

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, explotar 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 desviaciones 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 perceptor 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.

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 Inter-census Survey. Consequently, the purpose of this paper is to explore and, if appropriate, exploit the historical relationship between non-monetary variables and the

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

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 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.



Crowd forming a positive graph/Henrik Sorensens/Getty Images

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 monetarias 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

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	antigüedad
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

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 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 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 modelo 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 3*² 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 sobrerrepresentación de cada grupo en el MCS 2015 ha sido corregida y llevada a los niveles correspondientes a la EIC 2015.

La figura 2 hace evidente la sobrerrepresentación en la muestra de los hogares pequeños a costa de los grandes. Nuevamente, la posestratificación corrige las desviaciones en uno u otro sentido apegándose 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.

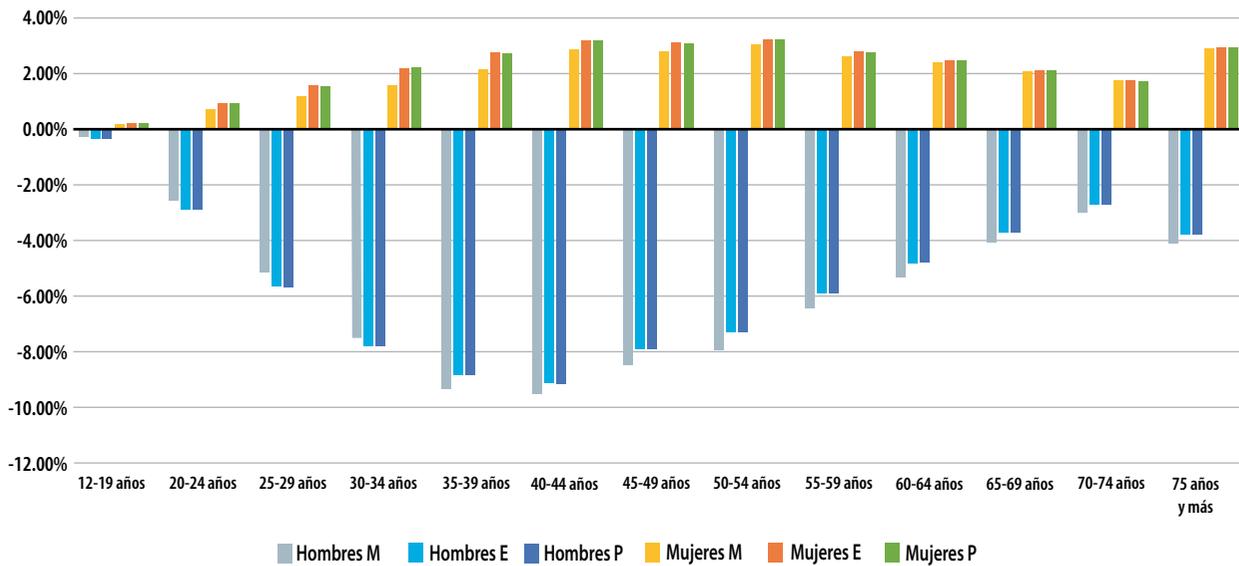
2.2 Introducción o no de variables a nivel estatal

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 recoge todas las diferencias presentadas entre las entidades. Es decir, se hace

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

Figura 1

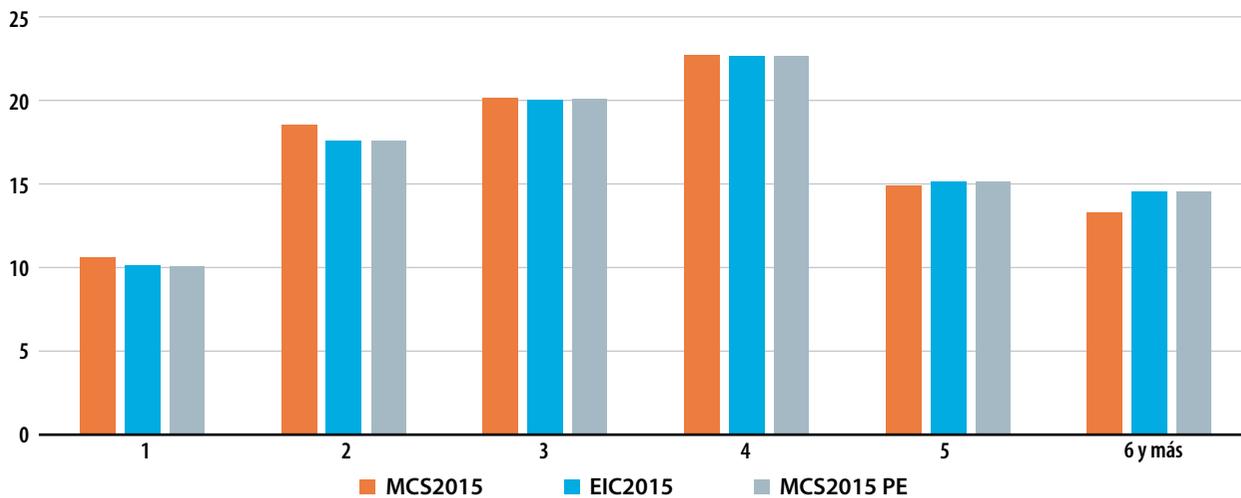
Efecto de la posestratificación en las estructuras relativas por edad y sexo del jefe del hogar, 2015



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.

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.

