

Modelo estadístico 2016 para la continuidad del MCS-ENIGH

2016 Statistical Model for MCS-ENIGH's Continuity

José Alejandro Ruiz Sánchez* y **Ana Miriam Romo Anaya****

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

** INEGI, miriam.romo@inegi.org.mx

Nota: los autores agradecen la colaboración de Lorenzo Cecilio Fernández en el proceso de integración del artículo.



Small figures whit money/Peter Muller/Getty Images.

En un proceso de mejora continua, el Instituto Nacional de Estadística y Geografía (INEGI) identificó áreas de oportunidad en el Módulo de Condiciones Socioeconómicas (MCS) de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH), que se materializaron en la implementación de acciones operativas en la ENIGH 2016, con el objetivo de alcanzar un mayor apego al diseño de la encuesta. Con ello, se inicia una nueva serie de la ENIGH.

Para dar continuidad a los resultados bienales de la serie del MCS-ENIGH, se propone una metodología de edición de microdatos basada en el ajuste de una función de distribución bajo restricciones. En una primera etapa, solo se modifican los ingresos por trabajo principal reportados en

la ENIGH 2016; el ajuste se realiza en función de las tasas de crecimiento observadas en la Encuesta Nacional de Ocupación y Empleo (ENOE) por entidad federativa. En la fase siguiente, se modifica el resto del ingreso total de los hogares; para ello, nos basamos en una relación constante que se ha observado desde el 2010 entre los ingresos totales y los ingresos por trabajo principal.

Como resultado de la modificación de los ingresos reportados, el crecimiento del ingreso total de los hogares entre el 2014 y los datos que arroja el modelo para el 2016 es de 2.1% (en comparación con 10.4%, si se utilizan datos de la ENIGH 2016). La variación para el primer decil es de 8.5% (mientras que con datos de la ENIGH 2016 fue de 29.9%).

Palabras clave: MCS-ENIGH; ENIGH 2016; imputación.

As a way to keep improving, INEGI identified some opportunity areas in the MCS-ENIGH, which resulted in the implementation of certain operative actions in the ENIGH 2016. This was aimed to get as close as possible to the survey's design. This is the start of a new ENIGH series.

We propose a new micro-data editing methodology in order to keep the continuous biennial results of the MCS-ENIGH series. This methodology is based on the adjustment of restricted distribution functions. On a first stage, only incomes per main labor reported in ENIGH 2016 are modified, with the adjustment done according to the growth rates observed in ENOE by Federal entity. In the next stage, we modified the rest of households' total income based on a constant relation observed since 2010 between total incomes and incomes derived from a person's main labor.

As a result of the modifications of reported incomes, households' total income growth, between 2014 and the data we get from the model designed for 2016, is of 2.1% (unlike the 10.4% we get when compared to data from ENIGH 2016 without adjustments). The variation for the first decile is of 8.5% (it was of 29.9% with data from ENIGH 2016 without adjustments).

Key words: MCS-ENIGH; ENIGH 2016; imputation.

1. Introducción

En un proceso continuo de mejora, el INEGI identificó áreas de oportunidad alrededor de la capacitación, supervisión y control operativo del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH). A partir de ello, se llevaron a cabo acciones de fortalecimiento operativo en el levantamiento de la ENIGH 2016 en un esfuerzo por alcanzar un mayor apego al diseño de la encuesta.

Con estas medidas implementadas, se inicia una nueva serie histórica para las encuestas de ingresos y gastos de los hogares. Sin embargo, y partiendo del reconocimiento de la necesidad de preservar la continuidad de sus ejercicios estadísticos, se presenta esta propuesta metodológica, cuyo objetivo es proveer de los insumos necesarios para la medición de la pobreza multidimensional, de tal manera que se mantenga la consistencia con los resultados de la serie bienal 2008-2014, tanto a nivel nacional como por entidad federativa.

En la medición de la pobreza multidimensional se incorporan variables no monetarias como la condición de acceso a salud, educación, seguridad social, etc., pero también otras que reflejan el ingreso de los hogares. Este documento está centrado solo en la modificación de este último grupo. En México, el encargado de medir la pobreza es el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL), cuya metodología toma como insumo la base de datos del MCS-ENIGH. Por lo tanto, una parte fundamental del proceso de ajuste consistirá en el desarrollo de un mecanismo que permita transitar de relaciones agregadas a nuevos microdatos de ingresos que sustituyan a los originalmente reportados.

El ingreso total de los hogares proviene de múltiples fuentes, por ejemplo: salarios, ganancias de empresas, recursos que reciben de otros hogares, etcétera. La ENIGH captura esa heterogeneidad a través de una serie de reactivos que son recabados a nivel individuo. La gran cantidad de preguntas referidas a los distintos canales de ingreso, entre otras características, hacen que la ENIGH tenga elementos que no poseen otros operativos similares levantados en el país.

A pesar de la diversidad de fuentes, la mayor parte del ingreso de los mexicanos proviene de su ingreso laboral, es decir, de su trabajo como asalariado y/o independiente (principal y secundario) que, en conjunto, ha representado cerca de 70% del ingreso corriente total (ICT) desde el 2010. Este elevado porcentaje, junto con la existencia de un levantamiento regular sobre condiciones laborales, como lo es la Encuesta Nacional de Ocupación y Empleo (ENOE), crean una ventana de oportunidad para utilizar el ingreso laboral que reporta esta encuesta como ancla para ajustar los ingresos laborales.¹

Para la modificación de los ingresos laborales reportados por los hogares, aproximamos una función de densidad de probabilidad a los datos empíricos del ingreso de la ENIGH 2016; posteriormente, se ajusta otra función de densidad pero ahora imponiéndole como condición que sea capaz de generar nuevos microdatos tales que arrojen un valor objetivo de un estadístico específico (por ejemplo, un determinado promedio muestral); en nuestro caso, ese valor proviene de la trayectoria para cada entidad federativa de los ingresos laborales de la ENOE.

Al modificar solo el ingreso laboral, la relación entre éste y el ingreso no laboral se ve alterada, rompiendo una trayectoria histórica que se venía observando desde el 2010. Para recomponer la relación, y en una segunda fase, se modifica el ingreso no laboral, de tal forma que se elimine el efecto colateral causado por el primer ajuste. Así, esta metodología modifica solo las variables de ingreso, dejando inalteradas las variables no-ingreso y los factores de expansión; esto último permite mantener el diseño muestral de la ENIGH.

El documento se encuentra dividido en siete secciones. En la segunda se presentan algunos estadísticos descriptivos que permiten conocer la magnitud y distribución de los efectos de las mejoras. En la tercera se desarrolla el marco teórico general de la metodología para modificar microdatos. En la siguiente se crean nuevos microdatos del ingreso por trabajo principal (ITP). En la quinta se presenta la segunda y última fase del ajuste a los ingresos de los hogares, relacionada con la modificación de los ingresos distintos al ITP (*ITP^c*). En la penúltima se muestran los principales resultados del ajuste y en la última, se concluye.

¹ Salvo en el primer trimestre del año, el ingreso laboral de la ENOE hace referencia al correspondiente originado por el trabajo principal.

Cuadro 2.1

Ingreso corriente promedio por decil (precios del 2016)

	2010	2012	2014	2016
Nacional	43 931	42 115	42 157	46 521
I	6 111	5 920	6 288	8 166
II	11 707	11 310	11 731	14 206
III	16 285	15 750	16 062	18 918
IV	20 876	20 326	20 336	23 556
V	26 098	25 357	25 258	28 813
VI	32 400	31 502	31 030	34 837
VII	40 352	39 343	38 489	42 431
VIII	52 028	50 494	49 463	53 383
IX	72 580	69 321	68 386	72 041
X	160 872	151 828	154 524	168 856

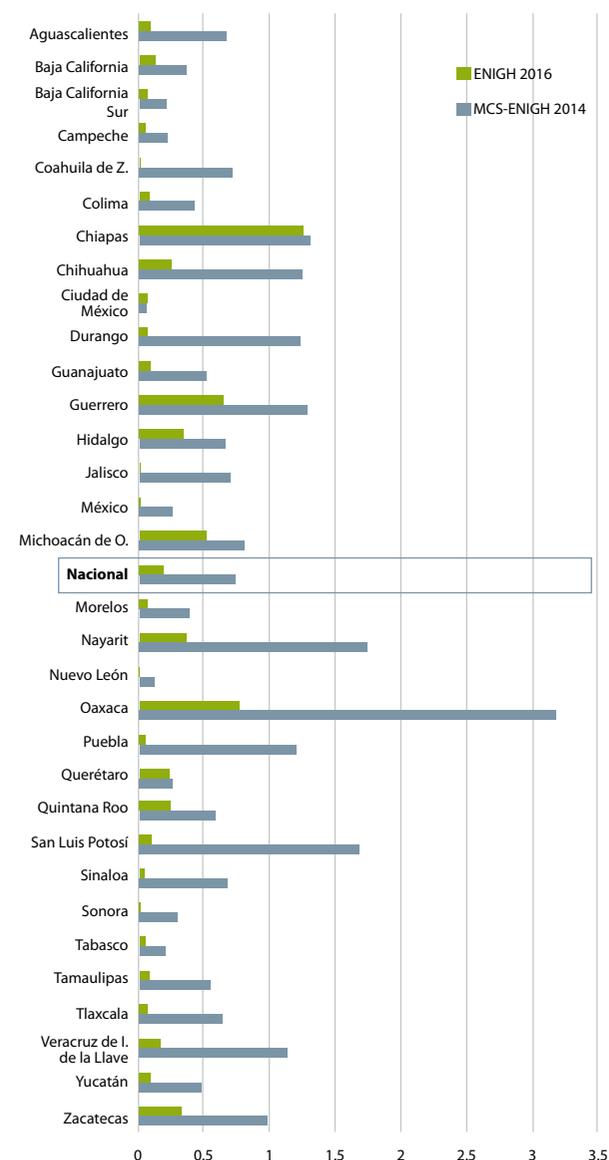
Cuadro 2.2

Variación porcentual del ICT por decil (promedio, precios del 2016)

	2010-2012	2012-2014	2014-2016
Nacional	-4.1%	0.1%	10.4%
I	-3.1%	6.2%	29.9%
II	-3.4%	3.7%	21.1%
III	-3.3%	2.0%	17.8%
IV	-2.6%	0.1%	15.8%
V	-2.8%	-0.4%	14.1%
VI	-2.8%	-1.5%	12.3%
VII	-2.5%	-2.2%	10.2%
VIII	-2.9%	-2.0%	7.9%
IX	-4.5%	-1.3%	5.3%
X	-5.6%	1.8%	9.3%

Figura 2.1

Porcentaje de hogares con ingresos inferiores a 3 mil pesos trimestrales (precios del 2016)



2. Análisis descriptivo

Como consecuencia de las acciones tomadas, el ICT que reportan los hogares en la ENIGH 2016 tiene incrementos sustanciales con respecto al 2014, en especial para los primeros deciles (ver cuadros 2.1 y 2.2).² Lo anterior se debe, en parte, a que ahora el problema de la subdeclaración que hacen los hogares es menor; ello se ve reflejado en el porcentaje que indicó recibir menos de 3 mil pesos trimestrales, el cual pasó de 0.7% en el 2014 a 0.2% en el 2016 a nivel nacional (ver figura 2.1). A nivel entidad federativa, los porcentajes muestran la misma tendencia decreciente (excepto en la Ciudad de México); incluso, para algunas entidades como Coahuila de Zaragoza, Jalisco, México, Nuevo León y Sonora, el porcentaje es cercano a cero.

Los cambios en tendencia mostrados en los cuadros anteriores son el resultado de las mejoras implementadas en el 2016, por lo que la información de la ENIGH 2016 no es consistente con la de años anteriores.

