lado del encuestador. No se requiere representatividad a nivel estado o municipio para explorar las principales hipótesis, una encuesta longitudinal con representatividad nacional es suficiente. Esto es posible lograrlo con tamaños de muestra moderados.

Más sustancialmente, debemos decidir cómo proceder en el futuro. En particular, es necesario responder dos preguntas fundamentales: 1) ¿el instrumento anterior medía el ingreso de forma adecuada? y 2) ¿es posible medir de forma correcta —y atribuir una interpretación sustantiva— variaciones de la pobreza a través del tiempo con datos de corte transversal?

Desde el punto de vista de los autores, calcular variaciones en la pobreza a través del tiempo con base en datos de corte transversal es inadecuado. El estado del arte nos invita a hacer uso de datos longitudinales. Solo así podremos medir tasas de transición dentro y fuera de la pobreza a través del tiempo.

6. Referencias

Biemer, P., S. Sudman, R. M. Groves, N. Mathiowetz y L. Lyberg. *Measurement Errors in Surveys*. Wiley Series in Probability and Statistics. Wiley, 2014.

- Bound, J., C. Brown, G. Duncan y W. Rodgers. "Evidence on the Validity of CrossSectional and Longitudinal Labor Market Data", en: *Journal of Labor Economics*, 12(3):345-368, 1994.
- Bound, J., C. Brown y N. Mathiowetz. "Measurement error in survey data", en: Heckman, J. and E. Leamer (editors). *Handbook of Econometrics, volume 5 of Handbook of Econometrics*, chapter 59, pages 3705-3843, 2001. Elsevier.
- Bound, J. y A. B. Krueger. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?", en: *Journal of Labor Economics*, 9(1):1-24, 1991.
- Bustos, A. "Estimation of the distribution of income from survey data, adjusting for compatibility with other sources", en: *Journal of the International Association for Official Statistics*, 31(4):565-577, 2015.
- ChangHwan, K. y C. Tamborini. "Response error in earnings", en: *Sociological Methods & Research*, 43(1):39-72, 2014.
- Chesher, A. y C. Schluter. "Welfare measurement and measurement error, en: *The Review of Economic Studies*, 69(2):357-378, 2002.
- Duncan, G. J. y D. H. Hill. "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data", en: *Journal of Labor Economics*, 3(4):508-532, 1985.
- Gottschalk, P. y M. Huynh. "Are Earnings Inequality and Mobility Overstated? The Impact of Nonclassical Measurement Error", en: *The Review of Economics and Statistics*, 92(2):302-315, 2010.
- Haughton, J. y S. Khandker. *Handbook on Poverty and Inequality*. Number 11985 in World Bank Publications. The World Bank, 2009.
- Hausman, J. "Mismeasured variables in econometric analysis: Problems from the right and problems from the left", en: *The Journal of Economic Perspectives*, 15(4):57-67, 2001.
- Heckman, J. J. "Sample Selection Bias as a Specification Error", en: *Econometrica*, 47(1):153-161, 1979.
- Hurst, E., G. Li y B. Pugsley. "Are Household Surveys Like Tax Forms? Evidence from Income Underreporting of the Self-Employed", en: *The Review of Economics and Statistics*, 96(1):19-33, 2014.
- Hyslop, D. y W. Townsend. "Earnings Dynamics and Measurement Error in Matched Survey and Administrative Data", en: *Working Papers 1618, MotuEconomicandPublicPolicyResearch*. 2016.
- Hyslop, D. R. y G. W. Imbens. "Bias from Classical and Other Forms of Measurement Error", en: *Journal of Business & Economic Statistics*, 19(4):475-481, 2001.
- Jenkins, S. "Recent trends in the uk income distribution: what happened and why?", en: *Oxford Review of Economic Policy*, 12(1):29-46, 1996.
- Kasprzyk, D. "Measurement error in household surveys: Sources and measurement", en: *Mathematica policy research reports, Mathematica Policy Research*. 2005.
- Little, R. J. A. y D. B. Rubin. *Statistical analysis with missing data*. Hoboken, NJ, Wiley, 2002.
- Mincer, J. "Investment in Human Capital and Personal Income Distribution", en: *Journal of Political Economy*, 66:281-281, 1958.
- Nicoletti, C., F. Peracchi y F. Foliano. "Estimating Income Poverty in the Presence of Missing Data and Measurement Error", en: *Journal of Business & Economic Statistics*, 29(1):61-72, 2011.
- Parmeter, C. "The effect of measurement error on the estimated shape of the world distribution of income", en: *Economics Letters*, 100(3):373-376, 2008.
- Rodgers, W. L., C. Brown y G. J. Duncan. "Errors in survey reports of earnings, hours worked, and hourly wages", en: *Journal of the American Statistical Association*, 88(424):1208-1218, 1993.
- Rubin, D. B. "Inference and missing data", en: *Biometrika*, 63:581-592, 1976.
- Ruiz, J. y A. Romo. *Armonización de los ingresos del MCS 2015 usando la ENOE como contrafactual*. Mimeo. 2017.