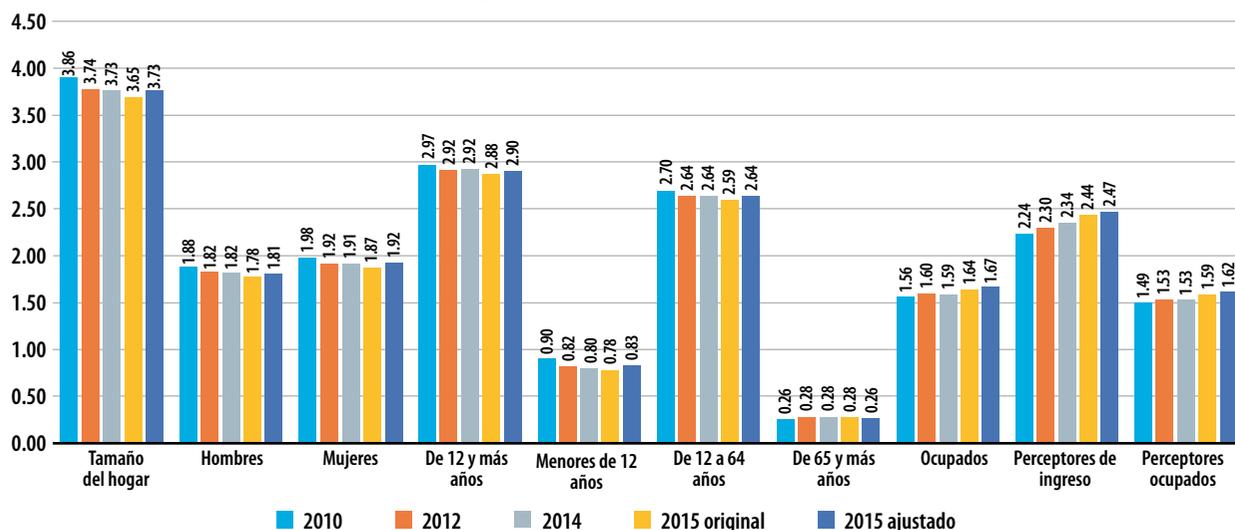
Parámetros del modelo: β_0

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

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

Figura 3

Efecto de la posestratificación en otras variables



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

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 $1 = e_1 + \dots + e_{32}$; o que, despejando, $e_1 = 1 - (e_2 + \dots + e_{32})$. 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:

$$X = \begin{bmatrix} \beta_0 + \beta_1 & \beta_2 - \beta_1 & \beta_{32} - \beta_1 & \alpha_1 \dots \alpha_q \\ \underline{1} & \underline{e}_2 & \dots & \underline{e}_{32} & \underline{x}_1 & \dots & \underline{x}_q \end{bmatrix}$$

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

$$X = \begin{bmatrix} \beta_0 + \beta_1 & \beta_0 + \beta_2 & \beta_0 + \beta_{32} & \alpha_1 \dots \alpha_q \\ \underline{e}_1 & \underline{e}_2 & \dots & \underline{e}_{32} & \underline{x}_1 & \dots & \underline{x}_q \end{bmatrix}$$

La adición de variables a nivel estatal no mejorará el ajuste del modelo. En efecto, la columna \underline{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 $1 - (e_2 + \dots + e_{32})$ 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., $y = y_1 * 1 + (y_2 - y_1) * e_2 + \dots + (y_{32} - y_1) * e_{32}$. 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* 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.

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>.

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>.

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 observación 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.

Instrucción

Continúa

....

```
replace ict_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(educacion_jefe==2,-0.0727829,0,0)+cond(educacion_jefe==  
3,-0.0126718,0,0)+cond(educacion_jefe==4,0.0522974,0,0)+cond(educacion_jefe==5,0.0541993,0,0)+  
cond(educacion_jefe==6,0.0984958,0,0)+cond(educacion_jefe==7,0.1205507,0,0)+  
cond(educacion_jefe==8,0.1860793,0,0)+cond(educacion_jefe==9,0.2323969,0,0)+cond  
(educacion_jefe==10,0.4338892,0,0)+cond(educacion_jefe==11,0.6434069,0,0)+cond(mat_pared==
```

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>.

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>.

```

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

```

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 ignora 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 dichos 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 comparación 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

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

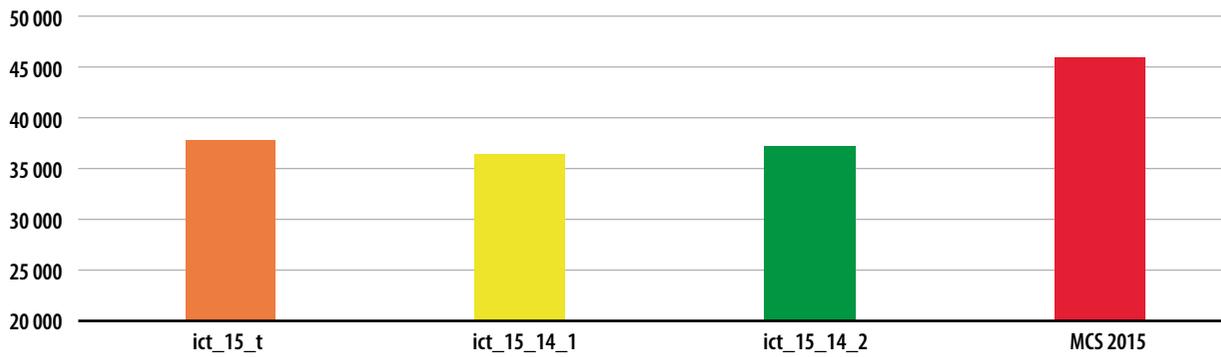
Estadísticas resumen del ajuste de modelos lineales para el levantamiento del MCS para el 2014

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

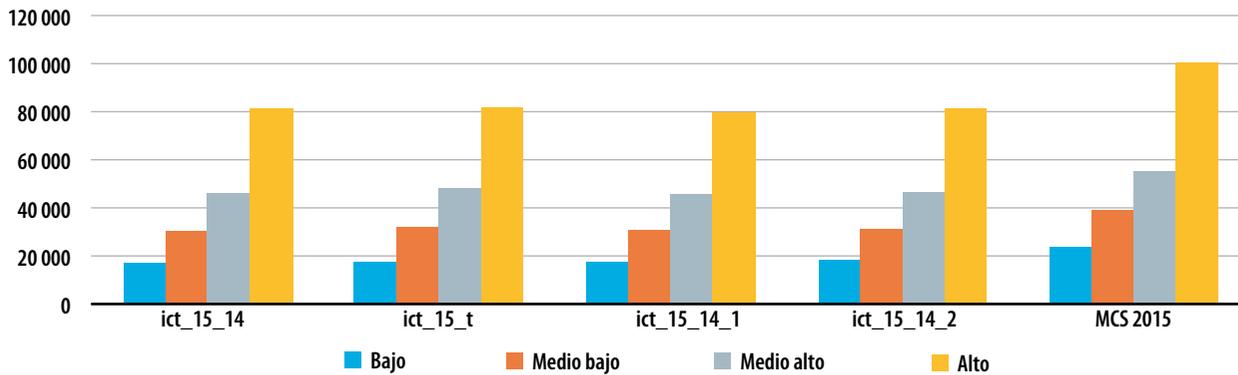
ICT promedio por hogar estimado por diferentes modelos, 2015