Por otra parte, y debido a que las instrucciones para una mejor recolección de la información se dieron a nivel nacional, la declaración de ingresos tuvo un efecto en las distintas entidades federativas, aunque de manera diferenciada. La figura 2.2 permite identificar regiones de acuerdo con su variación porcentual en el ingreso por hogar entre el 2014 y 2016. Observamos que, tanto en volumen como en promedio, la zona norte del país fue la que experimentó las mayores variaciones con respecto al 2014, llegando a niveles superiores a 20% en entidades como Nuevo León y Chihuahua.

Otro nivel de desagregación que se puede lograr con la información de la ENIGH es por fuentes de ingreso, es decir, aquellos rubros que componen el ingreso total de los hogares. Es deseable analizar el comportamiento con esta división ya que nos permite identificar y relativizar los cambios más sustanciales dependiendo de la procedencia de los ingresos. El cuadro 2.3 contiene el promedio nacional de cada una de las principales fuentes de ingreso, así como el peso que tienen con respecto al ingreso corriente total. Si bien el ingreso del trabajo no tuvo los incrementos de otras fuentes de ingreso, su peso dentro del ICT hace que sea un componente fundamental para recuperar la trayectoria histórica. Históricamente, la fuente de mayor importancia dentro del ICT son los ingresos del trabajo (principal y secundario), este componente ha representado poco menos de 70% del total.

Un aspecto a resaltar es que el comportamiento de estas dos fuentes de ingreso es distinto: los derivados por trabajo crecieron 2.9% entre el 2014 y 2016, mientras que el restante creció 26.9% en el mismo periodo (ver cuadro 2.4).

En términos del monto total de los ingresos por trabajo, la mayor variación porcentual con respecto al levantamiento previo se observó entre el 2014 y 2016 cuando aumentó 7.1% (a precios del 2016) (ver figura 2.3). Ello, en combinación con el cuadro 2.4, implica que el número de personas con ingresos laborales también tuvo un crecimiento sustancial.

Por otra parte, la gran diversidad de fuentes de ingresos captadas en la ENIGH no permite tener una contraparte para cada una de ellas, ya sea en encuestas o en agregados económicos;

² A menos que se especifique lo contrario, las cifras reportadas para años anteriores al 2016 son referidas a los datos conjuntos del MCS-ENIGH.

sin embargo, sí existe un levantamiento paralelo para el ingreso por trabajo: la ENOE; ésta, al igual que el MCS-ENIGH, tiene representatividad tanto a nivel nacional como por entidad federativa y ambas comparten el mismo marco muestral. La ENOE se levanta desde el 2005 con una periodicidad trimestral, cuya muestra rotatoria es de más de 120 mil viviendas; cada trimestre se sustituye 20% de las viviendas y cada vivienda es entrevistada durante cinco ocasiones antes de ser reemplazada; esto permite que en cada trimestre se mantenga 80% de la muestra. Otra ventaja de la ENOE es que no fue objeto de las mejoras implementadas en la ENIGH 2016. El objetivo de la ENOE es obtener información sobre las características ocupacionales de la población, por lo que no obtiene información de todas las fuentes de ingresos de los hogares. Un problema bien identificado en la ENOE es el elevado número de personas que no reportan ingresos; el porcentaje es cercano a 60% del total de individuos a los cuales se les aplicó el *Cuestionario de ocupación y empleo*, en distintos levantamientos desde el 2010 a la fecha. Para aminorar esta situación, el cuestionario agrega un reactivo de rescate que pregunta por un rango salarial a aquellas personas que, aun cuando se les preguntó el monto de sus ingresos, no contestaron un ingreso puntual. Para incorporar las respuestas por rangos salariales, se tienen diferentes opciones, la que usamos en esta metodología es la misma que propone el CONEVAL en la construcción del índice de la tendencia laboral de la pobreza y que se encuentra detallada en el anexo del documento publicado en internet (http://www.coneval.org.mx/Informes/ITLP-IS/Nota%20ingresos%20e%20ITLP_IS.pdf).

La información que reporta el MCS-ENIGH sobre los ingresos de los hogares es de hasta seis meses anteriores a la fecha del levantamiento y, debido a que la obtención de la información es de agosto a noviembre, el periodo reportado en la encuesta es de febrero a octubre (de estos nueve meses, cada hogar solo reporta seis, dependiendo del mes en el que es encuestado). Por otra parte, y dado que la realización de la ENOE es trimestral, aquellos periodos donde hay mayor coincidencia con la información del MCS-ENIGH son el segundo y el tercer trimestre; es por ello que, para los datos obtenidos de la ENOE, consideramos dichos periodos. La desventaja que esto conlleva es que en esos trimestres solo se recoge información del trabajo principal y se excluye lo correspondiente al secundario.

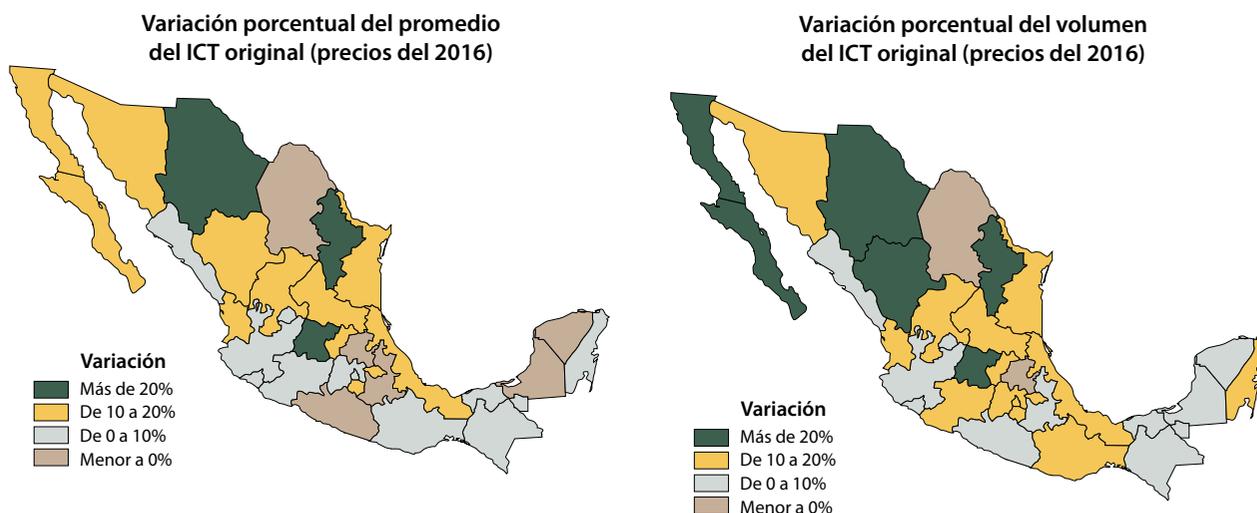
Como consecuencia de lo anterior, debemos identificar los ingresos obtenidos por el trabajo principal dentro del MCS-ENIGH y la ENIGH 2016 para hacerlos comparables con los de la ENOE. El MCS-ENIGH y la ENIGH 2016 generan información a nivel individuo y por tipo de ingreso recibido; este ingreso se encuentra clasificado por códigos y se pueden consultar en el documento metodológico de la encuesta. Los códigos que utilizaremos para generar la variable de ITP son los que se muestran en el cuadro 2.5 que, para fines prácticos, se pueden agrupar en ingresos por trabajo subordinado (columna a) y por trabajo independiente (columna b).

Para terminar de formar la variable ITP, tanto en el MCS-ENIGH como en la ENIGH 2016 y en la ENOE, sumamos los ingresos a nivel hogar. Para la ENOE se mantienen solo aquellos hogares en los que todos sus integrantes declaran algún ingreso (incluso si esa declaración es cero).

Como proporción del ICT, el ITP ha sido relativamente estable (61.5% en promedio) durante el periodo del 2010 al 2014, tal como se muestra en la figura 2.4. Dicha estabilidad será importante en nuestra metodología, como se explicará más adelante.

Figura 2.2

Variación porcentual del ICT del 2014 al 2016 (precios del 2016)



Cuadro 2.3

Promedio por hogar para distintas fuentes de ingreso y su peso relativo dentro del total (precios del 2016)

	MCS-ENIGH 2010	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016
Promedio ingreso corriente	43 931 100%	42 115 100%	42 157 100%	46 521 100%
Ingreso del trabajo	30 109 69%	29 370 70%	29 059 69%	29 906 64%
Ingreso distinto del trabajo	13 822 31%	12 745 30%	13 098 31%	16 615 36%

Cuadro 2.4

Variación porcentual por fuente de ingreso (promedio de los hogares)

	2010 al 2012	2012 al 2014	2014 al 2016
Ingreso corriente total	-4.1%	0.1%	10.4%
Ingreso del trabajo	-2.5%	-1.1%	2.9%
Ingreso distinto del trabajo	-7.8%	2.8%	26.9%

Apartados que forman la variable de ITP del MCS-ENIGH

a)

Código	Descripción
	Ingresos monetarios del trabajo principal para subordinados
P001	Sueldos, salarios o jornal
P002	Destajo
P003	Comisiones y propinas
P004	Horas extra
P005	Incentivos, gratificaciones o premios
P006	Bono, percepción adicional o sobresueldo
P007	Primas vacacionales y otras prestaciones en dinero

b)

Código	Descripción
	Ingresos por negocios del hogar, trabajo principal
P011	Sueldos o salarios
P012	Ganancias/utilidades
P013	Otros ingresos
	Ingresos por negocio propio, trabajo principal
P068	Por negocios con tipo de actividad industrial
P069	Por negocios con tipo de actividad comercial
P070	Por negocios prestadores de servicios
P071	Por negocios con actividades agrícolas
P072	Por negocios con actividades de cría y explotación de animales
P073	Por negocios con actividades de recolección, reforestación y tala de árboles
P074	Por negocios con actividades de pesca, caza y captura de animales

Figura 2.4

Proporción del ITP dentro del ICT en la ENIGH 2016

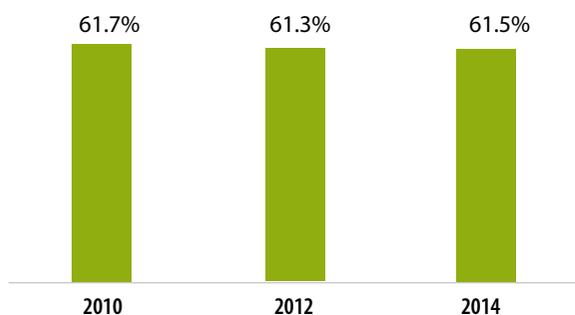
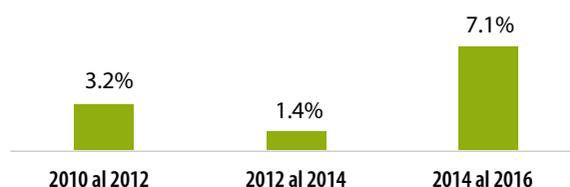


Figura 2.3

Variación porcentual del ingreso por trabajo (volumen, precios del 2016)

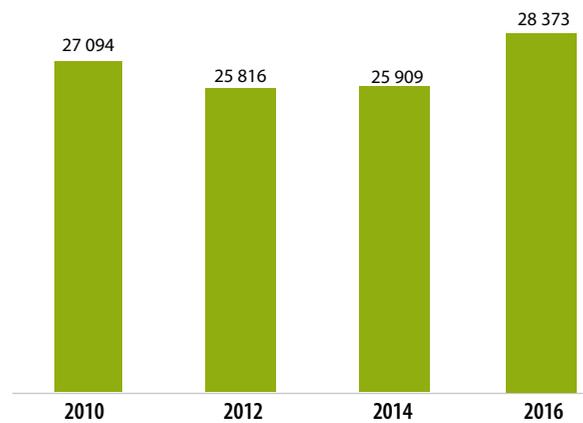


La figura 2.5 muestra el promedio nacional del ITP: en ella se observa una caída entre el 2010 y 2012 de más de 1 200 pesos; también se aprecia que el cambio más importante se dio entre el 2014 y 2016, con un crecimiento superior a los 2 000 pesos.