Colaboran en este número

Víctor Alfredo Bustos y de la Tijera

De nacionalidad mexicana, es licenciado en Actuaría por la Facultad de Ciencias de la Universidad Nacional Autónoma de México (UNAM), maestro en Estadística e Investigación de Operaciones por el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) de la UNAM y doctor en Estadística por la *London School of Economics*. Ha sido catedrático en la Facultad de Ciencias de la UNAM, profesor-investigador del IIMAS, así como profesor en la *School of Business Administration* de la *Temple University*, el Instituto Tecnológico Autónomo de México (ITAM) y el Instituto Tecnológico y de Estudios Superiores de Monterrey, además de catedrático-investigador en el Centro de Investigación y Docencia Económicas (CIDE) y profesor invitado en el Centro Científico de IBM de México. Ingresó al Instituto Nacional de Estadística y Geografía (INEGI) en 1991, donde actualmente es investigador en la Dirección General Adjunta de Investigación. Sus áreas de interés son, entre otras: el ajuste de modelos a partir de encuestas, donde destaca la estimación de la distribución del ingreso con datos de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), compatible con las cuentas nacionales; la estimación para dominios pequeños en temas como la desocupación y la migración; la clasificación multivariada con aplicación a la medición del bienestar y la marginación; y la construcción de índices temporales multivariados para seguir la evolución de una economía mediante criterios explícitos.

Contacto: alfredo.bustos@inegi.org.mx

Benito Durán Romo

Es de nacionalidad mexicana. Tiene la Licenciatura en Informática por el Instituto Tecnológico de Aguascalientes, con una primera maestría en la misma materia por la Universidad Autónoma de Aguascalientes (UAA) y una segunda en Análisis Estadístico por el Centro de Investigación en Matemáticas (CIMAT), AC. Desde 1989 labora en el INEGI, donde inició en el procesamiento de encuestas especiales en la Dirección de Estadísticas de Corto Plazo; posteriormente, colaboró en el procesamiento de la ENIGH para luego fungir como responsable del área de procesamiento de ésta; en la actualidad, es subdirector de Investigación en Indicadores Sociales, Demográficos y Económicos en la Dirección General Adjunta de Investigación del INEGI, donde proporciona apoyo técnico a diversos proyectos de investigación y desarrolla aplicaciones web interactivas que acercan a la población en general al tema del bienestar y el progreso de las sociedades.

Contacto: benito.duran@inegi.org.mx

Rodrigo Negrete Prieto

De nacionalidad mexicana, es licenciado en Economía por la UNAM. En el ámbito laboral, fue director de Encuestas en Hogares del INEGI y miembro de grupos de expertos internacionales por parte de la Organización Internacional del Trabajo, la Comisión Económica de las Naciones Unidas para

Europa y la Organización para la Cooperación y el Desarrollo Económicos con el fin de discutir y actualizar los marcos conceptuales para la medición estadística tanto del trabajo en general como del empleo informal en particular; a partir del 2009, es integrante del área de investigación del INEGI. Ha publicado más de una veintena de artículos en revistas especializadas y de debate a nivel nacional.

Contacto: rodrigo.negrete@inegi.org.mx

José Alejandro Ruiz Sánchez

Nació en la Ciudad de México. Cuenta con una Maestría en Economía por el Centro de Investigación y Docencia Económicas (CIDE) y una Maestría en Teoría Económica por el ITAM. En el ámbito profesional, su trayectoria se ha centrado en la administración pública federal, tanto en la Secretaría de Energía como en la Secretaría de Desarrollo Social. Desde enero de 2016 se incorporó al INEGI como Investigador dentro de la Dirección General Adjunta de Investigación. Sus áreas de especialización incluyen métodos de aprendizaje de máquina aplicados a las ciencias sociales, evaluaciones de impacto, econometría aplicada, productividad y crecimiento económico.