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

Figura 5

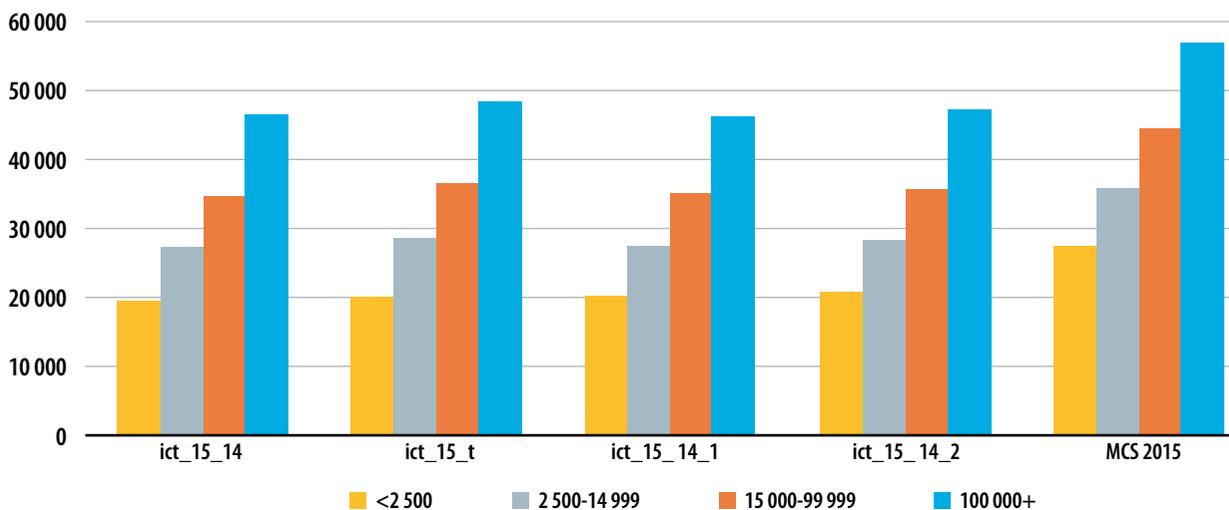
ICT promedio por hogar según estrato sociodemográfico y modelos ajustados, 2015



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

Figura 6

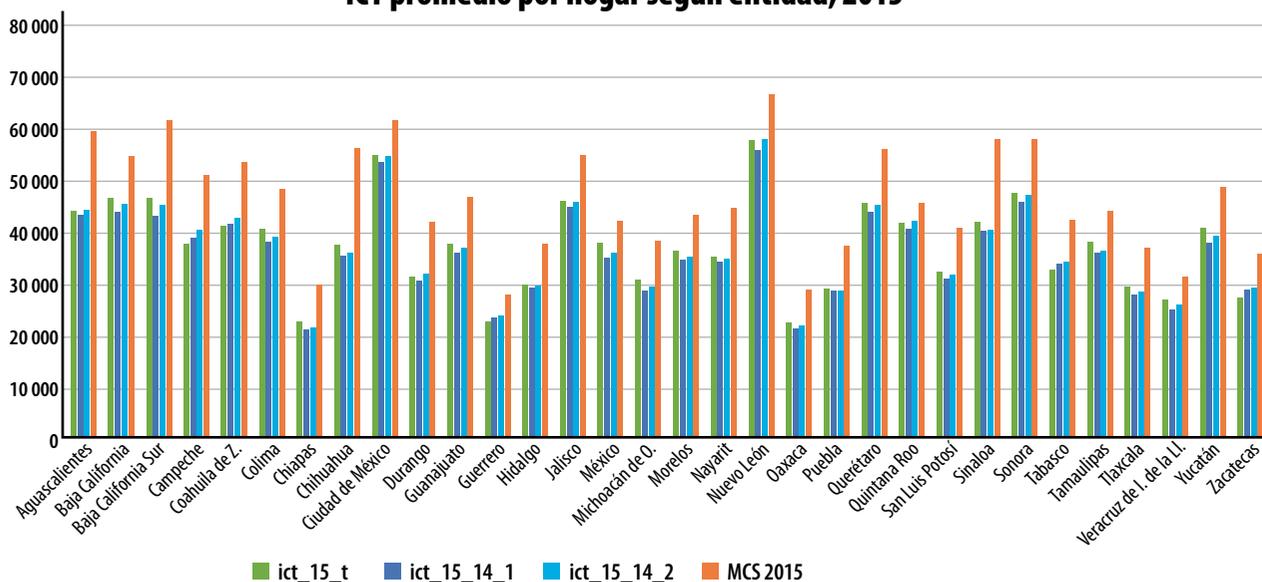
ICT promedio por hogar según tamaño de localidad, 2015



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.

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 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

Tabla 3

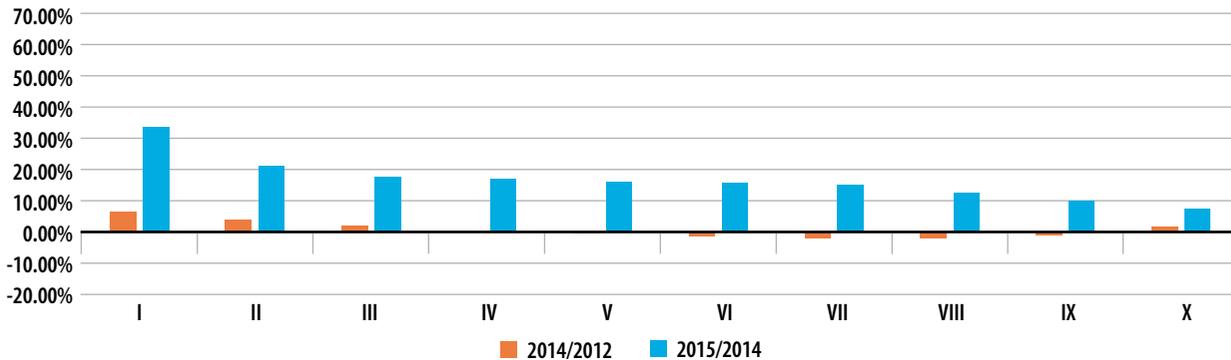
Promedio del ICT por deciles de hogares (precios constantes del 2015)