A grandes rasgos, la metodología usa la tasa de crecimiento del ITP_ENOE (ingreso por trabajo principal de la ENOE) dentro de cada entidad federativa, para ajustar el valor del ITP_ENIGH (ingreso por trabajo principal de la ENIGH) reportado por los hogares. Sin embargo, al ajustar solo el ITP, la proporción entre éste y el ICT se verá alterada. Si suponemos que esa relación debe ser similar a la observada históricamente (ver figura 2.4), ello nos permitiría tener un contrafactual para modificar el complemento del ITP (ITP^c). Así, el ajuste al ICT sería en dos fases: en la primera se modifica el ITP con base en el comportamiento de la ENOE, y en la segunda se modifica el ITP^c con base en el peso que a través del tiempo ha tenido el ITP dentro del ICT. En las siguientes secciones se detalla el proceso.

Figura 2.5

ITP promedio del MCS-ENIGH por hogar (montos trimestrales, precios del 2016)



3. Metodología

La metodología toma como marco teórico el trabajo desarrollado por Bustos (2015). En su artículo, el autor propone que, además de aproximar una función de distribución a los datos que arrojan encuestas como la ENIGH, se tomen datos puntuales de fuentes como las cuentas nacionales para imponerlos como restricciones al ajuste de la distribución teórica y, así, corregir problemas como el de la subdeclaración. De esta forma, se tendrían estimaciones más cercanas a la realidad sobre la distribución del ingreso de los hogares.

Comencemos suponiendo que la variable de interés se puede modelar a través de una función de densidad de probabilidad paramétrica. Es decir, si dicha variable la denotamos como y , entonces $f(y; \theta)$ es la función de densidad asociada a su función de distribución $F(y|\theta)$. Los valores del vector de parámetros $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ son desconocidos y deben estimarse procurando errores mínimos de medición, de tal manera que se pueda reconstruir lo mejor posible la *distribución empírica* (aquella obtenida a partir de datos recopilados por medio de una encuesta, por ejemplo).

Recordemos que una *función de distribución acumulada teórica* representa las probabilidades acumuladas obtenidas de la función de densidad. Es denotada por $F_Y(y|\theta)$ y se interpreta como la probabilidad de que la variable Y asuma un valor menor o igual a y . Se define como:

$$F_Y(y|\theta) = P(Y \leq y) = \int_0^y f_Y(t|\theta) dt = p \quad (1)$$

Así, para un cierto intervalo de ingresos $[a, b]$, tenemos que:

$$P(a \leq y \leq b) = \int_a^b f_Y(y|\theta) dy = F_Y(b) - F_Y(a) = p_{a-b} \quad (2)$$

Debido a que la función $F(y|\theta)$ es estrictamente creciente, entonces existe su inversa, la cual es llamada *función cuantil*. Ésta es la que usaremos para obtener los microdatos estimados.

$$F^{-1}(p) = \min\{y : F_Y(y) \geq p\}; p \in (0,1) \quad (3)$$

Por otro lado, la *distribución acumulada empírica* es la frecuencia acumulada de los datos observados ordenados de forma ascendente; en nuestro caso, y como los datos provienen de un diseño de muestreo, usamos los factores de expansión para definir dicha frecuencia como la proporción de elementos (hogares) observados que tienen un valor menor o igual a un valor observado y . Se representan por la siguiente expresión:

$$P_{m_acum} = \frac{W_{m_acum}}{\sum_{h=1}^n W_h}; W_{m_acum} = \sum_{h=1}^m W_h$$

$$\text{para cada } m = 1, 2, \dots, n; \quad (4)$$

donde w_h es el factor de expansión asociado al hogar h y n es el número total de hogares en la muestra.

Ahora bien, una vez elegida la forma funcional paramétrica $f(y; \theta)$ lo siguiente es establecer el método de optimización que se usará para encontrar el estimador del vector de parámetros θ . Existen diversos métodos que cumplen con esta finalidad; entre los más comunes se encuentran el de momentos, el de mínimos cuadrados y el de máxima verosimilitud. Decidimos usar este último ya que, además de arrojar estimaciones con propiedades de eficiencia cuando se cumplen ciertas condiciones, hace máxima la probabilidad de que el modelo —dado por $f(y; \theta)$ — genere la muestra que se observó; también, ya hay un gran avance en la implementación computacional del algoritmo que resuelven de manera eficiente sistemas de ecuaciones no lineales.

El método consiste en tomar los n datos muestrales de la variable de interés, representada por el vector $Y = (y_1, y_2, \dots, y_n)$, y suponer que esta variable tiene una función de distribución asociada $f(y|\theta)$; el estimador de θ (el de máxima verosimilitud) se obtiene calculando el máximo³ de la función pseudo-*log*-verosimilitud para datos con pesos (Skinner *et al.*, 1989) denotada por $l(\theta)$ y definida como:

$$l(\theta) = \log L(\theta|Y) = \sum_{h=1}^n w_h \log f(y_h|\theta) \quad (5)$$

donde w_h corresponde al peso de la observación (hogar) h , con $h = 1, \dots, n$.

El valor de θ que maximiza la función anterior es llamado estimador de θ y lo denotaremos por $\hat{\theta}_{SR}$ (el subíndice hace referencia a que es el estimador de un ajuste sin restricciones).

³ La mayoría de usuarios que elige este método de estimación prefiere usar la siguiente regla de equivalencia: maximizar una función es matemáticamente equivalente a minimizar la expresión negativa de la función objetivo, es decir, $\text{Max } [l(\theta)] = \text{Min } [-l(\theta)]$; la preferencia radica en que, computacionalmente, los algoritmos están adaptados para la minimización de funciones.

Con respecto al modelo de ajuste con restricciones, éste puede ser expresado de la siguiente manera:

$$\text{Max } l(\theta) = \sum_{i=1}^n w_i \log f(y_i|\theta) \quad (6)$$

Sujeta a:

Restricciones de igualdad	y/o	Restricciones de desigualdad
$H_1(y, \theta) = c_1$		$I_1(y, \theta) \leq d_1$
$H_2(y, \theta) = c_2$		$I_2(y, \theta) \leq d_2$
...		...
$H_p(y, \theta) = c_p$		$I_p(y, \theta) \leq d_p$

Con respecto al modelo de ajuste con restricciones, éste puede ser expresado de la siguiente manera:

$$\text{Max } l(\theta) = \sum_{i=1}^n w_i \log f(y_i|\theta) \quad (6)$$

Sujeta a:

Restricciones de igualdad	y/o	Restricciones de desigualdad
$H_1(y, \theta) = c_1$		$I_1(y, \theta) \leq d_1$
$H_2(y, \theta) = c_2$		$I_2(y, \theta) \leq d_2$
...		...
$H_p(y, \theta) = c_p$		$I_p(y, \theta) \leq d_p$

siempre y cuando $p < \text{número de parámetros a estimar}$.⁴ c_i, d_i son constantes (valores objetivos).

De manera iterativa se encuentra el valor de θ que arroje un valor máximo de $l(\theta)$ y que cumpla con las restricciones impuestas. Es éste al que consideramos el mejor estimador y lo denotaremos como $\hat{\theta}_R$.

Las funciones $H_i(\cdot), I_i(\cdot)$, pueden estar en términos de la variable de interés y/o de los valores del vector-parámetro θ . Así, por ejemplo, si la intención es que la variable Y en un cierto rango $[0, c]$ tenga una probabilidad no mayor que c_1 , entonces, una de las restricciones tomaría la siguiente forma:

$$H_1(y, \theta) = \int_{y=0}^{y=c} f(y|\theta) dy = c_1$$

Si el objetivo es que la distribución teórica que ajusta a Y tenga una esperanza (promedio) igual a c_2 , entonces, una forma de expresar la función de restricción es:

⁴ En caso contrario, el modelo expresado en (6) se convierte en la solución de un sistema de ecuaciones sin la dependencia de los valores observados de la variable aleatoria Y .

$$H_2(y, \theta) = \int_{y=0}^{\infty} yf(y|\theta)dy = c_2$$

Por otro lado, es posible a través de la especificación del modelo con restricciones, modificar recursivamente los microdatos muestrales (es decir, los datos de la encuesta) hasta obtener de ellos valores deseados de un estadístico muestral; así, por ejemplo, si deseamos que los nuevos microdatos (junto con sus factores de expansión w_h) tenga un promedio igual a c_3 , entonces la restricción de igualdad tendría la siguiente forma:

$$H_3(y, \theta) = \sum_{h=1}^n \frac{w_h}{N} \hat{y}_{h|\theta} = c_3$$

donde $\hat{y}_{h|\theta}$ es el nuevo valor para la observación h , dados los valores del parámetro θ (en la sección 3.1 se propone una forma de obtener $\hat{y}_{h|\theta}$).

Éstos son solo algunos ejemplos de la forma que pueden tener las funciones que definen las restricciones; pueden variar según sea el propósito del modelo y, como veremos enseguida, es factible que a través del modelo restringido se puedan editar los valores de los microdatos.

3.1 Generación de nuevos microdatos

Bajo el supuesto de que $Y \sim f(y|\theta)$, y dados los estimadores $\hat{\theta}_{SR}$ y $\hat{\theta}_R$, la forma de obtener los nuevos valores empíricos es la siguiente: se calcula la probabilidad acumulada, $\hat{p}_h = F(y_h|\hat{\theta}_{SR})$, que le correspondería al hogar muestral h de acuerdo con el valor y_h reportado; para ello, utilizamos la función de distribución acumulada teórica definida en (1). Una vez obtenida \hat{p}_h para cada hogar de la muestra, nos olvidamos de la función de probabilidad sin restricciones y usamos la función inversa de la función acumulada con restricciones (determinada por el vector $\hat{\theta}_R$); es decir, la *función cuantil* especificada en (3). Esta función toma un valor \hat{p}_h y arroja un valor de \hat{y}_h , $F^{-1}(\hat{p}_h|\hat{\theta}_R) = \hat{y}_h$; de esta forma obtenemos, para cada hogar muestral, un nuevo valor de la variable y a través de una función de probabilidad con restricciones.

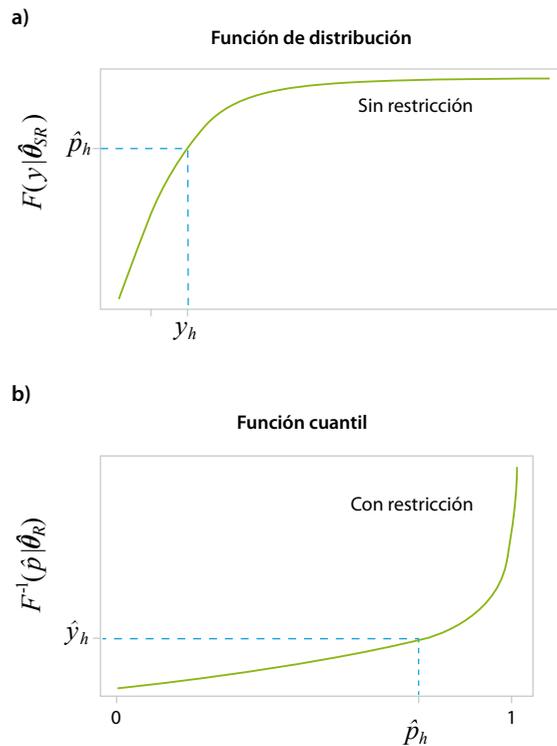
La figura 3.1 muestra de manera esquemática este proceso: la gráfica a) ejemplifica cómo a partir de un monto específico, y_h , se obtiene una probabilidad acumulada, \hat{p}_h , a través de la función acumulada *sin* restricciones; este proceso se repite para cada uno de los hogares de la muestra. Una vez fijadas las probabilidades acumuladas para cada hogar, y con ayuda de la *función cuantil* obtenida mediante el proceso de optimización *con* restricciones, se genera un nuevo valor \hat{y}_h , tal como se muestra en la gráfica b).