Contacto: jose.ruizs@inegi.org.mx

Ana Miriam Romo Anaya

Es licenciada en Matemáticas Aplicadas por la UAA, Maestra en Ciencias en Estadística Oficial por el Centro de Investigación en Matemáticas, AC. Desde el 2009 ha participado en la Dirección General Adjunta de Investigación del INEGI en el desarrollo y análisis de técnicas estadísticas, como: muestreo, modelos lineales y no lineales, técnicas de agrupamiento y clasificación, optimización de funciones paramétricas y el desarrollo computacional estadístico. Ha sido catedrática en la UAA y en la Universidad Politécnica de Aguascalientes impartiendo en diversos periodos cursos de Matemáticas y Estadística.

Contacto: miriam.romo@inegi.org.mx

Delfino Vargas Chanes

De nacionalidad mexicana, es profesor investigador del Programa Universitario de Estudios del Desarrollo en la UNAM. Obtuvo el doctorado en Sociología en la Universidad Estatal de Iowa (*Ph. D.*) en el 2000, los grados de Maestría en Sociología (1996) y en Estadística (1994) en la Universidad Estatal de Iowa y la Licenciatura en Matemáticas (1982) en la UNAM. Ha publicado más de 40 artículos en diversas revistas académicas arbitradas (nacionales e internacionales) en las áreas de Estadística, Medicina, Psicología, Sociología y Administración, entre otras. Es coautor de dos libros. Sus intereses de investigación son la desigualdad y la pobreza, así como el desarrollo de metodologías cuantitativas avanzadas para la investigación social. En la actualidad, pertenece al Sistema Nacional de Investigadores (SIN) con nivel II.

Contactos: dvchanes@unam.mx y dvchanes@gmail.com

Servando Valdés Cruz

De nacionalidad mexicana, obtuvo la Licenciatura en Actuaría por la Facultad de Ciencias de la UNAM en el 2017. Fue becario de investigación en el Centro de Estudios Demográficos Urbanos y Ambientales de El Colegio de México, ayudante de investigador tanto en el Programa Universitario de Estudios del Desarrollo (PUED) de la UNAM del 2015 al 2017 como en el Sistema Nacional de Investigadores del Consejo Nacional de Ciencia y Tecnología (CONACYT). Fue ponente en el 27.º Simposio Internacional de Estadística y en el 8th *International Workshop on Spatio-Temporal Modelling METMA VIII*. Impartió cursos de Estadística e Indicadores Sociales, Econometría y Estadística Espacial en el programa de posgrado de la Facultad de Economía y de Ciencias en la UNAM, así como en la Facultad Latinoamericana de Ciencias Sociales (FLACSO) sede México. Sus áreas de interés son: pobreza, desigualdad y modelos espaciales aplicados en pobreza. En la actualidad, es técnico académico en el PUED.

Contacto: servando.valdes.cruz@gmail.com

Edgar Vielma Orozco

De nacionalidad mexicana, es matemático por la Universidad de Guadalajara (UDG) y maestro en Economía por el ITAM. Hoy en día, labora en el INEGI como director general de Estadísticas Sociodemográficas, área responsable del Censo de Población y Vivienda, así como de la medición del empleo, la ocupación y la informalidad laboral, del ingreso monetario y no monetario de los hogares y de las variables que miden las carencias de los hogares, entre otros temas relativos al ámbito sociodemográfico. Anteriormente, fue director general adjunto de Modelos de Información Gubernamental y Encuestas Nacionales de Gobierno, Victimization, Seguridad y Justicia, así como director de Modelos de Información Gubernamental, ambos cargos en el INEGI, áreas en las que destacan la medición de los delitos y delitos no reportados ante las autoridades (*cifra negra*) a partir de encuestas probabilísticas tanto en hogares como en empresas, la calidad de los servicios públicos, la mejora regulatoria y la medición de la corrupción, además de la primera encuesta nacional a población penitenciaria en México. En la administración pública federal, fungió como director de Modelos y Estadísticas de Órganos de Vigilancia y Control y subdirector de Evaluación de Órganos de Vigilancia y Control en la Secretaría de la Función Pública. En el área académica, ha sido profesor en el Departamento de Matemáticas de la UDG.