	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

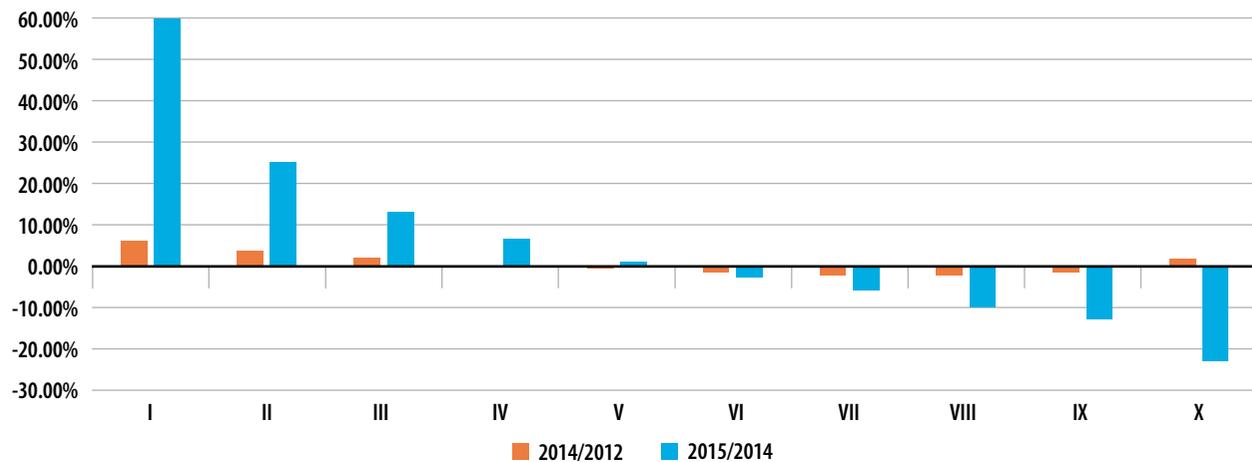
Crecimiento del ingreso corriente promedio, 2012-2014 y 2014-2015, según el INEGI



Fuente: INEGI, archivos del MCS 2015.

Figura 9

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



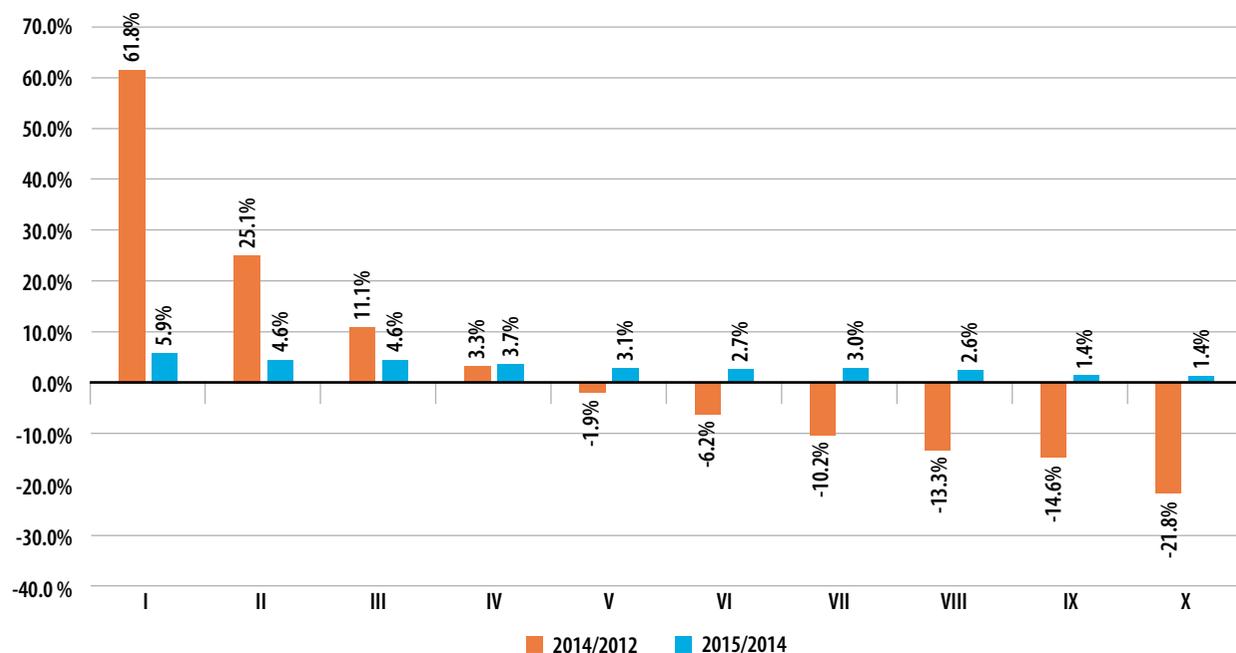
Fuente: INEGI, archivos del MCS 2015 y cálculos propios a partir de las bases de datos del MCS 2014 y 2015.

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 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

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.

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 recortando 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.