4. Ajuste al ingreso por trabajo principal

El ajuste al ingreso total de los hogares se realizará en dos etapas secuenciales, primero se modificarán los ingresos correspondientes al ITP y, después, los que forman el ICT, pero que son distintos al ITP (a los cuales llamamos complemento del ITP, ITP^c).

Figura 3.1

Visualización del proceso para la obtención de microdatos



Con base en la metodología desarrollada en el apartado anterior, en esta sección se detalla la modificación de microdatos relacionados con el ITP. El modelo requiere de un valor objetivo que debe alcanzar un estadístico deseado que, en nuestro caso, será la mediana del ICT de la ENIGH 2016. Para generar el valor objetivo que se debe alcanzar, se siguen estos pasos: en cada entidad federativa por separado se toman las medianas del ingreso reportado dentro del segundo y tercer trimestres de la ENOE para el 2014 y 2016. Con ellas se obtienen los promedios entre los dos trimestres del mismo año. Posteriormente, se calculan las variaciones porcentuales del promedio de las medianas del 2016 con respecto al del 2014. De esta manera, tenemos 32 tasas de crecimiento de las medianas entre el 2014 y 2016. Enseguida, y para cada entidad federativa, se calculan las medianas del ICT del MCS-ENIGH 2014; a éstas les aplicamos la tasa de crecimiento obtenida con los datos de la ENOE para, finalmente, obtener una mediana objetivo para cada entidad en el 2016 que denotaremos como $Me = (Me_1, Me_2, \dots, Me_{32})$.

Una vez que hemos determinado tanto el estadístico deseado como su valor objetivo a alcanzar generamos nuevos valores del ICT para cada hogar, de acuerdo con la metodología de la sección anterior. Sin embargo, y para lograr ajustes diferenciados entre fuentes de ingreso, los valores originales del ICT y los nuevos valores son usados solo para modificar el ITP (como más adelante se explica), mientras que el ITP^c se modificará en función de la relación histórica entre el ITP y el ICT (ver figura 2.4).

Una de las primeras tareas a resolver es la elección de la forma funcional que mejor se ajusta a los ingresos observados. En el *Anexo 1* se especifican las formas funcionales de cuatro funciones paramétricas⁵ elegidas para ser sometidas a pruebas de ajuste: *Gama Generalizada (GG)*, *Beta Generalizada del segundo tipo (GB2)*, *LogNormal (LN)* y *Dagum*. En el *Anexo 2* se incluyen los resultados de los estadísticos de ajuste a partir de los cuales elegimos la función de distribución con la que trabajaremos. Con base en dichos criterios, se determinó que la *GB2* con cuatro parámetros (μ, σ, v, τ) es la que mejor se ajusta a los ingresos reportados del ICT a nivel nacional.

Definida la función $f(y|\theta) = GB2(\theta) = GB2(\mu, \sigma, v, \tau)$, el proceso formal para ajustar el ingreso ICT en cada entidad es como se expresa a continuación.

Sea $Y^i = (y_1^i, y_2^i, \dots, y_n^i)$ el vector de ingreso corriente total para la entidad i ; sus pesos (factores de expansión) correspondientes están dados por $w^i = (w_1^i, w_2^i, \dots, w_n^i)$. Entonces, el ajuste de la distribución a los datos empíricos que corresponde a maximizar la *log*-verosimilitud de la *GB2*(θ) se expresa de la siguiente manera:

Para cada entidad $i = 1, 2, 3, \dots, 32$

$$\max l(\theta^i | Y^i) = \sum_{h=1}^{n_i} w_h^i \log f(y_h^i | \theta^i) \quad (7)$$

Resultado de estos ajustes obtendremos 32 estimadores de los parámetros de la distribución, los cuales representamos como $\hat{\theta}_{SR} = (\hat{\theta}_{SR}^1, \hat{\theta}_{SR}^2, \dots, \hat{\theta}_{SR}^{32})$, donde $\hat{\theta}_{SR}^i = (\hat{\mu}_{SR}^i, \hat{\sigma}_{SR}^i, \hat{v}_{SR}^i, \hat{\tau}_{SR}^i)^T$.

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

Es decir, sea $Y_o^i = (y_{(1)}^i, y_{(2)}^i, \dots, y_{(n)}^i)$ el vector ordenado de ingresos para la entidad i y sus correspondientes pesos muestrales (factores de expansión) denotados por $w_o^i = (w_{(1)}^i, w_{(2)}^i, \dots, w_{(n)}^i)$; y sea el vector de medianas objetivo $Me = (Me_1, Me_2, \dots, Me_{32})$; para cada entidad $i = 1, 2, 3, \dots, 32$ se resuelve el siguiente problema de optimización restringida:

$$\max l(\theta^i) = \sum_{h=1}^{n_i} w_h^i \log f(y_h^i | \theta^i) \quad (8)$$

sujeto a:

a) Restricciones de igualdad:

$$1. F^{-1}(F(y_{(k)}^i | \hat{\theta}_{SR}^i) | \theta^i) = Me_i$$

donde:

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

Me_i es la mediana objetivo de la entidad federativa i .

5 Estas fueron elegidas debido a que han sido muy usadas en la modelación del ingreso en distintos países con resultados satisfactorios.

$\hat{\theta}_{SR}^i$ es el estimador de los parámetros de la función teórica de densidad sin restricciones para la entidad i .

$F(\cdot)$ es la función de probabilidad acumulada de $GB2$.

$F^{-1}(\cdot)$ es la función cuantil de $GB2$ y

2.

$$\int_0^{max^i} f(y|\theta^i) dy = p^{max}(\hat{\theta}_{SR}^i)$$

donde:

$max^i = y_{(n)}^i$ es el valor máximo del ingreso de ICT en la entidad i .

$p^{max}(\hat{\theta}_{SR}^i) = F(y_{(n)}^i|\hat{\theta}_{SR}^i)$ es la probabilidad acumulada (bajo la función teórica sin restricciones) correspondiente al valor max en la entidad i .

b) Restricciones de desigualdad: las propias del dominio de los valores de los parámetros de la función densidad, para $GB2$ (μ, σ, v, τ):

$$\mu, v, \tau > 0; \quad -\infty < \sigma < \infty; \quad -v < \frac{1}{\sigma} < \tau.$$

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

Derivado de los modelos optimizados, obtenemos un vector que contiene los estimadores de los parámetros restringidos de la función $GB2$, que denotamos como $\hat{\theta}_R = (\hat{\theta}_R^1, \hat{\theta}_R^2, \dots, \hat{\theta}_R^{32})$, donde $\hat{\theta}_R^i = (\hat{\mu}_R^i, \hat{\sigma}_R^i, \hat{v}_R^i, \hat{\tau}_R^i)^T \forall i$.

Lo siguiente es calcular los nuevos microdatos del ICT; para ello, y como se explicó en la sección 3, utilizamos tanto la función de distribución acumulada como la cuantil. Primero, tomamos el valor reportado del ICT para cada hogar y , de acuerdo con la entidad federativa de pertenencia, se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_h^i = F(ICT_h^i|\hat{\theta}_{SR}^i) \forall h = 1, 2, \dots, n_i$; donde n_i es el número de hogares en la entidad i . Enseguida, se obtiene el valor estimado del ICT_fase1 para cada hogar usando la función cuantil: $ICT_fase1_h^i = F^{-1}(\hat{p}_h^i|\hat{\theta}_R^i) \forall h = 1, 2, \dots, n_i$.

De esta manera, obtenemos para cada hogar en toda la muestra un valor del ICT imputado que está determinado en función de los resultados del ajuste por entidad, $ICT_{fase1-h} = \hat{y}_h \forall h = 1, 2, \dots, n$.

Finalmente, el nuevo vector del ingreso por trabajo principal ($ITP_ajustado$) se obtiene como sigue:

- Para cada hogar de la muestra tenemos dos valores: $ICT_original$ y el ICT_fase1 . Esto nos permite obtener un factor de corrección para cada hogar h ,

$$f = \frac{ICT_{fase1}}{ICT_{original}} \triangleq \left(\frac{\hat{y}_1}{y_1}, \frac{\hat{y}_2}{y_2}, \dots, \frac{\hat{y}_n}{y_n} \right) \\ = (f_1, f_2, \dots, f_n)$$

- Calculamos un nuevo valor del ITP para cada hogar de la siguiente manera:

$$\widehat{ITP}_{ajustado} = ITP_{original} * f \triangleq (y_{ITP_1} f_1, y_{ITP_2} f_2, \\ \dots, y_{ITP_n} f_n) \\ \widehat{ITP}_{ajustado} = (\hat{y}_{ITP_1}, \hat{y}_{ITP_2}, \dots, \hat{y}_{ITP_n}) \quad (9)$$

El factor de ajuste f solo es aplicado al ITP debido a que buscamos ajustes diferenciados (no en la misma proporción) para el ITP y para el ITP^C , ya que las mejoras operativas pudieron afectarlos de manera distinta (ver cuadro 2.4).

4.1 Resultados del ajuste al ITP de los hogares⁶

En este apartado presentamos los resultados del ajuste realizado al ingreso por trabajo principal (definido en el cuadro 2.5) de los hogares. El ITP promedio por hogar fue de 25 909 en el 2014, mientras que en la ENIGH 2016 fue de 28 373, ello representaría un incremento de 9.5% en términos reales. Con el ajuste descrito en el apartado anterior al ITP 2016, el promedio baja a 26 467, lo que equivale a un aumento de 2.1% con respecto al 2014.

Si solo ajustamos el ITP y mantenemos el ITP^C originalmente reportado, el promedio por hogar del ICT sería de 44 615 pesos, equivalente a un aumento de 5.8% con respecto al MCS-ENIGH 2014. La figura 4.1 muestra las variaciones del ICT por decil que se tendrían al ajustar solo el ITP, es decir, bajo un ajuste parcial.

⁶ El *software* estadístico que se utilizó para implementar la metodología descrita fue *R* versión 3.4.0 (de uso libre) a través de su IDE *RStudio* versión 1.0.143. Las librerías usadas son:

- `foreign_0.8-67`
- `gamlss_5.0-2`
- `alabama_2015.3-1`
- `survey_3.32`
- `plyr_1.8.4`
- `dplyr_0.7.0`
- `GB2_2.1`

Los resultados pueden variar dependiendo de las características del sistema de la computadora utilizada; el procesamiento del algoritmo fue realizado en un equipo con las siguientes características: Lanix con procesador AMD FX-8370 de ocho núcleos y 4.00 GHz, sistema operativo de 64-bits.

