

# 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 consiste 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.

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 Monte

\* 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

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 Socio-económicas (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);<sup>1</sup> 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 comparabilidad 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**

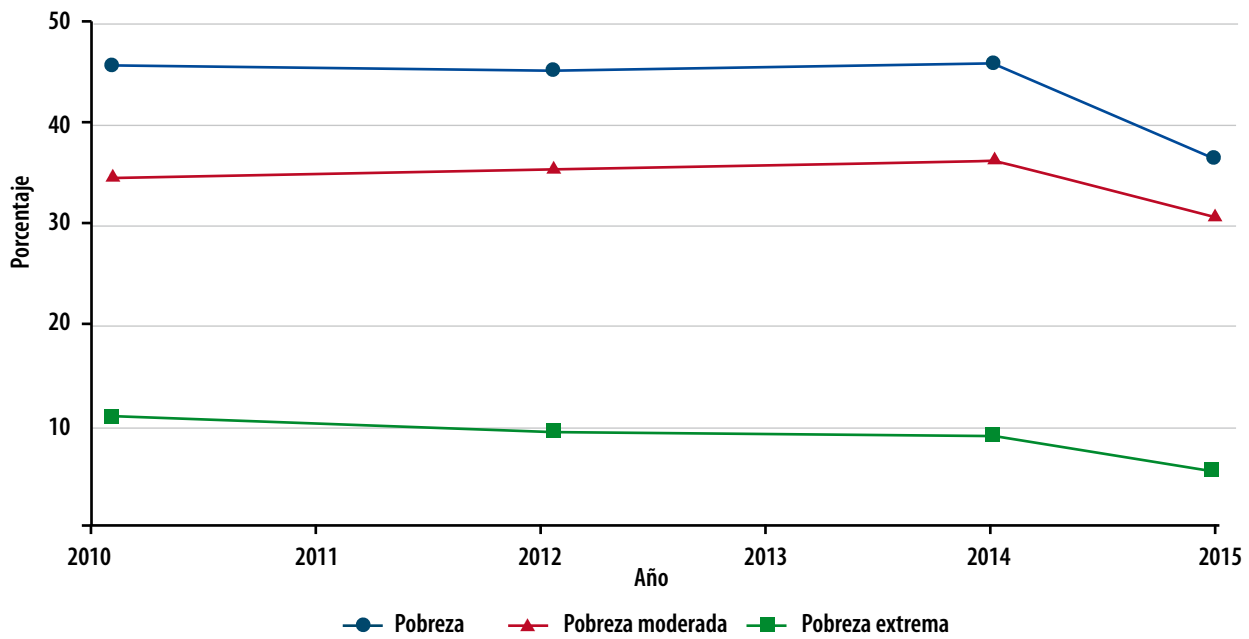
El CONEVAL utiliza el soporte normativo que le da la *Ley General de Desarrollo Social (LGDS)*, la cual utiliza dos enfoques: el de bienestar —que incluye los ingresos— y el de los derechos humanos —que incorpora las carencias sociales— (CONEVAL, 2011). La figura 1 muestra la evolución de la pobreza 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.

<sup>1</sup> 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

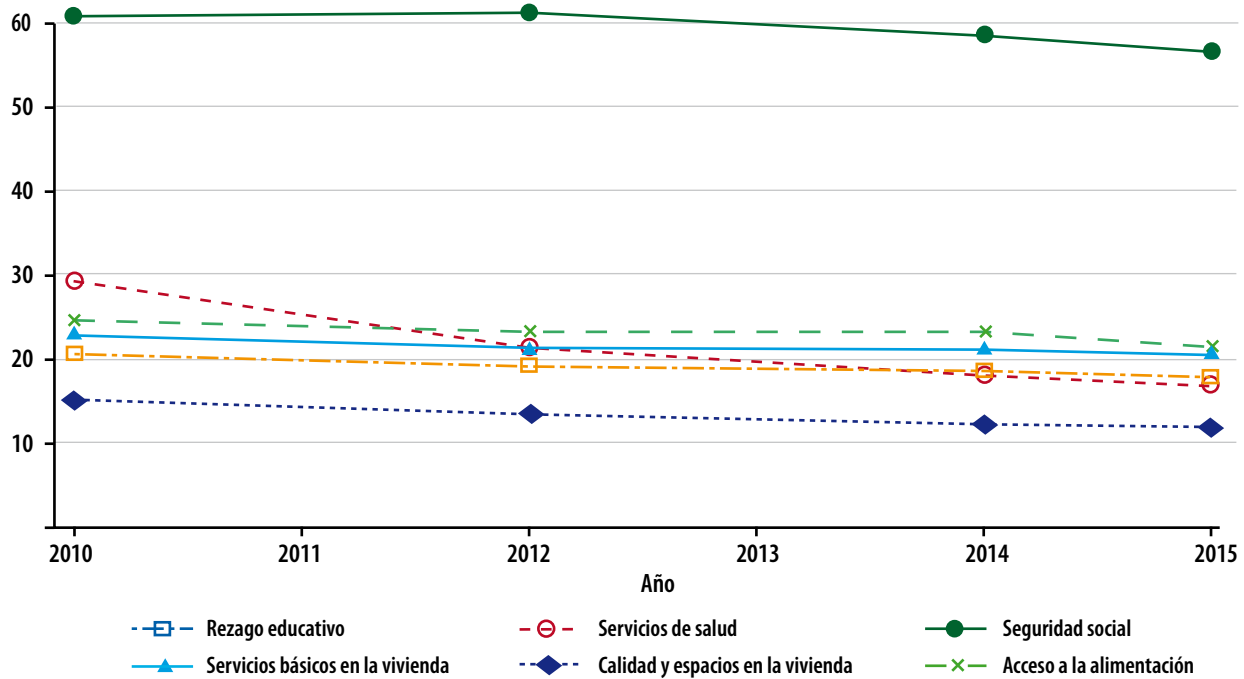
### Evolución de la pobreza en México con ingreso inferior a la línea de bienestar y línea de bienestar mínimo, 2010-2015