Contacto: edgar.vielma@inegi.org.mx

Octavio Heredia Hernández

Nació en México. Estudió la Licenciatura en Matemáticas Aplicadas en la UAA y cursó la Maestría en Análisis Estadístico en el CIMAT. Como docente, ha impartido materias relacionadas con probabilidad, teoría estadística, modelación matemática, investigación de operaciones y estadística para diferentes niveles y carreras en la UAA y en la Universidad Politécnica de Aguascalientes.

Dentro del sector público se desempeñó como director de Encuestas Nacionales de Gobierno y como director de Diseño y Marcos Estadísticos en el INEGI; en la actualidad, es titular de la Dirección General Adjunta de Encuestas Sociodemográficas del INEGI.

Contacto: octavio.heredia@inegi.org.mx

José Vences Rivera

Es de nacionalidad mexicana. Estudió la carrera de Física y Matemáticas en el Instituto Politécnico Nacional, tiene la Maestría en Estadística Experimental por el Colegio de Postgraduados y cursó un posgrado en Demografía en el Centro Latinoamericano de Demografía (CELADE)-Universidad de Costa Rica. Como docente, ha impartido cursos en diferentes maestrías de la UAA. Ha sido conferencista sobre estadística oficial y superación de la pobreza en diversas universidades del país y el extranjero. Es autor del libro *Estadística multivariada análisis de factores*. Fue miembro del Grupo de Río-CEPAL sobre Estadísticas de Pobreza y del Comité Técnico para la Medición de la Pobreza en México, así como representante del INEGI en este tema para la Conferencia Estadística de las Américas. Es miembro activo de la Sociedad Matemática Mexicana y de la Asociación Mexicana de Estadística. Fue cofundador de la Maestría en Ciencias en Estadística Oficial INEGI-CIMAT y coordinador académico de la misma. En la actualidad, se desempeña como director de Desarrollo de Procesos Estadísticos en el INEGI.

Contacto: jose.vences@inegi.org.mx

Alida M. Gutiérrez-Landeros

Nació en México. Es licenciada en Economía, así como maestra y doctora en Economía Social por la Universidad Autónoma Metropolitana (UAM). Actualmente, desempeña el cargo de directora general adjunta de Medición de Pobreza y Seguimiento de Indicadores en el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL). Sus temas de interés son la medición de la pobreza, econometría, economía del sector público y economía laboral.

Contacto: amgutierrez@coneval.org.mx

Marco A. Romero-Navarro

De nacionalidad mexicana, es licenciado en Economía por la UNAM. En la actualidad, desempeña el cargo de subdirector de Monitoreo de Fuentes en el CONEVAL. Sus temas de interés son la docencia de métodos cuantitativos aplicados a la ciencia económica, el análisis dinámico de la pobreza y la aplicación de métodos econométricos a las encuestas en hogares.

Contacto: mromero@coneval.org.mx

Mariana Galindo-Orozco

Es de nacionalidad mexicana. Estudió la Licenciatura en Economía en el ITAM. Actualmente, desempeña el cargo de subdirectora de Análisis Económico en

el CONEVAL. Sus temas de interés son la relación entre crecimiento económico, pobreza y desigualdad, así como economía laboral.

Contacto: mrgalindo@coneval.org.mx

Carolina Chávez-Ruelas

Nació en México. Es licenciada en Economía por el ITAM. En la actualidad, desempeña el cargo de jefa del Departamento de Desarrollo de Técnicas y Herramientas para el Análisis de la Pobreza en el CONEVAL. Sus temas de interés son la economía ambiental, el desarrollo social y la desigualdad, en particular desde una perspectiva de género y conductual.

Contacto: cchavez@coneval.org.mx

David Rojas-Rosey

Es de nacionalidad mexicana. Obtuvo la Licenciatura en Economía por la UAM. Actualmente, desempeña el cargo de subdirector de Medición de la Pobreza en el CONEVAL. Sus temas de interés son la estadística aplicada, el análisis de la desigualdad y métodos multivariados aplicados al desarrollo social.

Contacto: drojas@coneval.org.mx

David López-Lira Bayod

De nacionalidad mexicana, es licenciado en Economía y cuenta con la especialidad en Economía Ambiental y Ecología, ambas por la UNAM; además, es maestro en Estudios Urbanos por El Colegio de México. En la actualidad, desempeña el cargo de jefe del Departamento de Seguimiento de Indicadores Económicos en el CONEVAL. Su tema de interés es el estudio de los condicionantes socioeconómicos y geográficos que inciden en la calidad de vida de las personas, con énfasis en la movilidad urbana y el transporte.