Figura 4.1

Variación porcentual del ICT del 2014 al 2016

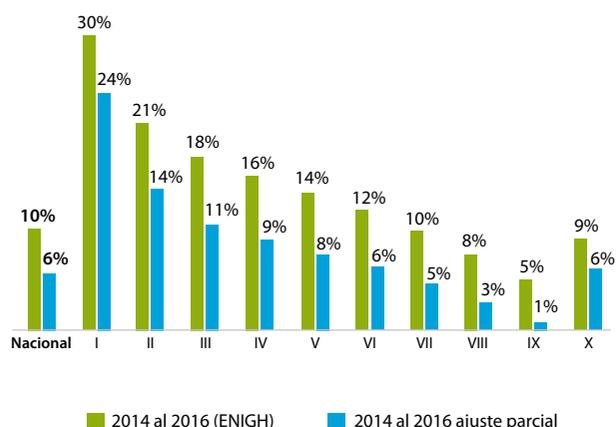
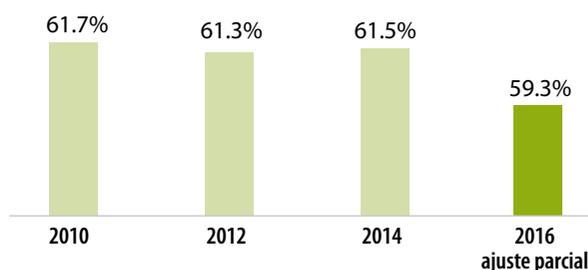


Figura 5.1

Proporción del ITP dentro del ICT



5. Ajuste al complemento del ITP

Un efecto colateral del ajuste realizado al ingreso por trabajo principal es que la proporción que históricamente guardaba (hasta antes del 2014) dentro del ICT se rompe con el ajuste de la sección anterior (ver figura 5.1). Este efecto es consecuencia de que solo se modifica una parte del ICT.

Si suponemos que en ausencia de las mejoras implementadas en la ENIGH 2016 la proporción entre el ITP y el ICT hubiese sido similar a la histórica (61.5% en promedio del 2010 al 2014),⁷ entonces podemos calcular el total del ingreso que no es ITP, tal que la proporción entre el ITP e ICT sea consistente con lo histórico. Es decir, llamemos ITP^c a la diferencia entre el ICT y el ITP. Después del ajuste al vector del ITP ($ITP^{ajustado}$) observamos que, a nivel nacional, los totales de cada ingreso cumplen con la siguiente relación:

$$\frac{T_{ITP^{ajustado}}}{T_{ICT^{ajuste\ parcial}}} \% = \frac{T_{ITP^{ajustado}}}{T_{ITP^{ajustado}} + T_{ITP^c}} \% = 59.3\%$$

Sin embargo, esa proporción debería ser cercana a 61.5% para hacerla consistente con lo que se venía observando hasta el 2014. De ser así, y dado que el valor del $T_{ITP^{ajustado}}$ es conocido, lo único que tendríamos que hacer para conocer el monto del T_{ITP^c} ajustado es el siguiente despeje:

$$\frac{T_{ITP^{ajustado}}}{T_{ICT^{ajuste\ total}}} \% = \frac{T_{ITP^{ajustado}}}{T_{ITP^{ajustado}} + T_{ITP^c\ ajustado}} \% = 61.5\%$$

$$\Rightarrow T_{ITP^c\ ajustado} = \frac{T_{ITP^{ajustado}} (1 - 0.615)}{0.615}$$

⁷ La razón de tomar la proporción histórica (y no la observada en la ENIGH 2016) es que el modelo tiene como objetivo hacer los datos de la ENIGH 2016 comparables con la trayectoria que se venía observando, incluida la relación entre el ITP y su complemento.

Si además dividimos lo anterior entre el número total de hogares, tendríamos el promedio por hogar que garantiza obtener 61.5%:

$$prom_ITP^c \text{ ajustado} = \frac{T_{ITP^c} \text{ ajustado}}{\text{número de hogares poblacionales}} \quad (10)$$

Para distribuir el ajuste del ITP^C a los hogares, utilizamos de nuevo el marco metodológico desarrollado en la sección 3 y realizamos un procedimiento simétrico al de la sección anterior, aunque en este caso se hace un solo ajuste a nivel nacional; es decir, ajustamos una función de distribución $GB2$ al vector ITP^C para obtener un conjunto de parámetros $\hat{\theta}_{SR}^{ITP^c}$; después imponemos como restricción que el nuevo promedio de todos los microdatos del ITP^C sea igual al $prom_ITP^c \text{ ajustado}$, de esta forma obtenemos el segundo conjunto de parámetros $\hat{\theta}_R^{ITP^c}$.

Formalmente, el modelo con restricciones a resolver es el siguiente: sea $H = \{h: 1, 2, \dots, n\}$ conjunto de hogares en la muestra, si suponemos que $ITP^c \sim GB2(z|\theta)$, entonces, el modelo con restricciones quedaría de la siguiente forma:

$$max_{\theta} l(\theta) = \sum_{i=h}^n w_h \log f(ITP_h^c | \theta) \quad (11)$$

sujeto a:

a) Dos restricciones de igualdad.

$$\frac{\sum_H w_h F^{-1}(F(ITP_h^c | \hat{\theta}_{SR}^{ITP^c}) | \theta)}{\sum_H w_h} = prom_ITP^c \text{ ajustado}$$

$$\int_0^{max} f(ITP_h^c | \theta) dITP_h^c = p^{max}(\hat{\theta}_{SR}^{ITP^c})$$

b) Restricciones de desigualdad: $\mu, v, \tau > 0; -\infty < \sigma < \infty; -v < -\frac{1}{\sigma} < \tau$.

Donde max es el máximo valor del ITP^C y $p^{max}(\hat{\theta}_{SR}^{ITP^c})$ es la probabilidad acumulada (bajo la función teórica sin restricciones de ITP^C) correspondiente al valor max . F y f son las funciones acumulada y de densidad de la $GB2$, respectivamente. De este proceso se obtienen los parámetros restringidos $\hat{\theta}_R^{ITP^c}$.

La generación de los nuevos microdatos del ITP^C se realizó de la siguiente manera (ver figura 3.1): se toma el valor reportado del ITP^C para cada hogar y se calcula su probabilidad acumulada según la distribución teórica ajustada sin restricción alguna, $\hat{p}_h^{ITP^c} = F(ITP_h^c | \hat{\theta}_{SR}^{ITP^c}) \forall h = 1, 2, \dots, n$, donde n es el número de hogares en la muestra. El nuevo valor estimado del ITP^C para cada hogar será entonces $\hat{ITP}_h^c = F^{-1}(\hat{p}_h^{ITP^c} | \hat{\theta}_R^{ITP^c}) \forall h = 1, 2, \dots, n$.

Así, el nuevo vector del ICT, \widehat{ICT} , tendría un ajuste secuencial en dos etapas: 1) una donde se ajusta solo el vector de ITP y 2) otra donde solo se ajusta el vector de ITP^C . Los datos por hogar del nuevo vector del ICT, \widehat{ICT}_h , están formados por cada uno de los dos procesos:

$$\widehat{ICT}_h = \widehat{ITP}_h + \widehat{ITP}_h^c \quad (12)$$

Por otra parte, y dado que el objetivo del ajuste es poder medir pobreza, es necesario trasladar el ajuste del ICT a todas las fuentes de ingreso monetario y no monetario que lo componen. Para transferir el efecto del ajuste, se calculan dos factores por hogar: $\alpha_h = \frac{\widehat{ITP}_h}{ITP_h}$ y $\beta_h = \frac{\widehat{ITP}_h^c}{ITP_h^c}$. Cada fuente de ingreso que forma el ITP se multiplica por α_h , mientras que las fuentes de ingreso que forman el ITP^C se multiplican por β_h . Así se generan nuevas versiones de las tablas usadas por la metodología del CONEVAL para el cálculo de la pobreza.

6. Resultados

En los cuadros 6.1 y 6.2 se observa cómo con la ENIGH 2016 se obtienen ingresos considerablemente mayores a los históricamente reportados en los distintos deciles; de igual forma, podemos ver la magnitud del ajuste con el modelo estadístico 2016 y cómo éste recupera la continuidad histórica del ingreso. Si bien el primer decil es el que mayor crecimiento presentó en términos porcentuales, el aumento es similar al registrado entre el 2012 y 2014 (6.2%), periodo donde también fue el de mayor crecimiento.

Con respecto al comportamiento entre entidades, la ENIGH 2016 logra captar dos hogares cuyos ingresos son muy elevados, uno de ellos fue en Nuevo León y otro en Guanajuato. Esto es consecuencia del tamaño de muestra usado (por primera vez la ENIGH 2016 tiene representatividad a nivel entidad federativa y para los ámbitos rural y urbano), lo cual aumenta la probabilidad de captar hogares ricos o con ingresos muy elevados. Tanto en el cuadro 6.3 como en la figura 6.1 se observa la corrección del modelo estadístico para las distintas entidades federativas.

Con respecto a los cálculos de pobreza (ver cuadro 6.4), se consideró el ajuste a la carencia de seguridad social (adicional a los ajustes al ICT desarrollados en secciones anteriores) realizado por el CONEVAL: previo al levantamiento de la ENIGH 2016, el CONEVAL solicitó el cambio de fraseo para algunas preguntas relacionadas con el acceso a los servicios médicos y con otras dos prestaciones laborales que son consideradas para la identificación de la condición de carencia por seguridad social de la población trabajadora subordinada. Dicho cambio tuvo un efecto sobre la consistencia histórica de las respuestas que dan los informantes. La base de datos original de la ENIGH 2016 reporta tal cual los resultados del levantamiento en campo que resultan del nuevo fraseo. Sin embargo, buscando recuperar la continuidad de las series, se incorporan los resultados de un ejercicio estadístico que el CONEVAL diseñó para tal propósito. El ajuste se focaliza en la población trabajadora subordinada con acceso a servicios médicos como prestación laboral para recuperar la consistencia de la serie histórica de la carencia por acceso a la seguridad social.⁸

8 Para mayores detalles véase: http://www.coneval.org.mx/Medicion/MP/Documents/Pobreza_16/Notas_Pobreza_2016/Nota_tecnica_2_ajuste_seguridad_social.pdf

Cuadro 6.1

**Promedio del ICT por deciles de hogares
(precios constantes 2016)**

Decil	MCS-ENIGH 2008	MCS-ENIGH 2010	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016	Modelo estadístico 2016
Nacional	46 901	43 931	42 115	42 157	46 521	43 036
I	6 270	6 111	5 920	6 288	8 166	6 820
II	12 318	11 707	11 310	11 731	14 206	12 350
III	17 230	16 285	15 750	16 062	18 918	16 677
IV	22 162	20 876	20 326	20 336	23 556	21 029
V	27 599	26 098	25 357	25 258	28 813	25 920
VI	34 228	32 400	31 502	31 030	34 837	31 650
VII	43 037	40 352	39 343	38 489	42 431	38 841
VIII	55 119	52 028	50 494	49 463	53 383	49 218
IX	76 315	72 580	69 321	68 386	72 041	67 034
X	174 735	160 872	151 828	154 524	168 856	160 820

