INEGI
Dirección General de Integración, Análisis e Investigación
Dirección General Adjunta de Investigación

MODELO ESTADÍSTICO 2018 PARA LA CONTINUIDAD DEL MCS-ENIGH

NOTA TÉCNICA

30 DE JULIO DE 2019

INSTITUTO NACIONAL
DE ESTADÍSTICA Y GEOGRAFÍA

1. Introducción

En 2017 el INEGI dio a conocer los resultados de la Encuesta Nacional de Ingresos y Gastos de los Hogares Nueva Serie (ENIGH-NS) 2016, la cual constituye el inicio de un conjunto de levantamientos con mejoras en la recolección de información y con una mayor representatividad y precisión estadística. A la par, y con el propósito de generar los insumos necesarios en las mediciones de pobreza realizadas por el CONEVAL, el INEGI implementó el Modelo Estadístico 2016 para la continuidad del MCS-ENIGH (MEC 2016), cuyo objetivo fue ajustar el vector de ingresos para mantener su consistencia con lo observado históricamente en los levantamientos del Módulo de Condiciones Socioeconómicas de la ENIGH (MCS-ENIGH) desde 2008.

La ENIGH-NS 2018 mantiene el mismo diseño que la ENIGH-NS 2016, de manera que se garantiza la consistencia entre ambos levantamientos. Una de las implicaciones de lo anterior es que, para preservar la trayectoria de los ingresos reportados por el MCS-ENIGH de 2008 a 2014 y por el MEC 2016, los ingresos de la ENIGH-NS 2018 deben ser ajustados.

A solicitud del CONEVAL, el INEGI llevó a cabo la correspondiente reconciliación de los ingresos de la ENIGH-NS 2018. Para ello, se desarrolló un modelo estadístico que ajusta dos funciones de distribución, una de las cuales contiene restricciones que deben ser cumplidas en el proceso de modelación; estas restricciones provienen de la variación observada en el Ingreso Corriente Total (ICT) de la ENIGH-NS entre 2016 y 2018, para cada entidad federativa; es decir, los insumos necesarios para el modelado de los ingresos en 2018 se toman de la ENIGH Nueva Serie. Es importante subrayar que al mantener el mismo diseño entre 2016 y 2018, la ENIGH-NS refleja la dinámica de los ingresos de los hogares en este periodo y, por lo tanto, se convierte en la mejor información de la que disponemos para el ajuste que se realiza a través del MEC 2018.

2. Modelo Estadístico

En Ruiz y Romo (2018) se propone una metodología para la edición e imputación de datos a partir del ajuste paramétrico de dos funciones de distribución a datos empíricos; uno de los ajustes contempla la incorporación de restricciones que debe cumplir la función de distribución modelada (Bustos, 2015). Con el ajuste paramétrico de la función sin restricciones, se calcula el cuantil al que corresponde determinado valor de la variable que se quiere editar, posteriormente se genera un nuevo valor con ayuda de la función inversa de la distribución restringida y del cuantil previamente calculado.

Con el MEC 2016 se estimaron valores alternos a los del primer levantamiento de la ENIGH-NS, a efecto de recuperar la consistencia con la serie previa del MCS-ENIGH. Dado que en 2016 sólo se contaba con un levantamiento de la ENIGH-NS, fue necesario recurrir a la ENOE, ya que ésta ofrecía un referente contrafactual de la dinámica de 2014 a 2016 del ingreso de los hogares proveniente de una fuente confiable libre de cambios. Sin embargo, para el MEC 2018 existe información adicional que podemos incorporar en el proceso de modelación: al tener dos levantamientos de la ENIGH-NS (2016 y 2018) con el mismo diseño estadístico, y dado que la ENIGH es una encuesta especializada en la captación del ingreso y que, a diferencia de la ENOE, capta tanto el ingreso laboral como el no laboral, es de esperar que los cambios en el ingreso corriente total (laboral más no laboral) observados entre las ENIGH-NS reflejen más cercanamente el verdadero comportamiento de los ingresos de los hogares entre 2016 y 2018. De esta forma, con las ENIGH-NS 2016 y 2018 disponemos de un referente natural para modelar la trayectoria de ingresos ajustados del MEC, puesto que es evidente que el uso de las variaciones del ingreso que de ellas resultan son un mejor insumo para aproximar los cambios del ingreso del MEC 2016 al MEC 2018. Específicamente, la estrategia que seguimos para aproximar este comportamiento de la ENIGH-NS dentro del MEC consiste en imponer que sus microdatos modelados incorporen la variación porcentual observada de 2016 a 2018 en la mediana del ingreso corriente total por entidad federativa de la ENIGH-NS. **Es decir, el crecimiento de las medianas del ingreso corriente total entre la ENIGH-NS 2016 y la ENIGH-NS 2018, debe ser el mismo que el de las medianas del ingreso corriente total entre el MEC 2016 y el MEC 2018, para cada entidad federativa.** Una segunda restricción es incluida para controlar el carácter no finito de la función de distribución empleada en los cálculos, para lo cual se utiliza el valor del ingreso corriente total más alto observado en la ENIGH-NS 2018 para cada entidad federativa, como cota superior del ingreso corriente total mayor admisible para el MEC 2018.