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

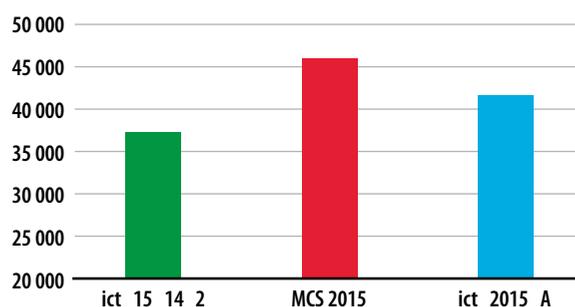
En esta sección se presenta una forma alternativa de *emparejar el piso*, es decir, de construir una nueva base de datos que sea en efecto comparable con las anteriores. En el apartado anterior, dicha comparación se logró cuando tanto los ingresos del módulo 2014 como los del 2015 fueron reemplazados 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 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

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>

Figura 4.a

ICT promedio por hogar estimado por diferentes modelos, 2015



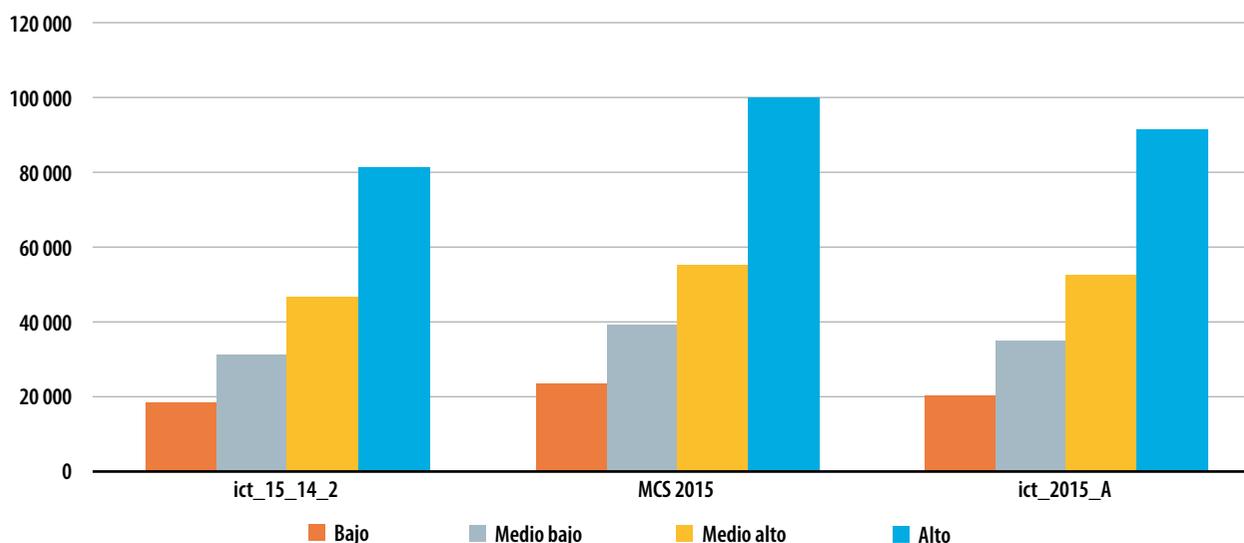
Fuente: cálculos propios a partir de las bases de datos del MCS 2008-2015.

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 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.

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.

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.

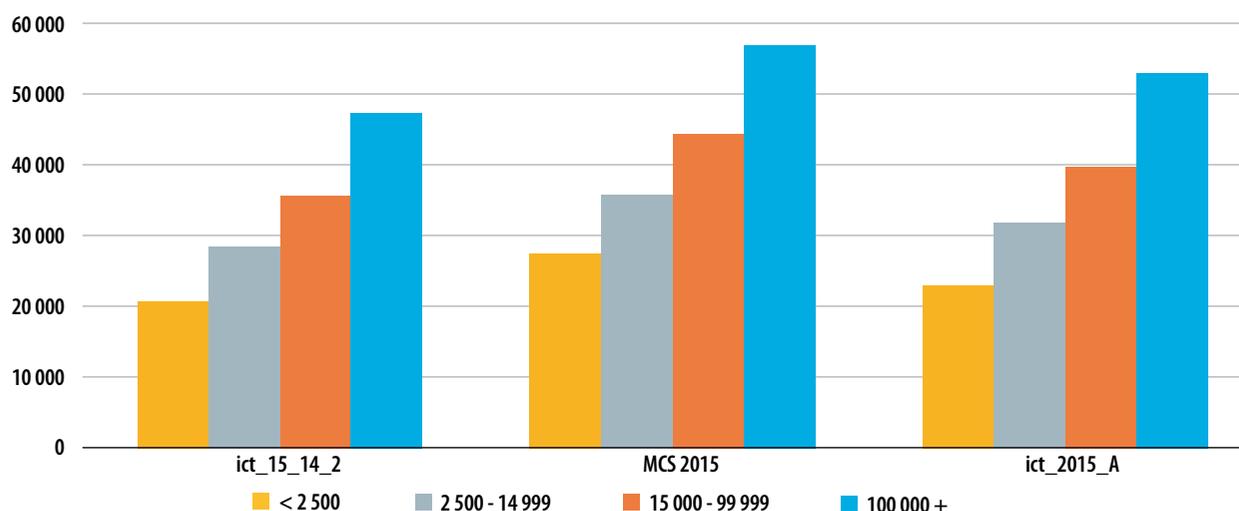
Tabla 3.a

**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.

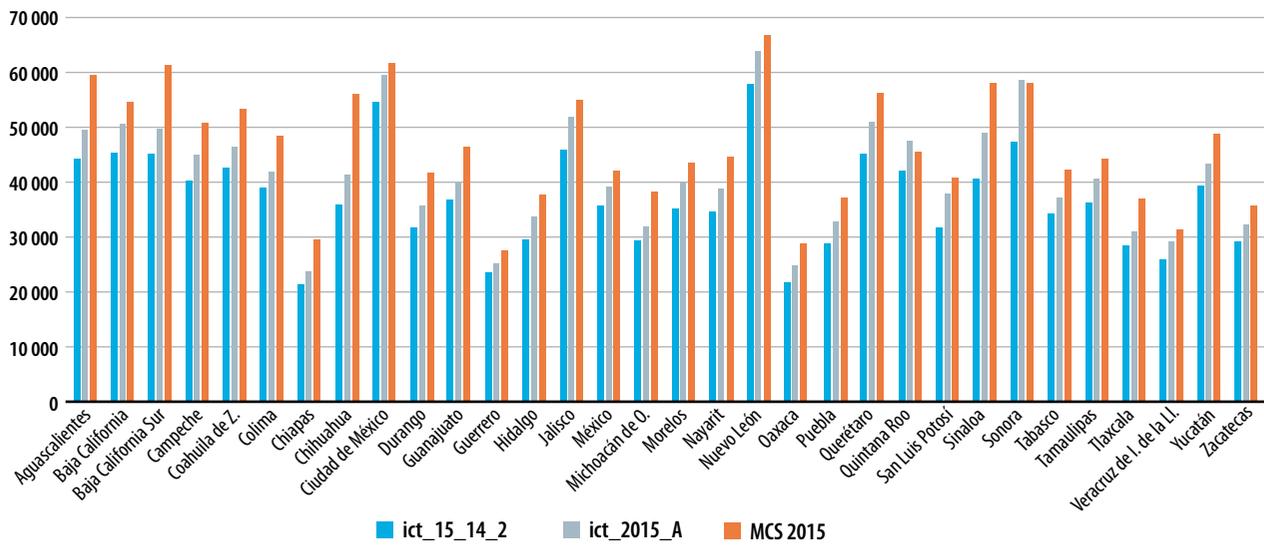
5. Validación y evaluación de la metodología