Cuadro 6.2

**Variación porcentual del ICT por deciles de hogares
(precios constantes 2016)**

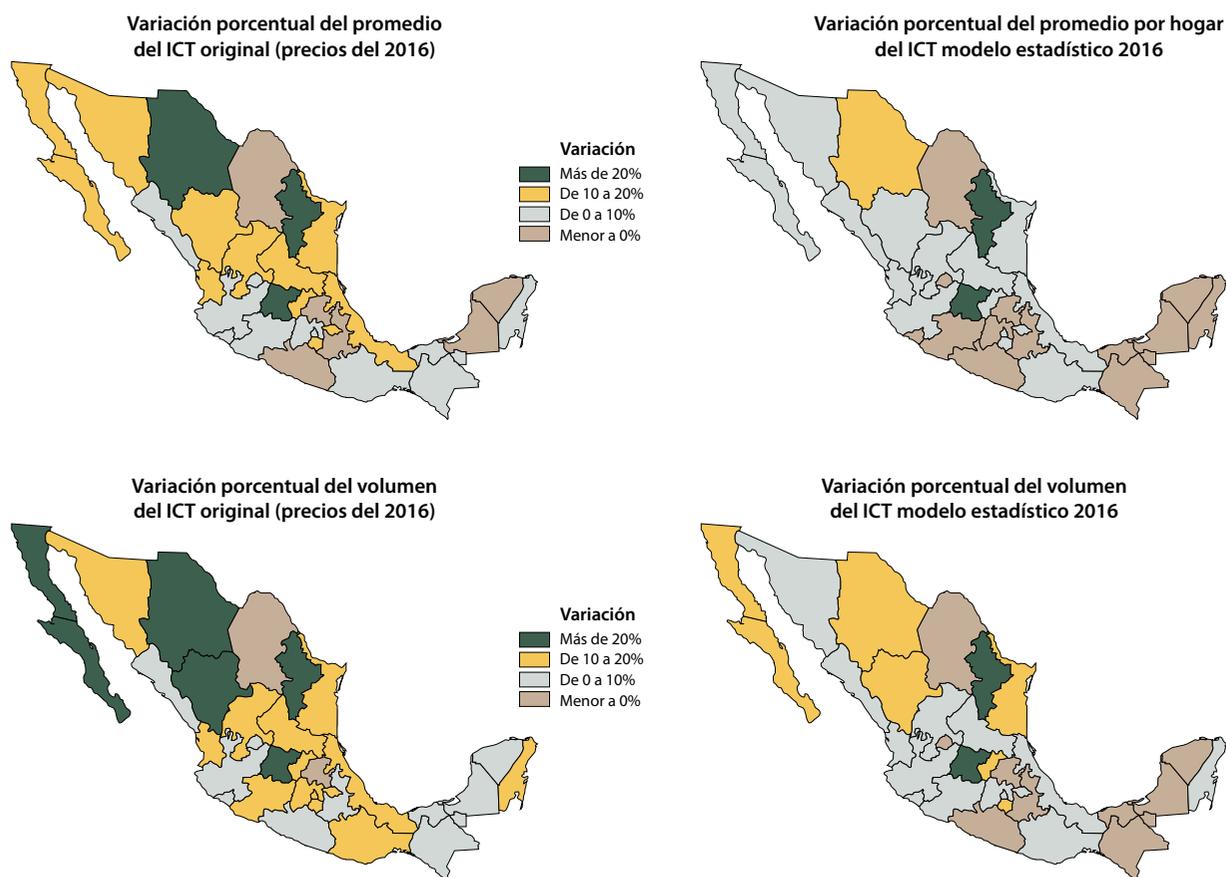
Decil	2008-2010	2010-2012	2012-2014	2014-2016	2014-Modelo estadístico 2016
Nacional	-6.3%	-4.1%	0.1%	10.4%	2.1%
I	-2.5%	-3.1%	6.2%	29.9%	8.5%
II	-5.0%	-3.4%	3.7%	21.1%	5.3%
III	-5.5%	-3.3%	2.0%	17.8%	3.8%
IV	-5.8%	-2.6%	0.1%	15.8%	3.4%
V	-5.4%	-2.8%	-0.4%	14.1%	2.6%
VI	-5.3%	-2.8%	-1.5%	12.3%	2.0%
VII	-6.2%	-2.5%	-2.2%	10.2%	0.9%
VIII	-5.6%	-2.9%	-2.0%	7.9%	-0.5%
IX	-4.9%	-4.5%	-1.3%	5.3%	-2.0%
X	-7.9%	-5.6%	1.8%	9.3%	4.1%

Promedio del ingreso corriente de hogares por entidad

Entidad federativa	MCS-ENIGH 2008	MCS-ENIGH 2012	MCS-ENIGH 2014	ENIGH 2016	Modelo estadístico 2016
1 Aguascalientes	52 572	45 394	48 870	49 247	45 938
2 Baja California	60 507	50 648	50 401	57 687	54 462
3 Baja California Sur	71 989	53 564	50 719	56 351	53 270
4 Campeche	46 521	44 145	45 641	44 211	42 055
5 Coahuila de Zaragoza	48 333	47 841	52 858	46 588	46 157
6 Colima	56 067	44 829	46 726	47 733	43 398
7 Chiapas	24 185	24 798	23 049	23 258	21 836
8 Chihuahua	55 842	44 097	39 872	48 111	45 228
9 Ciudad de México	73 834	58 490	66 738	70 834	65 241
10 Durango	37 355	34 499	34 305	40 384	37 998
11 Guanajuato	43 780	42 319	36 747	56 974	51 233
12 Guerrero	28 545	24 521	27 538	26 980	25 051
13 Hidalgo	33 689	33 173	33 717	32 535	31 418
14 Jalisco	54 407	47 333	48 705	52 367	47 864
15 México	45 204	44 732	40 992	43 319	39 217
16 Michoacán de Ocampo	35 279	33 787	31 475	33 788	30 113
17 Morelos	40 802	38 827	36 907	41 196	36 105
18 Nayarit	40 537	35 616	38 818	43 064	40 031
19 Nuevo León	71 316	63 974	63 108	87 653	90 394
20 Oaxaca	29 315	27 924	25 324	27 704	23 382
21 Puebla	32 458	30 915	37 759	35 369	31 094
22 Querétaro	57 072	49 385	49 759	56 605	51 163
23 Quintana Roo	55 184	44 497	49 387	50 568	49 623
24 San Luis Potosí	37 998	35 334	34 950	39 051	34 371
25 Sinaloa	52 145	43 785	45 085	48 591	44 361
26 Sonora	59 988	54 736	53 185	58 813	53 882
27 Tabasco	35 967	38 318	36 845	37 383	35 134
28 Tamaulipas	46 734	43 032	42 332	48 049	44 110
29 Tlaxcala	33 487	31 923	30 903	34 050	32 178
30 Veracruz de l. de la Llave	36 242	32 262	29 794	32 983	28 716
31 Yucatán	45 876	39 383	42 681	42 457	41 067
32 Zacatecas	36 712	32 469	31 938	36 246	33 903
Nacional	46 907	42 115	42 157	46 521	43 036

Figura 6.1

Variación porcentual del ICT ajustado del 2014 al 2106 (precios del 2016)



6.1 Estimación de los errores del modelo

La metodología desarrollada en este documento no es parte de las metodologías tradicionales de imputación/edición de microdatos, por lo que hasta la fecha no se cuenta con procedimientos estadísticos para la cuantificación de los errores en las estimaciones que arroja. Sin embargo, y aprovechando procesos como el de *bootstrap*, podemos generar mediciones que permitan tener un acercamiento a errores del modelo. Para ello, los cuadros 6.5 y 6.6 muestran el resultado de 500 repeticiones del algoritmo de ajuste del ingreso, con remuestras distintas obtenidas a través de *bootstrap*.

Como se observa en ambos cuadros, el promedio de las 500 repeticiones es muy cercano a los datos reportados en los cuadros 6.1 y 6.3. Las entidades donde se observan las mayores distancias entre ambas estimaciones son Guanajuato, Nuevo León y Tamaulipas; de ellas, y como ya se ha mencionado, en las dos primeras se captaron hogares con ingresos muy elevados, lo que puede estar influyendo en las estimaciones.

Cuadro 6.4

Porcentaje de población en pobreza multidimensional

Entidad federativa	2010	2012	2014	Modelo estadístico 2016
Aguascalientes	38.1	37.8	34.8	28.2
Baja California	31.5	30.2	28.6	22.2
Baja California Sur	31.0	30.1	30.3	22.1
Campeche	50.5	44.7	43.6	43.8
Coahuila de Zaragoza	27.8	27.9	30.2	24.8
Colima	34.7	34.4	34.3	33.6
Chiapas	78.5	74.7	76.2	77.1
Chihuahua	38.8	35.3	34.4	30.6
Ciudad de México	28.5	28.9	28.4	27.6
Durango	51.6	50.1	43.5	36.0
Guanajuato	48.5	44.5	46.6	42.4
Guerrero	67.6	69.7	65.2	64.4
Hidalgo	54.7	52.8	54.3	50.6
Jalisco	37.0	39.8	35.4	31.8
México	42.9	45.3	49.6	47.9
Michoacán de Ocampo	54.7	54.4	59.2	55.3
Morelos	43.2	45.5	52.3	49.5
Nayarit	41.4	47.6	40.5	37.5
Nuevo León	21.0	23.2	20.4	14.2
Oaxaca	67.0	61.9	66.8	70.4
Puebla	61.5	64.5	64.5	59.4
Querétaro	41.4	36.9	34.2	31.1
Quintana Roo	34.6	38.8	35.9	28.8
San Luis Potosí	52.4	50.5	49.1	45.5
Sinaloa	36.7	36.3	39.4	30.8
Sonora	33.1	29.1	29.4	27.9
Tabasco	57.1	49.7	49.6	50.9
Tamaulipas	39.0	38.4	37.9	32.2
Tlaxcala	60.3	57.9	58.9	53.9
Veracruz de I. de la Llave	57.6	52.6	58.0	62.2
Yucatán	48.3	48.9	45.9	41.9
Zacatecas	60.2	54.2	52.3	49.0
Nacional	46.1	45.5	46.2	43.6

Cuadro 6.5

Resultados por decil del proceso *bootstrap*

Decil	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
I	6 820	6 802	27.58
II	12 350	12 367	29.08
III	16 677	16 719	30.27
IV	21 029	21 082	31.99
V	25 920	25 978	33.57
VI	31 650	31 717	35.55
VII	38 841	38 915	37.99
VIII	49 218	49 338	42.82
IX	67 034	67 237	54.91
X	160 820	160 927	374.41

Cuadro 6.6

Continúa

Resultados por entidad federativa del proceso *bootstrap*

Entidad	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
Aguascalientes	45 938	46 022	51.92
Baja California	54 462	54 484	62.41
Baja California Sur	53 270	53 265	60.29
Campeche	42 055	42 205	55.68
Coahuila de Zaragoza	46 157	46 258	52.13
Colima	43 398	43 694	56.25
Chiapas	21 836	21 928	36.06
Chihuahua	45 228	45 473	64.34
Ciudad de México	65 241	65 592	79.63
Durango	37 998	38 144	43.88
Guanajuato	51 233	50 718	419.04
Guerrero	25 051	25 192	41.21
Hidalgo	31 418	31 426	52.13
Jalisco	47 864	47 781	70.13
México	39 217	39 300	49.45

Resultados por entidad federativa del proceso *bootstrap*

Entidad	Modelo estadístico 2016	Resultado de las 500 repeticiones	
		Promedio	Error estándar
Michoacán de Ocampo	30 113	30 234	40.65
Morelos	36 105	36 233	52.32
Nayarit	40 031	40 136	69.32
Nuevo León	90 394	89 831	880.87
Oaxaca	23 382	23 594	47.56
Puebla	31 094	31 290	52.01
Querétaro	51 163	51 386	66.7
Quintana Roo	49 623	49 818	57.53
San Luis Potosí	34 371	34 543	45.52
Sinaloa	44 361	44 375	52.13
Sonora	53 882	54 106	138.13
Tabasco	35 134	35 054	48.55
Tamaulipas	44 110	44 540	99.72
Tlaxcala	32 178	32 272	40.82
Veracruz de I. de la Llave	28 716	28 888	48.67
Yucatán	41 067	41 181	73.82
Zacatecas	33 903	34 006	73.55

7. Conclusiones

En un proceso continuo de mejora, el INEGI identificó áreas de oportunidad alrededor de la capacitación, supervisión y control operativo del MCS-ENIGH. A partir de ello, se llevaron a cabo acciones de fortalecimiento operativo en el levantamiento de la ENIGH 2016, en un esfuerzo por alcanzar un mayor apego al diseño de la encuesta.

La mayoría de los componentes del ICT tuvo un crecimiento considerable en la ENIGH 2016 con respecto a la información del MCS-ENIGH 2014. Uno de ellos es el ingreso del trabajo, el cual creció en este periodo 7.1% (para 2010-2012 y 2012-2014, las tasas de crecimiento fueron de 3.2 y 1.4%, respectivamente). Este rubro del ingreso ha representado más de 60% del ICT históricamente.