De manera formal, sea $Y^i = \left(y_1^i, y_2^i, \dots, y_{n_i}^i\right)$ el vector de ingreso corriente total para la entidad *i*; sus pesos (factores de expansión) correspondientes están dados por $\boldsymbol{w}^i = (w_1^i, w_2^i \dots, w_{n_i}^i)$. Entonces, el ajuste de la distribución a los datos empíricos que corresponde a maximizar la log-verosimilitud de la función $f(\boldsymbol{Y}|\boldsymbol{\theta}) = \boldsymbol{GB2}(\boldsymbol{\theta})$, se expresa de la siguiente manera**:**

Para cada entidad i = 1,2,..,32:

$$\max\ l\left(\boldsymbol{\theta^i}\middle|\boldsymbol{Y}^i\right) = \sum_{h=1}^{n_i} w_h^i \log f\left(y_h^i\middle|\boldsymbol{\theta^i}\right)$$

Resultado de estos ajustes obtendremos 32 estimadores de los parámetros de la distribución, los cuales representamos como $\widehat{\boldsymbol{\theta}}_{SR} = (\widehat{\boldsymbol{\theta}}_{SR}^1, \widehat{\boldsymbol{\theta}}_{SR}^2, .., \widehat{\boldsymbol{\theta}}_{SR}^{32})$.

Posteriormente al ajuste obtenido, se realiza un nuevo ajuste en el que incluimos como restricción que los nuevos microdatos generen la *mediana objetivo* para la entidad federativa correspondiente. Es decir:

Sea $Y_o^i = \left(y_{(1)}^i, y_{(2)}^i, \dots, y_{(n)}^i\right)$ el vector ordenado de ingresos para la entidad *i* y sus correspondientes pesos muestrales (factores de expansión) denotados por $w_o^i = (w_{(1)}^i, w_{(2)}^i \dots, w_{(n)}^i)$, y sea el *vect*or de *medianas objetivo* $\mathbf{Me} = (\mathrm{Me}_1, \mathrm{Me}_2, .., \mathrm{Me}_{32})$. Para cada entidad *i* =1,2,..,32 se resuelve el siguiente problema de optimización restringida

$$\max\ l\left(\boldsymbol{\theta^i}\right) = \sum_{h=1}^{n_i} w_h^i \log f\left(y_h^i\middle|\boldsymbol{\theta^i}\right)$$

Sujeto a:

a) Restricciones de igualdad

1. $F^{-1}\left(F\left(y_{(k)}^{i}\middle|\widehat{\boldsymbol{\theta}}_{SR}^{i}\right) \,\middle|\, \boldsymbol{\theta}^{\boldsymbol{i}}\right) = Me_{i}$

   donde:

   $y_{(k)}^{i}$ es el Ingreso Corriente Total asociado al subíndice *k* tal que $0.5 \leq \frac{W_{(k)}}{W_{(n)}}\ y\ 0.5 \leq 1 - \frac{W_{(k-1)}}{W_{(n)}}$
   con $w_{(j)} = \sum_{h=1}^{j} w_{(h)}^{i}$

   Me_{i} es la mediana de la entidad federativa *i.*
   $\widehat{\boldsymbol{\theta}}_{SR}^{i}$ es el estimador de los parámetros de la función teórica de densidad sin restricciones para la entidad *i*.
   $F(\cdot)$ es la función de probabilidad acumulada de GB2.
   $F^{-1}(\cdot)$ es la función cuantil de GB2 y

2. $\int_{0}^{\max^{i}} f(\boldsymbol{Y}|\, \boldsymbol{\theta}^{i})dy = \ p^{max}(\widehat{\boldsymbol{\theta}}_{SR}^{i})$

   donde:

   $max^{i} = y_{(n)}^{i}$, es el valor máximo del ingreso de ICT en la entidad *i*.
   $p^{max}\left(\widehat{\boldsymbol{\theta}}_{SR}^{i}\right) = F(y_{(n)}^{i}|\widehat{\boldsymbol{\theta}}_{SR}^{i})$ es la probabilidad acumulada (bajo la función teórica sin restricciones) correspondiente al valor max en la entidad *i*.

b) *Restricciones de desigualdad*: las propias del dominio de los valores de los parámetros de la función densidad, por ejemplo, para $GB2(\mu,\sigma,\upsilon,\tau)$: $\mu,\upsilon,\tau > 0$; $-\infty < \sigma < \infty$; $-\upsilon < \frac{1}{\sigma} < \tau$.