Contacto: dlopez@coneval.org.mx

Ricardo Aparicio-Jiménez

Nació en México. Es licenciado en Actuaría por la UNAM y maestro en Estadística por la *London School of Economics*. Actualmente, desempeña el puesto de director general adjunto de Análisis de la Pobreza en el CONEVAL. Su tema de interés es el análisis de los fenómenos económicos y sociales desde una perspectiva de derechos humanos.

Contacto: rcaparicio@coneval.org.mx

Alfonso Miranda

Nació en México. Es licenciado en Economía por la UNAM y cuenta con dos maestrías —una por el CIDE y otra por la Universidad de Warwick (UW), Reino Unido—, además del Doctorado en Economía por la UW. Es profesor-investigador en la División de Economía del CIDE desde agosto del 2012 y miembro del Sistema Nacional de Investigadores con nivel II desde enero del 2015; asimismo, se desempeña como director del Programa de Estudios Longitudinales, Experimentos y Encuestas del CIDE, investigador afiliado del Grupo de Salud, Econometría y Datos de la Universidad de York e investigador asociado del

Instituto para el Estudio del Trabajo, Alemania; anteriormente, fue profesor en Economía en la Universidad de Keele, Reino Unido (2004-2008), así como en Ciencia Social Cuantitativa en el Instituto de Educación de la *University College London*, Reino Unido (2008-2012). Trabaja en temas de economía de la educación, de la salud y laboral. Ha publicado en revistas académicas internacionales de prestigio, incluyendo *PLoS One*, *Health Economics*, *Economics Letters*, *Journal of Population Economics*, *Journal of the Royal Statistical Society Series A*, *Cancer*, *Fiscal Studies* y *Research in Labor Economics*, and *Stata Journal*. Tiene experiencia en el diseño y análisis de encuestas sociales con diseño longitudinal y de corte transversal, así como en el uso de datos administrativos ligados a encuestas para la evaluación de políticas públicas en el Reino Unido y México. Ha desarrollado métodos econométricos para corregir problemas de selección de muestra, tratamiento endógeno y participación endógena en modelos econométricos no lineales.

Contacto: alfonso.miranda@cide.edu

Jaime Sainz Santamaría

De nacionalidad mexicana, estudió la Licenciatura en Ciencia Política en la UAM, Unidad Iztapalapa, obtuvo la Maestría en Políticas Públicas en el ITAM y el Doctorado en Ciencias Ambientales en la *University of California Santa Barbara* (UCSB). Es profesor-investigador de la División de Administración Pública y director de la sede Región Centro del CIDE, así como investigador principal del Programa de Estudios Longitudinales, Experimentos y Encuestas del CIDE. Ha publicado sobre economía política de desastres naturales, conflictos por agua, cambio climático y valoración económica. Ha participado en el diseño, implementación, coordinación en campo y análisis de encuestas en ejidos y comunidades con el fin de apoyar el diseño de políticas públicas para la conservación ambiental.

Contacto: jaime.sainz@cide.edu

Política y lineamientos editoriales

REALIDAD, DATOS Y ESPACIO REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA es una publicación cuatrimestral que sirve de enlace entre la generación de la información estadística y geográfica oficial y la investigación académica para compartir el conocimiento entre especialistas e instituciones con propósitos similares.

Se publicarán sólo artículos inéditos y originales relacionados con la situación actual del uso y aplicación de la información estadística y geográfica a nivel nacional e internacional.

Es una revista técnico-científica, bilingüe, cuyos trabajos son arbitrados por pares (especialistas), bajo la metodología doble ciego, con los siguientes criterios de evaluación: trabajos inéditos, originalidad, actualidad y oportunidad de la información, claridad en la definición de propósitos e ideas planteadas, cobertura de los objetivos definidos, estructura metodológica adecuada y congruencia entre la información contenida en el trabajo y las conclusiones.

El resultado del proceso de dictaminación se comunica por correo electrónico y contempla tres variantes: recomendado ampliamente (con modificaciones menores), recomendado (pero condicionado a modificaciones sugeridas) y no recomendado (rechazado). Dos dictámenes aprobados, se notifica al autor que se publica y se envía a corrección de estilo; un aprobado y uno rechazado, se le solicita realizar cambios; y dos rechazados, se notifica la no publicación.