Los ingresos no laborales también crecieron de manera significativa; en términos del promedio por hogar, su crecimiento fue de 27% entre el 2014 y 2016, mientras que para el periodo 2012 al 2014 fue de solo 2.8 por ciento.

Debido al carácter único de la información que recopila la ENIGH, no existen otras encuestas con las cuales podamos comparar el comportamiento de sus distintas fuentes de ingreso. Sin embargo, la ENOE nos permite hacerlo para la fuente principal del ingreso de los hogares: los ingresos por trabajo principal. Una característica fundamental de la ENOE es que la recolección de información ha sido similar en los distintos levantamientos y que, por lo tanto, no sufrió cambios importantes en la captación de ingresos.

La metodología que aquí se propone edita los microdatos reportados a través de funciones de distribución. Primero, se ajusta una función de distribución conocida (*GB2*, en nuestro caso) a los datos muestrales y, con ello, se obtienen estimaciones de los parámetros que determinan la forma funcional de la función de distribución. Después, se realiza un proceso similar, pero imponiendo un conjunto de restricciones que debe cumplir el ajuste; estas condiciones involucran la obtención de valores puntuales (valores objetivo) de estadísticos muestrales, lo cual implica la modificación de los datos reportados en la muestra.

El proceso de ajuste se realizó por separado para el ITP y para los ingresos distintos a estos últimos (*ITP^c*). Para obtener los valores objetivo de las restricciones impuestas, nos basamos en el crecimiento de la mediana de ingresos reportada en la ENOE; con ello modificamos el ITP. Los valores objetivo para las restricciones que entran en el ajuste al *ITP^c* se obtuvieron de una relación empírica que se ha venido observando desde el 2010 entre el *ITP^c* y el ICT, y la cual asumimos que se debe mantener.

En términos de resultados, y considerando ambos ajustes (tanto al ITP como al *ITP^c*), se logra que la variación entre el 2014 y 2016 del ICT promedio por hogar pase de 10.4 a 2.1 por ciento. El crecimiento para el primer decil es de 8.5% (en comparación con 29.9% al utilizar los datos de la ENIGH 2016). Cuando estos resultados se desagregan por entidad federativa, también se observan ajustes a la baja en las tasas de crecimiento para la mayoría de las entidades.

Referente a la medición de pobreza, una implicación de la disminución de los ingresos es el aumento en pobreza con respecto a los datos de la ENIGH 2016. Con el ajuste propuesto, la población en situación de pobreza multidimensional es de 43.6% (para el 2014 fue de 46.2%). Cabe mencionar que esta cifra de pobreza incorpora un ajuste adicional a la carencia por seguridad social realizada por el CONEVAL.

Fuentes

- Bandourian, Ripsy; James McDonald and Robert Turley. "A Comparasion of Parametric Models of Income Distribution Across Countries and Over Time Across Countries and Over Time", en: *Luxembourg Income Study Working Paper No. 305*. 2002.
- Bustos, Alfredo. "Estimation of the distribution of income from survey data, adjusting for compatibility with other sources", en: *Statistical Journal of the IAOS. Journal of the International Association for Official Statistics*. Vol. 31, no. 4, noviembre del 2015, pp. 565-577.
- García, Carmelo; Mercedes Prieto e Hipólito Simón. *La modelización paramétrica de las distribuciones salariales. Un estudio aplicado al caso español*. Universidad de Alicante, Departamento de Análisis Económico Aplicado, junio del 2012, 31 pp.
- Graf, Monique; Desislava Nedyalkova; Ralf Münnich; Jan Seger and Stefan Zins. *Parametric estimation of income distributions and indicators of poverty and social exclusion. Advanced Methodology for European Laeken Indicators*. Sevent Framework Programme, 2011, 74 pp.
- McDonald, James B. "Some generalized functions for the size distribution of income", en: *Modeling Income Distributions and Lorenz Curves*. Capítulo 3. Mayo de 1984, pp. 37-55.
- McDonald, James B. and Michael Ransom. "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality", en: *Modeling Income Distributions and Lorenz Curves*. Cap. 8. Economic Studies in Inequality Social Exclusion and Well-Being. Springer, 2008, pp. 147-164.
- McDonald, James B. and Yexiao J. X. "A generalization of the beta distribution with applications", en: *Journal of Econometrics*. Vol. 66, Issues 1-2, marzo-abril de 1995, pp. 133-152.
- Rigby, R. A. and D. M. Stasinopoulos. "Generalized additive models for location, scale and shape", en: *Appl. Statist.* 54, Part 3. 2005, pp. 507-554.
- Singh, S. K. and G. S. Maddala. "A function for size distribution or incomes", en: *Modeling Income Distributions and Lorenz Curves*. Vol. 44. Septiembre de 1976, pp. 963-970.
- Skinner, C.; D. Holt and T. Smith (editors). *Analysis of Complex Surveys*. New York, USA, John Wiley and Sons, 1989.
- Stasinopoulos, Mikis; Bob Rigby and Calliope Akantziliotou. *Instructions on how to use the gamlss package in R*. Second edition. Enero del 2008, 206 pp.

Anexo 1

En estudios previos, y para distintos países, se han probado familias de distribuciones para ajustar variables, como ingreso y salario, con resultados satisfactorios. Algunas de las funciones más utilizadas se muestran en la figura A.1 obtenida de Bandourian *et al.* (2002).

La siguiente parametrización de las formas funcionales elegidas para el estudio puede ser encontrada en el documento *Instructions on how to use the gamlss package in R*.

Beta Generalizada tipo 2: GB2 (μ, σ, v, τ)

$$f_Y(y|\mu, \sigma, v, \tau) = |\sigma| y^{\sigma v - 1} \{ \mu^{\sigma v} B(v, \tau) [1 + (\frac{y}{\mu})^\sigma]^{v + \tau} \}^{-1}$$

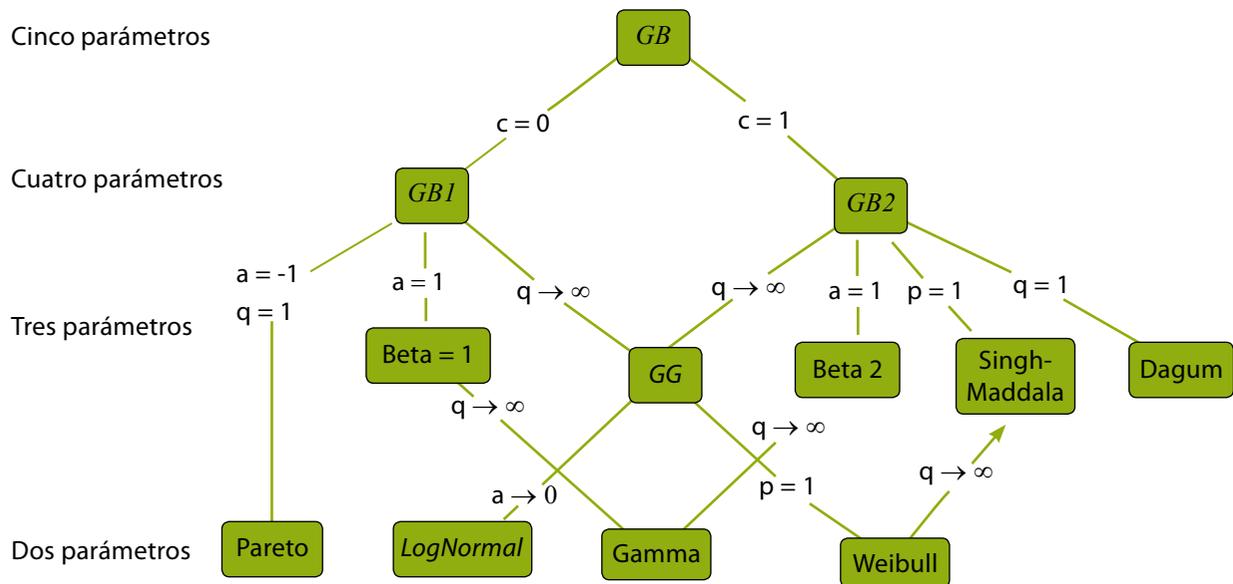
Para $y > 0$, donde $\mu, v, \tau > 0$, $-\infty < \sigma < \infty$ $y - v < \frac{1}{\sigma} < \tau$.

Dagum

$$Dagum(y|\mu, \sigma, v) = GB2(y|\mu, \sigma, v, \tau = 1)$$

Figura A.1

Esquema de las funciones que se usan en la modelación del ingreso



Gamma Generalizada: GG (μ, σ, v)

$$f_Y(y|\mu, \sigma, v) = \frac{|v|\theta^\theta z^\theta \exp\{-\theta z\}}{\Gamma(\theta)y}$$

Para $y > 0$, donde $\mu > 0, \sigma > 0, y - \infty < v < \infty$. Y donde $z = (\frac{y}{\mu})^v$ y $\theta = 1/(\sigma^2 v^2)$.

LogNormal distribución: LogNor (μ, σ)

$$f_Y(y|\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{y} \exp\left\{-\frac{[\log(y) - \mu]^2}{2\sigma^2}\right\}$$

Para $y > 0$, donde $\mu > 0$ y $\sigma > 0$.

Aquí $E(Y) = \omega^{1/2} e^\mu$ y $Var(Y) = \omega(\omega - 1)e^{2\mu}$, donde $\omega = \exp(\sigma^2)$.

Anexo 2

Para determinar la función paramétrica a utilizar, realizamos algunas pruebas de bondad de ajuste y elementos de visualización que nos permitieran elegir la función que mejor se ajusta a los datos empíricos. El resultado fue la selección de la función *GB2* para nuestro modelo.

Estadísticos para pruebas de bondad de ajuste

Definimos:

$$p_i(\hat{\theta}) = F(y_i|\hat{\theta}) - F(y_{i-1}|\hat{\theta})$$

$$N = \sum_{i=1}^n w_i$$

w_i = número de hogares que representa el ingreso y_i .

SSE suma del cuadrado de errores

$$SSE = \sum_{i=1}^n \left(\frac{w_i}{N} - p_i(\hat{\theta}) \right)^2$$

$$SSE_{acum} = \sum_{i=1}^n \left(\frac{\sum_{m=1}^i w_m}{N} - F(y_i|\hat{\theta}) \right)^2$$

Estadístico χ^2

$$\chi^2 = N \sum_{i=1}^n \left[\frac{\left(\frac{w_i}{N} - p_i(\hat{\theta}) \right)^2}{p_i(\hat{\theta})} \right]$$

Log-verosimilitud

$$l(\theta_k)$$

Aikaike

$$AIC(k) = 2k - 2l(\theta_k)$$

BIC

$$BIC(k) = \ln(n)k - 2l(\theta_k)$$

K = número de parámetros.

Visualización

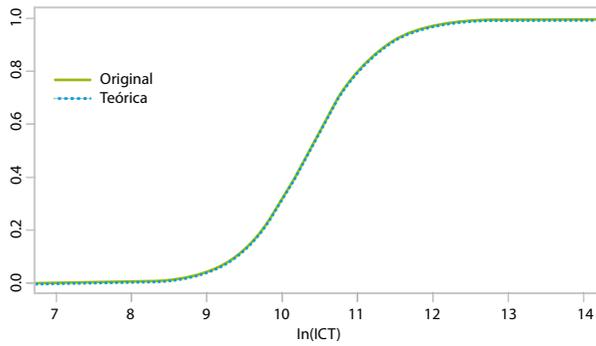
Función de densidad

En primera instancia se puede comprobar si la función elegida se adecua al histograma de los datos observados en la encuesta (ver grupo 2 de gráficas).