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 referencia 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.

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.

Figura 7.a

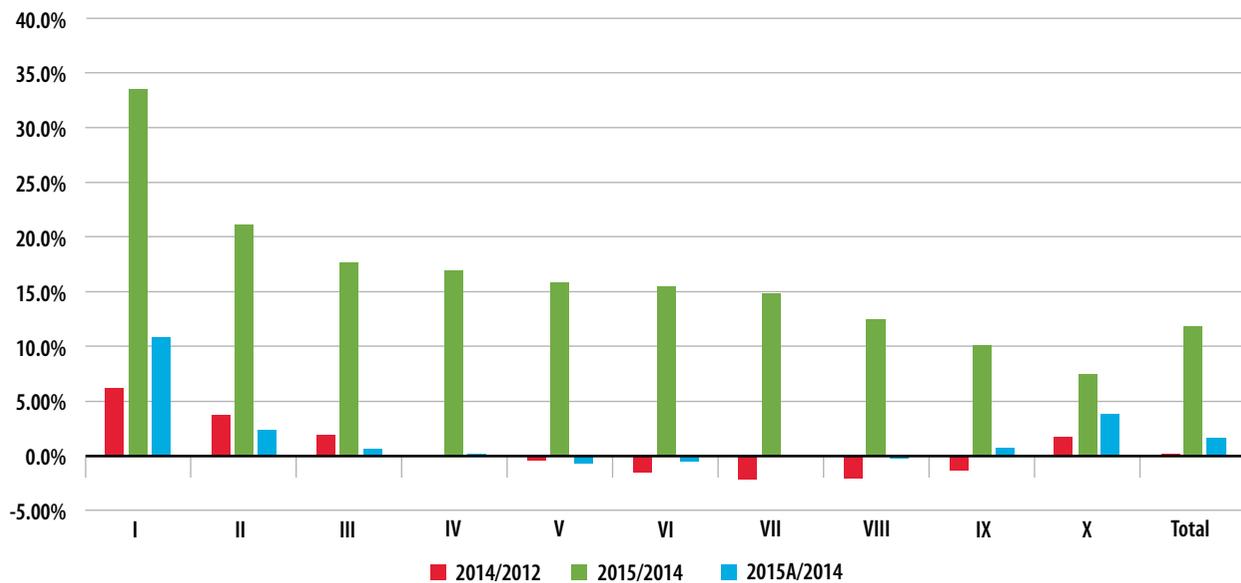
ICT promedio por hogar según entidad, 2015



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.