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 x1 a x35<sup>2</sup> 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.<sup>3</sup> 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.

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 <sup>a</sup>	2014-2015 (I-VI) <sup>b</sup>
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%

<sup>a</sup> Tasa de cambio del 2014 al 2015, datos sin corregir.

<sup>b</sup> 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.

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.

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 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  $\theta$  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  $\phi$  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  $\theta$  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

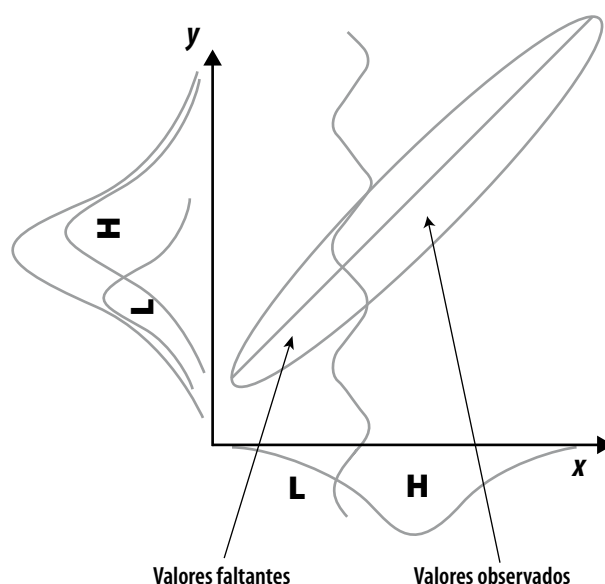
Con la finalidad de calibrar el modelo de regresión obtenido a partir de la imputación múltiple se probaron dos enfoques más. El primero fue el de la imputación vía regresión, el cual completa los datos faltantes usando un enfoque Gaussiano (Gelman, Carlin, Stern y Rubin, 2004; Rubin, 1987; Shenker y Taylor, 1996). Este método parte de la idea de que los datos están completos en



las covariables<sup>4</sup> 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.

**Figura 3**  
**Imputaciones vía regresión bajo el supuesto *MAR***



$$y_i = \alpha + \beta x_i + \zeta_i,$$

$$E(\zeta) = 0, V(\zeta) = \sigma_\zeta^2$$

$$E(x) = \mu_x, V(x) = \sigma_x^2$$

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

<sup>4</sup> 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.

datos observables. En realidad, el método FIML no produce 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.

## 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.<sup>5</sup> 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:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon_i,$$

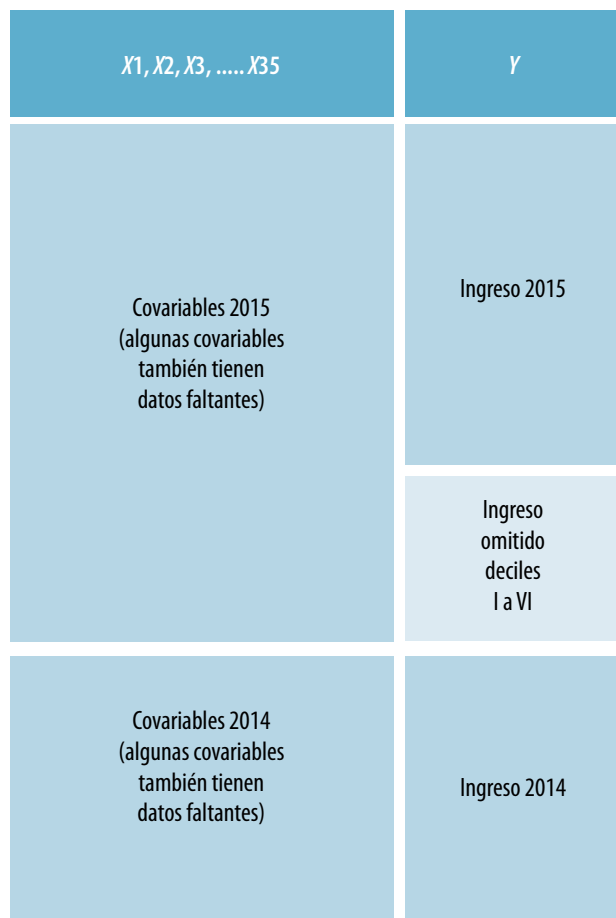
donde  $Y_i = \ln(\text{ingreso del hogar} + k)$  es el logaritmo del ingreso corriente total mensual del hogar. Las variables predictoras se muestran en el *Anexo 1*. Se busca la mejor transformación del ingreso mediante el  $\log(y + k)$  y encontramos que el valor de