Indizaciones y registros

- LATINDEX Catálogo (Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal).
- CLASE (Citas Latinoamericanas en Ciencias Sociales y Humanidades).
- REDIB (Red Iberoamericana de Innovación y Conocimiento Científico).

Lineamientos para publicar

Se publicarán trabajos en español e inglés: artículos de investigación, revisión y divulgación; ensayos; metodologías; informes técnicos; comunicaciones cortas; reseñas de libros; revisiones bibliográficas y estadísticas, entre otros.

1. El artículo —o cualquier otro tipo de escrito de los mencionados— deberá entregarse con una carta dirigida al editor responsable de REALIDAD, DATOS Y ESPACIO. REVISTA INTERNACIONAL DE ESTADÍSTICA Y GEOGRAFÍA en la que se proponga el texto para su publicación, que se declare que es inédito y que no ha sido postulado de manera paralela en otro medio. Asimismo, deben incluirse los datos completos del(os) autor(es), nacionalidad(es), institución(es) de adscripción y cargo(s) que ocupa(n), domicilio(s) completo(s), correo(s) electrónico(s) y teléfono(s). Esto debe dirigirse a la atención de la M. en C. Virginia Abrin Batule, virginia.abrin@inegi.org.mx (tel. 5278 10 00, ext. 1161).
2. El trabajo se debe presentar en versión electrónica (formato *Word* o compatible) con: a) extensión no mayor de 20 cuartillas; b) letra Helvética, Arial o Times de 12 puntos y c) interlineado de 1.5 líneas. El material adicional al texto se requiere por separado: a) las imágenes, con resolución de 300 ppp y un tamaño no menor a 17 centímetros de base (ancho) en formato JPG o TIF —no remuestrear (ampliar) imágenes de menor resolución—; si son líneas o mapas, deben entregarse en formato vectorial (EPS o Ai), en caso de incluirse imágenes en mapa de bits, incrustarlas o enviarlas con el nombre con el cual se creó el vínculo (conservando los requerimientos de resolución y tamaño estipulados); y para fotografías, éstas no deben ser menores a 5 megapíxeles; b) las fórmulas o expresiones matemáticas tienen que elaborarse con el editor de ecuaciones propio de *Microsoft*[™], pero en caso de usar *software* de terceros, incluir en la entrega PDF testigo en el cual figuren exactamente cómo deben representarse; c) las gráficas, que incluyan el archivo en *Excel* con el cual se desarrollaron o, en su defecto, la imagen JPG legible, de origen, en alta resolución; y d) los cuadros, que sean editables, no se deben insertar como imagen.
3. La colaboración debe incluir: título del trabajo (en español e inglés o viceversa); resúmenes del trabajo en español e inglés (que no excedan de un párrafo de 10 renglones); palabras clave en español e inglés (mínimo tres, máximo cinco); bibliografía u otras fuentes; así como breve(s) semblanza(s) del(os) autor(es) que no exceda(n) de un párrafo de cinco renglones y que incluya(n) nacionalidad(es), grado(s) académico(s), principal(es) experiencia(s) profesional(es), adscripción(es) laboral(es) actual(es) y dirección(es) electrónica(s) de contacto.
4. Las referencias bibliográficas u otras fuentes deberán presentarse al final del artículo de la siguiente manera: nombre(s) del(os) autor(es) comenzando por el(los) apellido(s); título de la publicación con cursivas (si se trata de un artículo, debe estar entrecomillado, seguido de coma y la preposición en con dos puntos y, enseguida, el título de la revista o libro donde apareció publicado, con cursivas); país de origen; editorial; lugar y año de edición; página(s) consultada(s). En el caso de las fuentes electrónicas (páginas web) se debe seguir el mismo orden que en las bibliográficas, pero al final se pondrá entre paréntesis DE (dirección electrónica), la fecha de consulta y la liga completa. Se tienen que omitir aquellas que se mencionen como notas a pie de página. Si se aplica la opción de incluir en cuerpo de texto la referencia de nombre de autor y año de la fuente consultada entre paréntesis, sí deben aparecer todas las referencias mencionadas.