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%

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

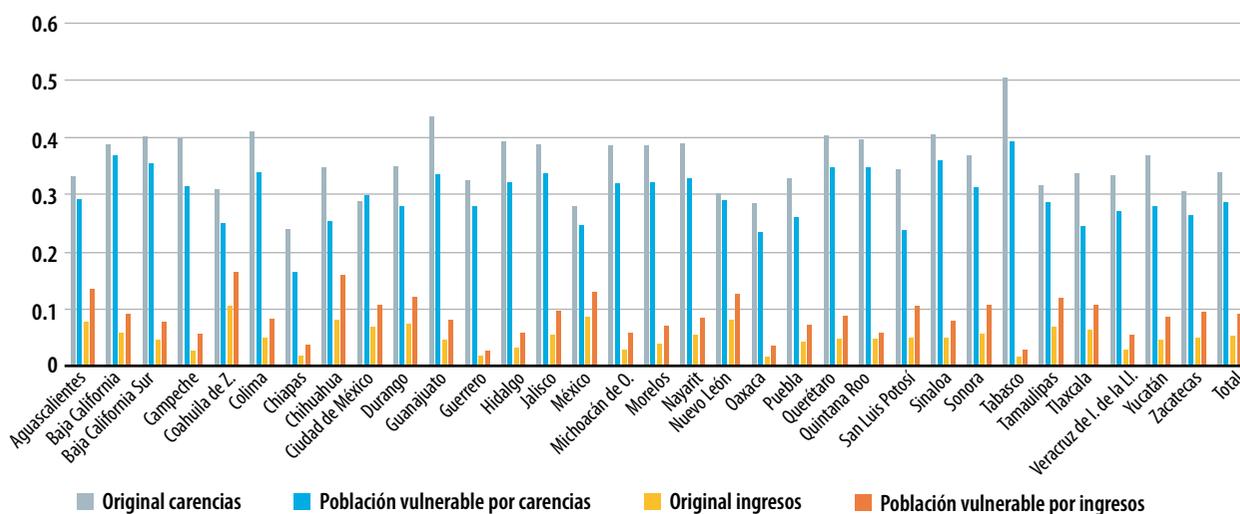
Tabla 5

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

Figura 11

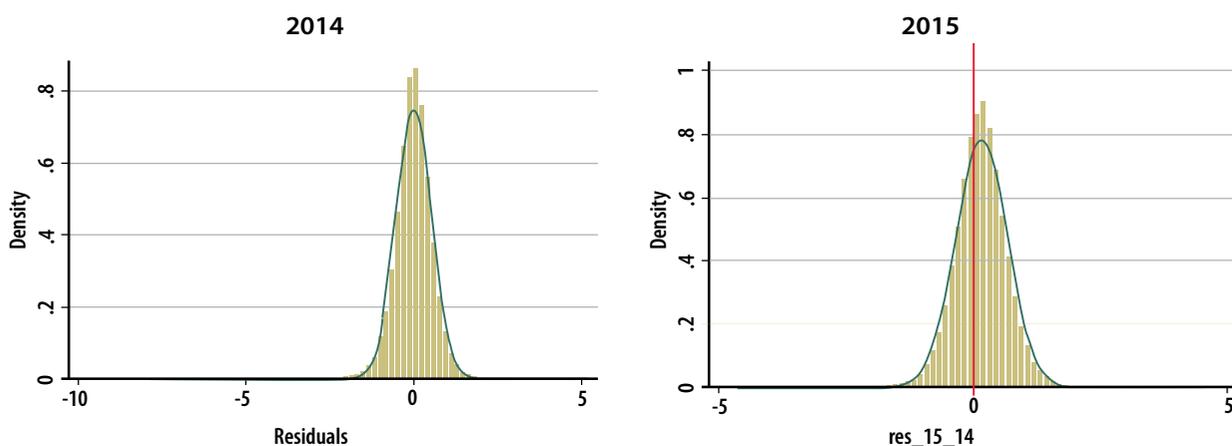
Vulnerabilidad por entidad federativa, 2015



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

Figura 12

Residuos del ajuste, nacional



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

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.

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.

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

5.1 Modelo alternativo

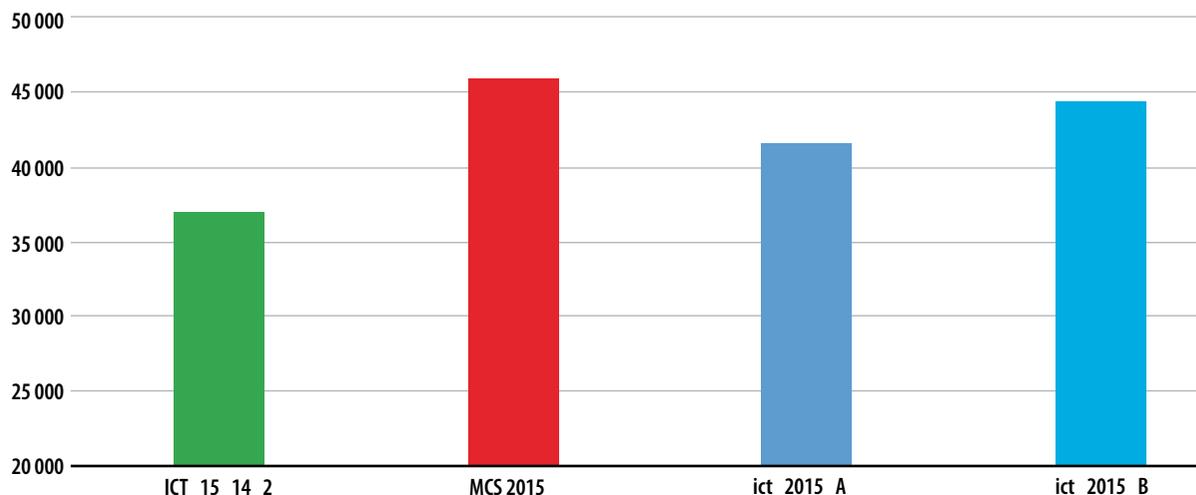
Con el fin de añadir evidencia en favor o en contra del modelo elegido, y siguiendo sugerencias recibidas de quienes escucharon diversas presentaciones del modelo, se llevó a cabo el ajuste de un segundo modelo cuyas características destacables son las siguientes:

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 ($p_{\text{hab_ind}}$).
 - 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.

Figura 4.b

ICT promedio por hogar estimado por diferentes modelos, 2015

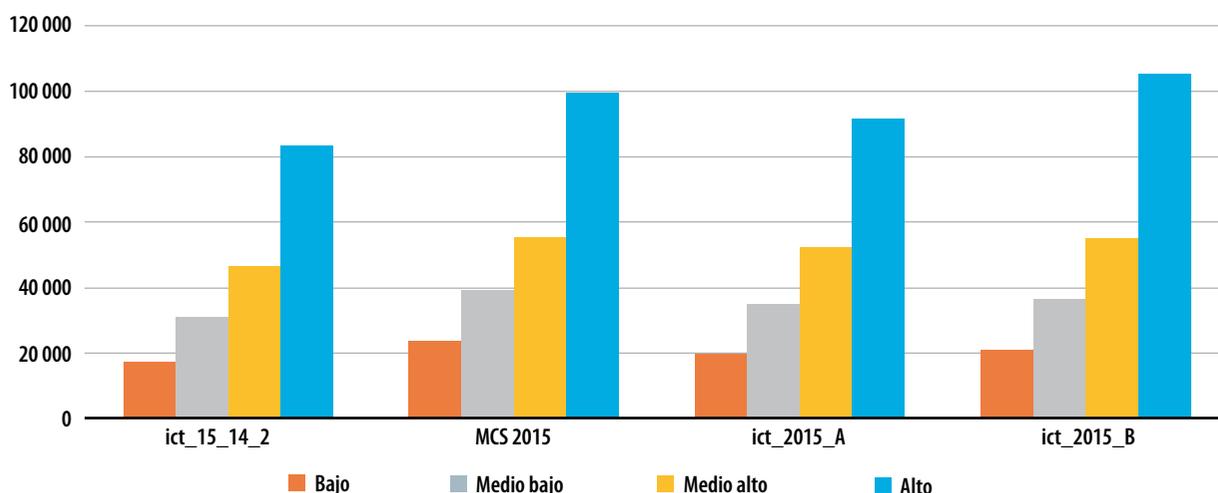


Fuente: INEGI, MCS 2015 y cálculos propios de las bases de datos del MCS 2014 y 2015.

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; enseguida, 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 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.