<sup>5</sup> 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).

**Figura 4**  
**Esquema de preparación de archivos que incorporan los MCS 2014 y 2015**

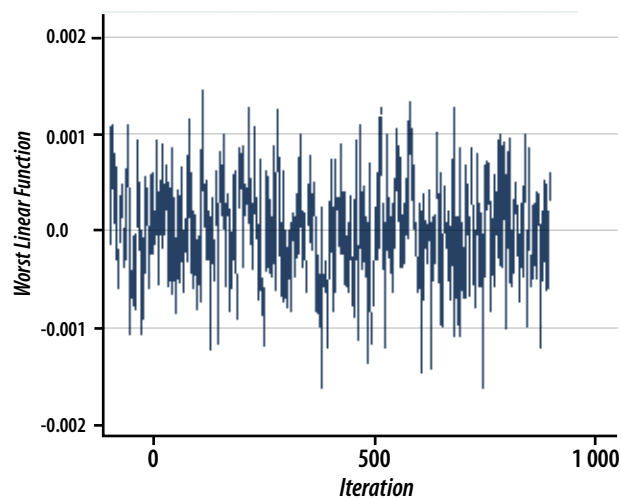


**Nota:** para el 2015, se omite el ingreso en los deciles I a VI.

**Cuadro 2**  
**Prueba de normalidad de Lilliefors (Kolmogorov-Smirnov) y medidas de dispersión**

Variable	Estadístico D	Valor de $p$	Asimetría	Curtosis
Ingreso sin transformar ( <i>ing2015</i> )	0.275	<0.001	48.17	5 606.83
$\text{Log}(\text{Ingreso}+k)$ ( <i>lning2015</i> )	0.010	<0.001	0	0.58

**Figura 5a**  
**Peor función lineal (WLF)**



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.<sup>6</sup> La garantía de que hay convergencia en las simulaciones y en la parte de imputación de los datos se juzga a partir de la peor función

<sup>6</sup> 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.

Figura 5b

### Función de autocorrelación de la WLF

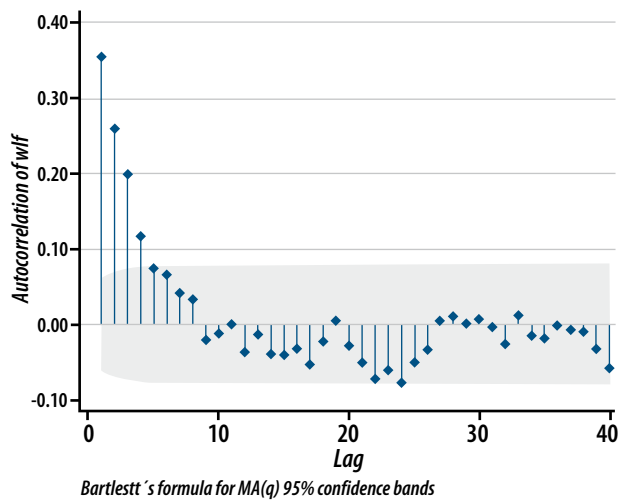
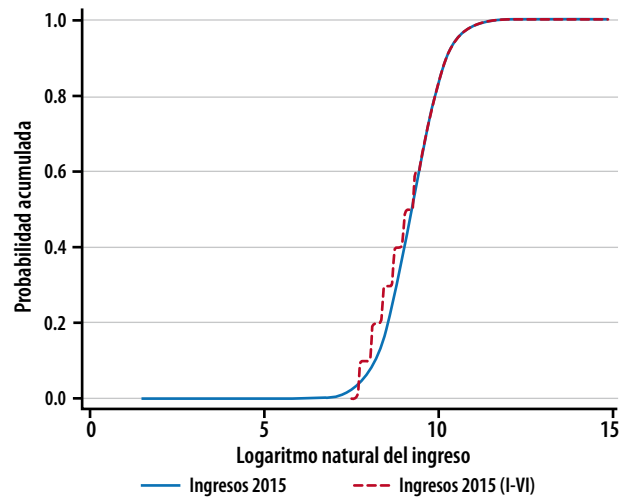


Figura 6

### 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.

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.<sup>7</sup>

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 3<sup>8</sup>) 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 tienen los resultados que a continuación se muestran en el siguiente subapartado de esta sección. La figura 6 presenta la función de distribución acumulada (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.