La segunda restricción de igualdad tiene como objetivo controlar el carácter no finito en el dominio de las funciones de distribución. Para ello, tomamos el valor más grande del ICT que se obtuvo en la encuesta en cada entidad federativa y establecemos que los nuevos ingresos estimados estén en [0, max^{i}]

Derivado de los modelos optimizados, obtenemos un vector que contiene los estimadores de los parámetros restringidos de la función GB2 que denotamos como $\widehat{\boldsymbol{\theta}}_{R} = (\widehat{\boldsymbol{\theta}}_{R}^{1}, \widehat{\boldsymbol{\theta}}_{R}^{2}, .., \widehat{\boldsymbol{\theta}}_{R}^{32})$.

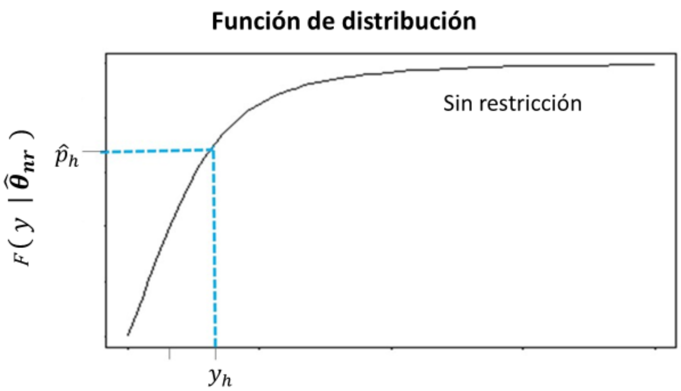
2.1 Generación de nuevos microdatos

Para encontrar el ingreso corriente total ajustado correspondiente al MEC 2018 (ICT ajustado), en cada hogar hacemos uso de los vectores $\widehat{\boldsymbol{\theta}}_{SR}^{i}$ Y $\widehat{\boldsymbol{\theta}}_{R}^{i}$ de la siguiente manera: tomamos el valor reportado del ingreso corriente total para cada hogar, y de acuerdo a la entidad federativa de pertenencia, se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_{h}^{i} = F\left(y_{h}^{i}\middle|\widehat{\boldsymbol{\theta}}_{SR}^{i}\right)\ \forall\ h = 1,2, .. n_{i}$; donde n_{i} es el número de hogares en la entidad *i*. El valor estimado del ICT ajustado (restringido) para cada hogar será entonces $\hat{y}_{h}^{i} = F^{-1}\left(\hat{p}_{h}^{i}\middle|\widehat{\boldsymbol{\theta}}_{R}^{i}\right)\ \forall\ h = 1,2, .. n_{i}$.

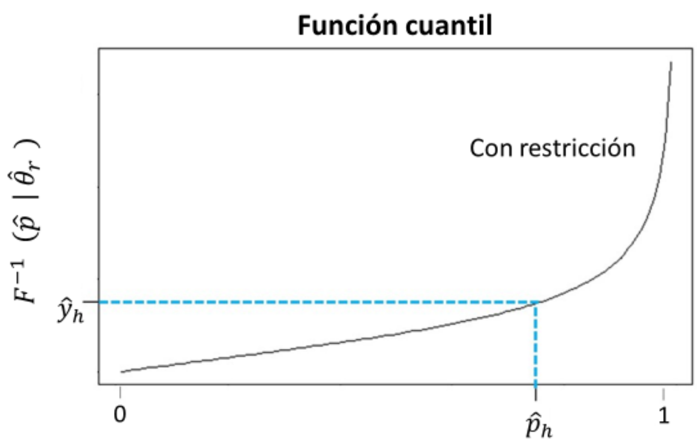
De esta manera, obtenemos para cada hogar en toda la muestra, un valor del ICT imputado que está determinado en función de los resultados del ajuste por entidad, $ICT\ ajustado_h = \hat{y}_h \ \ \forall \ \ h = 1,2,..,n$

La Figura 1 muestra esquemáticamente este proceso: la gráfica a) ejemplifica cómo a partir de un monto específico, y_h, se obtiene una probabilidad acumulada, $\hat{p}_h$, a través de la función acumulada *sin* restricciones, este proceso se repite para cada uno de los hogares de la muestra. Una vez fijadas las probabilidades acumuladas para cada hogar, y con ayuda de la función cuantil obtenida mediante el proceso de optimización *con* restricciones, se genera un nuevo valor $\hat{y}_h$, tal como se muestra en la gráfica b).

Figura 1. Visualización del proceso para la obtención de microdatos



(a)



(b)

Por último, y para distribuir las modificaciones entre cada uno de los componentes del ICT que se encuentran en las distintas tablas (insumo para la medición de la pobreza), cada componente del ICT reportado por el hogar *h* es modificado en una proporción igual a la que resulta de dividir el ICT ajustado entre el ICT original para ese hogar *h* ($p = \frac{ICT\ ajustado_h}{ICT\ original_h} = \frac{\hat{y}_h}{y_h}$).