Figura 5.b

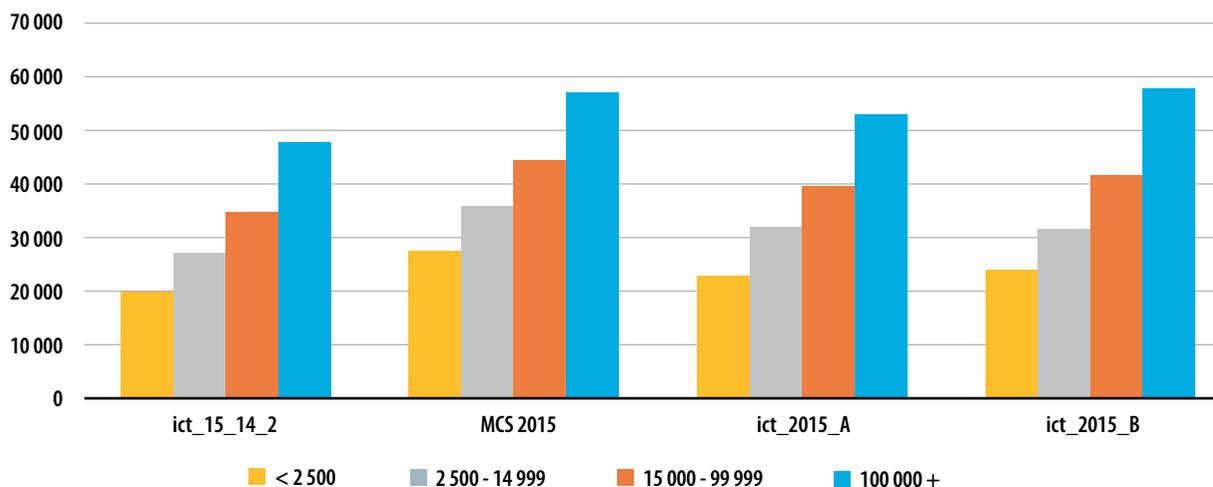
ICT promedio por hogar según estrato sociodemográfico y modelos ajustados, 2015



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.

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.

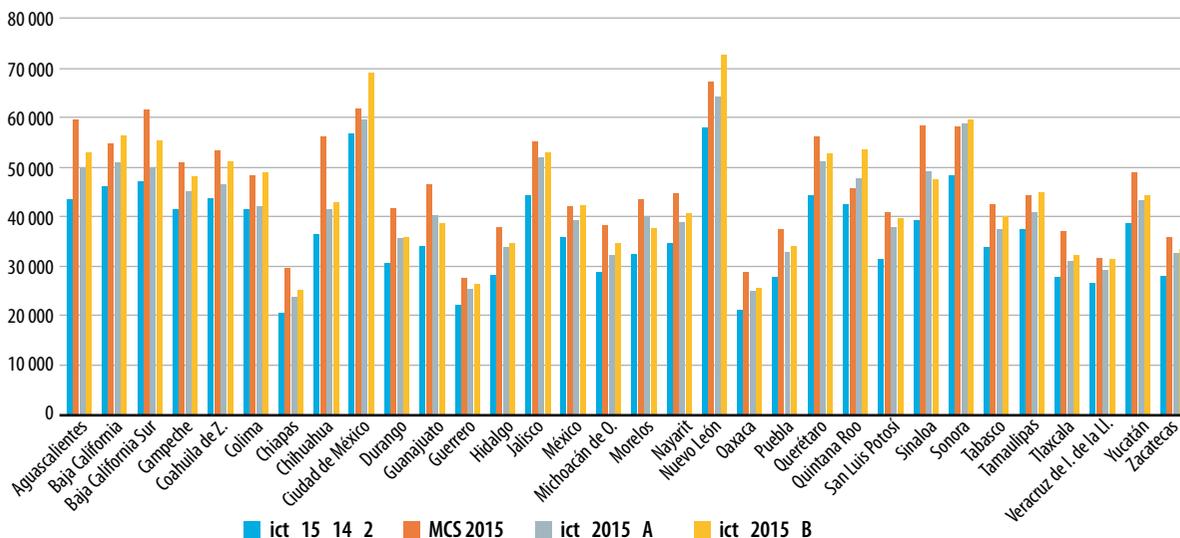
La tabla 1.b muestra la desagregación por deciles de ingreso para los mismos promedios para diversos años. A nivel de los deciles de hogares se tendrían tanto algunos incrementos en los promedios como también algunos decrementos. Esta circunstancia no parece ser inusual ya que algo semejante se observa, 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 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.

Figura 7.b

ICT promedio por hogar según entidad, 2015



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

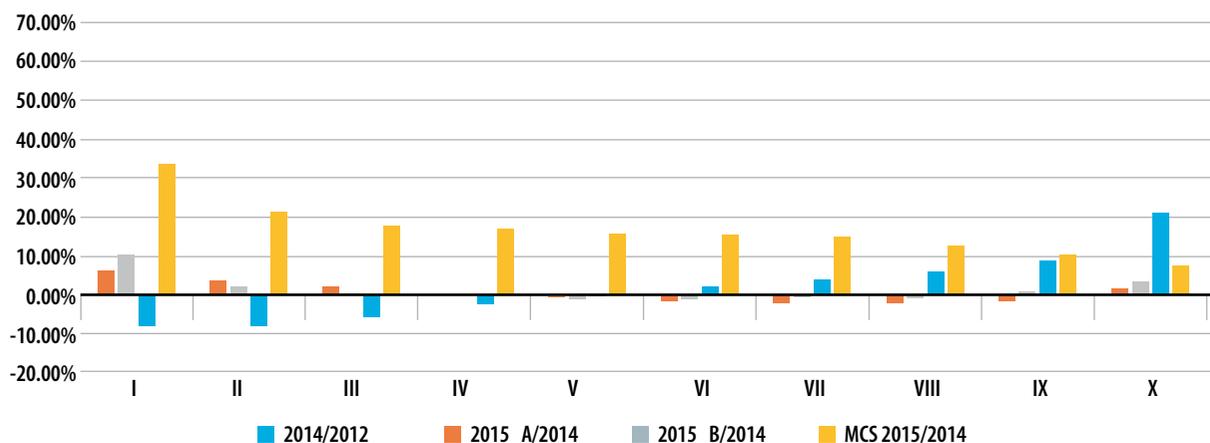
Tabla 1.b

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
Coeficiente 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



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 metodología 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.

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.

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 *decilizaciones* alternativas para las observaciones del mismo periodo. Por una parte, formamos los deciles 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.