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 al VI.

Cuadro 3

**Ingreso promedio corriente total por deciles, Estados Unidos Mexicanos, 2010-2015<sup>a</sup>**

	2010	2012	2014	2015 <sup>b</sup>	2015 (I-IV) <sup>c</sup>	2015 (I-V) <sup>d</sup>	2015 (I-VI) <sup>e</sup>
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

<sup>a</sup> Calculado a pesos constantes del 2015.

<sup>b</sup> Valores observados no comparables.

<sup>c</sup> Se imputan los deciles I-IV.

<sup>d</sup> Se imputan los deciles I-V.

<sup>e</sup> Se imputan los deciles I-VI.

**Fuente:** estimaciones propias con base en las ENIGH-MCS 2010, 2012, 2014 y 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 corregida 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,

### Medición de la pobreza, población general, Estados Unidos Mexicanos, 2010-2015

Pobreza multidimensional	Porcentaje						Millones de personas							
	2010	2012	2014	2015 <sup>a</sup>	2015 (I-IV) <sup>b</sup>	2015 (I-V) <sup>c</sup>	2015 (I-VI) <sup>d</sup>	2010	2012	2014	2015 <sup>a</sup>	2015 (I-IV) <sup>b</sup>	2015 (I-V) <sup>c</sup>	2015 (I-VI) <sup>d</sup>
<b>Pobreza</b>														
Población en situación de pobreza	46.1	45.5	46.2	36.7	40.0	41.7	42.4	52.8	53.3	55.3	44.5	48.6	50.6	51.4
Población en situación de pobreza moderada	34.8	35.7	36.6	30.8	32.3	33.5	34.0	39.8	41.8	43.9	37.4	39.2	40.7	41.2
Población en situación de pobreza extrema	11.3	9.8	9.5	5.9	7.7	8.2	8.4	13.0	11.5	11.4	7.1	9.4	9.9	10.2
Población vulnerable por carencias sociales	28.1	28.6	26.3	33.9	33.9	33.9	33.9	32.1	33.5	31.5	41.2	41.2	41.2	41.2
Población vulnerable por ingresos	5.9	6.2	7.1	5.5	6.2	6.7	7.0	6.7	7.2	8.5	6.6	7.5	8.2	8.4
Población no pobre y no vulnerable	19.9	19.8	20.5	23.9	23.2	22.6	22.4	22.8	23.2	24.6	29.0	28.2	27.5	27.2
<b>Privación social</b>														
Población con al menos una carencia social	74.2	74.1	72.4	70.6	70.6	70.6	70.6	85.0	86.9	86.8	85.7	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	20.5	32.4	28.1	26.5	24.9	24.9	24.9	24.9
<b>Indicadores de carencia social</b>														
Rezagó educativo	20.7	19.2	18.7	17.9	17.9	17.9	17.9	23.7	22.6	22.4	21.8	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	16.9	33.5	25.3	21.8	20.5	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	56.6	69.6	71.8	70.1	68.7	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	12.0	17.4	15.9	14.8	14.6	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	20.5	26.3	24.9	25.4	24.9	24.9	24.9	24.9
Carencia por acceso a la alimentación	24.8	23.3	23.4	21.7	21.7	21.7	21.7	28.4	27.4	28.0	26.4	26.4	26.4	26.4
<b>Bienestar</b>														
Población con ingreso inferior a la línea de bienestar mínimo	19.4	20.0	20.5	12.2	18.3	19.8	20.5	22.2	23.5	24.6	14.8	22.2	24.0	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	59.6	60.6	63.8	51.1	56.1	58.8	59.9
<b>Coefficiente de Gini</b>	0.484	0.479	0.480	0.453	0.478	0.487	0.492							

<sup>a</sup> Valores observados no comparables; <sup>b</sup> se imputan los deciles I-IV; <sup>c</sup> se imputan los deciles I-V; <sup>d</sup> se imputan los deciles I-VI.  
**Fuente:** estimaciones propias con base en las ENIGH-MCS 2008, 2010, 2012, 2014 y 2015.

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* (2<sup>nd</sup> 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*. (2<sup>nd</sup> 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*. Thousand 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



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.	<i>P</i> > <i>z</i>
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_segsoc</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