Página electrónica: <http://rde.inegi.org.mx>

Editorial Guidelines and Policy

REALITY, DATA AND SPACE INTERNATIONAL JOURNAL OF STATISTICS AND GEOGRAPHY is a four-monthly publication that connects statistics and geographic official information with academic research in order to share knowledge among specialists and institutions with similar aims.

We will publish only original and unpublished articles related to the current use and appliance of statistical and geographical information at both national and international levels.

It is a technical-scientific and bilingual magazine, with articles previously peer-reviewed by specialists under a double-blind methodology with the following evaluation criteria: unpublished works, originality, information related to opportunity and current affairs, we expect clarity in the definition of aims and ideas stated, defined objectives coverage, accurate methodological structure and coherence between the information of the paper as well as its conclusions.

The result of the paper-assessment process is delivered by email, and it involves three possibilities: fully recommended (with slight modifications), recommended (on condition of suggested modifications) and not recommended (i.e. rejected). When there are two reports of approval, the author gets notified that his/her paper will be published and it is sent to a style editing process. When one report approves the paper for publication and another one rejects it, the author is requested to make some changes for the text to be published. If the text submitted receives two non-favourable reports, the author is notified that the text will not be published.

Index and Registers

- LATINDEX Catalogue (Online Regional Information System for Scientific Journals from Latin America, the Caribbean, Spain and Portugal).
- CLASE (Latin American Quotations in Humanities and Social Sciences)
- REDIB (Latin American Net of Innovation and Scientific Knowledge)

Publishing Guidelines

Articles will be published in Spanish or English: research, revision and scientific-spreading articles; methodologies; technical reports; short texts; book reviews; and bibliographical and statistical revisions, among others.

1. The article —or any other kind of text from those aforementioned— must be delivered with an attached letter addressed to the chief editor of Reality, Data and Space. International Statistics and Geography Magazine in which the text intended for publication will be submitted. There it must be stated that the text has not been published, and that it has not been submitted for publication in any other media. The names in full of the authors must be included, as well as their nationalities, adscription institutions, position in those institutions, postal address, e-mail address, and telephone numbers. This must be addressed to MSc Virginia Abrin Batule, Virginia.abrin@inegi.org.mx (tel (+52) (55) 52.78.10.00, extension 1161).
2. The article must be submitted in an electronic version (a Microsoft Word file or a compatible one) with the following format: a) the text should not exceed the 20 pages of length; b) typography must be Helvetic, Arial or Times (12 points); and c) there should be a 1.5 line spacing in each paragraph. Additional material to the text will be delivered separately: a) images with a resolution of 300 ppp and no smaller than 17 cm width will be delivered in format JPG or TIF —please do not amplify images with lower resolution—. If the added materials are lines or maps, these must be delivered in vectorial format (EPS or Ai). If there are images in bits map, these must be embedded or attached with the name of the original file with which the link was created (keeping the resolution and size requirements above stated). As regards to photographs, these should not be inferior as 5 megapixels; b) mathematical expressions or formulae have to be created with the equations editor by Microsoft[™], but in case of using third-parties software, please attach a witness PDF in which the exact representation of mathematical formulae or expressions is contained; c) graphics must include the Excel file in which they were created or a legible image in the original JPG format in high resolution; and d) charts must be editable, and must not be inserted as images.
3. The text must include the following: the article's title (both in English and Spanish); the abstract of the article—both in English and Spanish (not longer than a 10-line paragraph); key words—both in English and Spanish (three as minimum and five as maximum); bibliography and other sources; as well as brief biographical sketches of the authors not exceeding a five-line paragraph each including nationalities, academic titles, main professional experiences, current work-related affiliations, and electronic addresses for the authors to be contacted.
4. Bibliographical references and other sources must be included at the end of the article in the following way: author's name (Surname first), and publication's title (in italics). If it is an article, the title must be in quotation marks followed by a comma and the preposition "en" with semicolon (in Spanish), then it should appear the title of the book or magazine in which the article was published (in italics); country of origin; publishing house, edition year, and consulted pages. As regards to electronic sources (web pages) the same order of the bibliographical references must be followed, but at the end the word "EA" (as for Electronic Address) ("DE" in Spanish) must be added within parenthesis followed by consultation date and the complete reference link. Those web links referred previously as footnotes, must be omitted in this section. However, if the name of the author and the year of the consulted source were included in the main body of the text within parenthesis, all these must be included as part of the bibliographical references.

Webpage: <http://rde.inegi.org.mx>