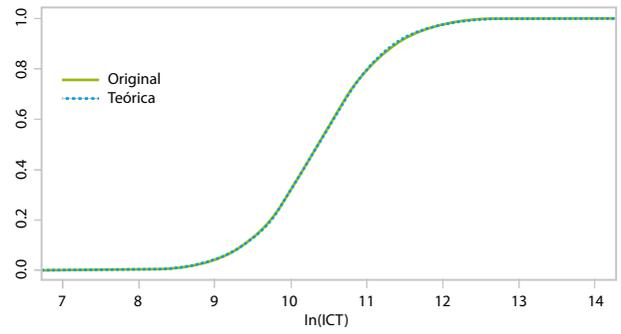
Grupo 1

Distribución acumulada, teórica (ajustada) vs. ENIGH 2016 (empírica)

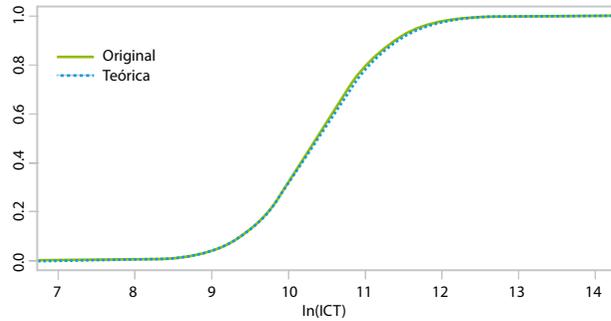
Ajuste *Dagum*: muestra expandida
Distribución acumulada, original vs. teórica



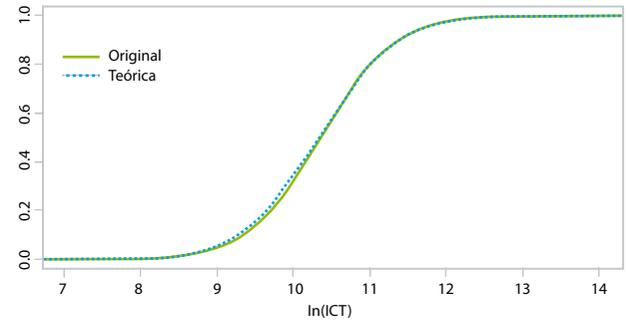
Ajuste *GB2*: muestra expandida
Distribución acumulada, original vs. teórica



Ajuste *GG*: muestra expandida
Distribución acumulada, original vs. teórica



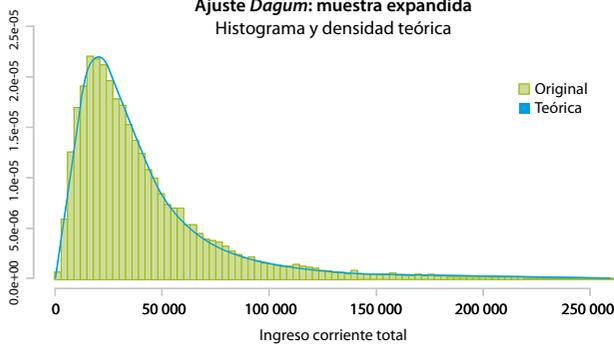
Ajuste *LogNor*: muestra expandida
Distribución acumulada, original vs. teórica



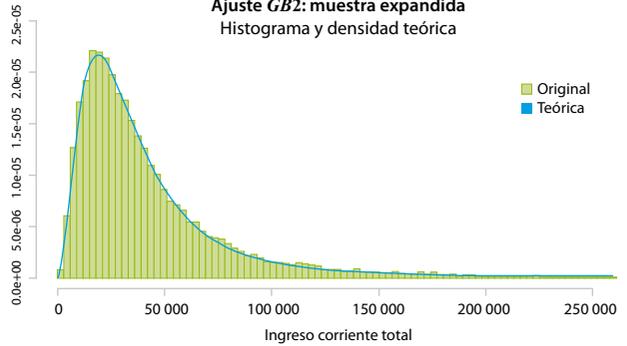
Grupo 2

Densidad, teórica (ajustada) vs. ENIGH 2016 (empírica)

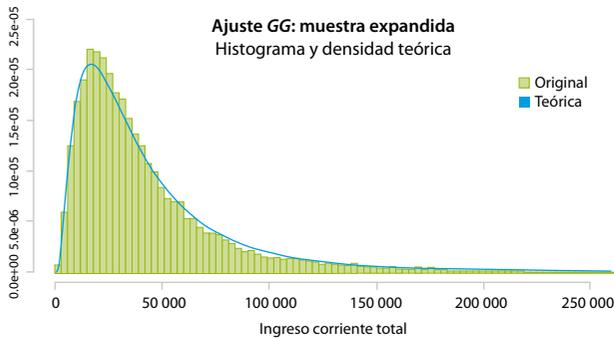
Ajuste *Dagum*: muestra expandida
Histograma y densidad teórica



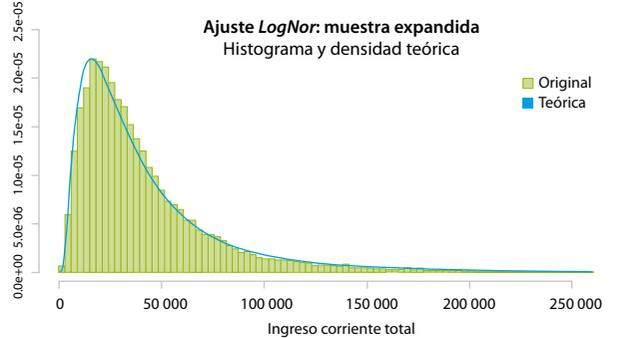
Ajuste *GB2*: muestra expandida
Histograma y densidad teórica



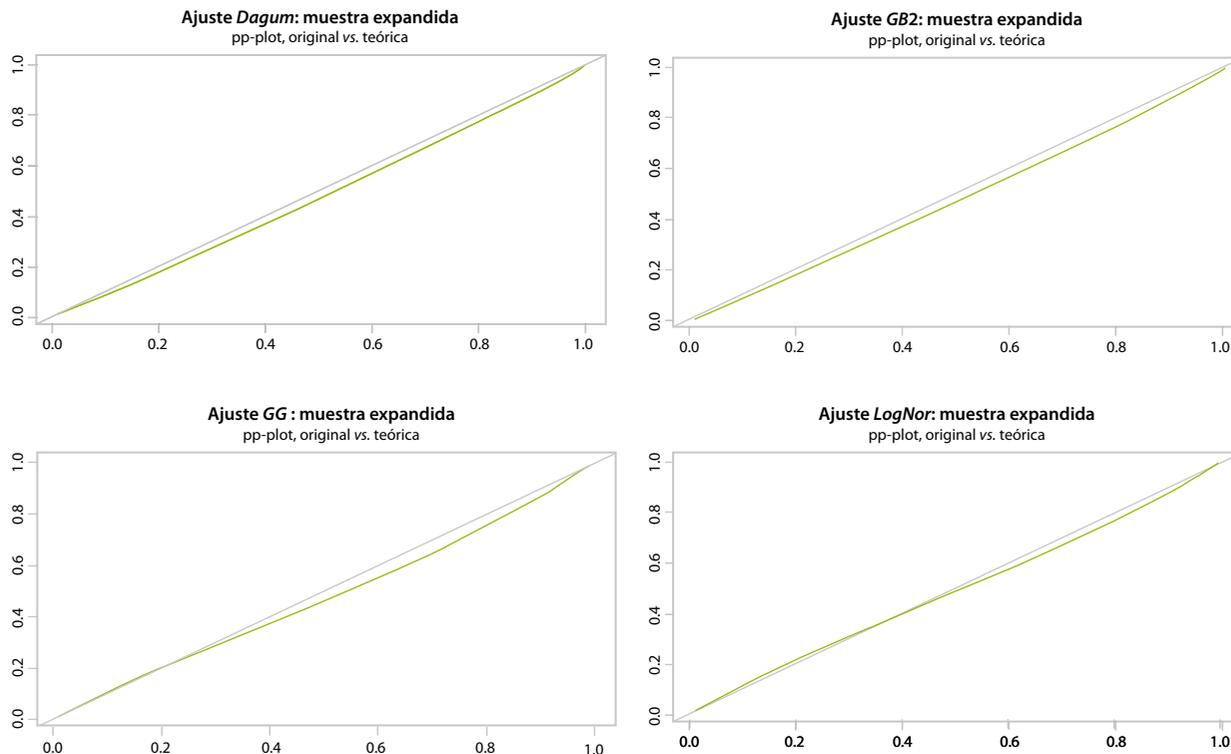
Ajuste *GG*: muestra expandida
Histograma y densidad teórica



Ajuste *LogNor*: muestra expandida
Histograma y densidad teórica



pp-plots, teórica (ajustada) vs. ENIGH 2016 (empírica)



Estadísticos

Suma cuadrática de errores

Decil	GB2	GG	LogNor	Dagum
I	7.54	199.33	366.29	23.35
II	2.65	581.95	1 974.89	62.28
III	6.05	291.24	2 365.92	98.51
IV	26.30	19.68	1 550.56	80.60
V	3.20	156.51	904.04	15.55
VI	0.40	731.64	219.12	63.93
VII	0.81	1 397.93	20.16	111.35
VIII	8.07	1 295.13	92.24	147.25
IX	8.41	546.10	103.47	52.57
X	4.18	48.22	15.85	22.66

Log-verosimilitud y BIC

	GB2	GG	LogNor	Dagum
Log-verosimilitud	387 168 080	387 702 106	387 956 925	387 212 144
BIC	774 336 167	775 404 219	775 913 854	774 423 900

Función de distribución acumulada

En el grupo 1 se muestran las probabilidades acumuladas que asigna la función teórica a los ingresos observados $F(Y|\theta) = P(Y \leq y); y > 0$.

pp-plot

En el grupo de gráficas 3 refleja la discrepancia para cada observación de la muestra con respecto a los valores que la función del modelo teórico pronostica; cuanto más se aproximan a la diagonal del cuadrante la calidad del ajuste es mejor.

Promedio del ICT por decil

Decil	Empíricos	GB2	GG	LogNor	Dagum
I	8 166.343	8 088.938	7 658.307	7 611.810	8 090.211
II	14 206.059	14 229.802	13 548.606	13 090.490	14 424.132
III	18 918.086	18 977.140	18 470.364	17 684.250	19 166.112
IV	23 555.613	23 690.323	23 552.246	22 467.080	23 780.692
V	28 812.506	28 827.502	29 216.253	27 851.860	28 753.483
VI	34 836.654	34 845.130	35 912.906	34 294.500	34 555.896
VII	42 431.256	42 464.582	44 358.019	42 536.250	41 932.233
VIII	53 383.482	53 215.095	56 017.062	54 122.790	52 488.715
IX	72 040.802	71 708.933	75 008.577	73 490.730	71 238.109
X	168 855.531	153 052.109	136 041.401	140 650.320	169 229.232

Cuantiles

Cuantiles	Empíricos	GB2	GG	LogNor	Dagum
1	4 776.380	4 396.711	4 341.613	4 516.718	4 208.322
2	6 017.870	5 798.776	5 556.088	5 658.439	5 684.630
5	8 524.280	8 511.483	7 995.246	7 932.096	8 541.460
10	11 587.230	11 661.153	10 984.335	10 707.107	11 811.650
20	16 616.400	16 657.933	16 025.386	15 396.725	16 868.953
30	21 184.660	21 296.024	20 945.848	20 006.730	21 448.162
40	26 072.200	26 153.967	26 253.846	25 025.791	26 171.796
50	31 657.340	31 643.406	32 346.301	30 849.512	31 468.989
60	38 285.170	38 302.901	39 760.637	38 028.336	37 893.551
70	47 059.050	47 112.358	49 459.267	47 567.238	46 462.043
80	60 687.960	60 424.185	63 645.591	61 808.619	59 673.107
90	87 032.500	86 942.300	89 793.806	88 886.264	87 158.768
95	120 797.390	120 120.403	118 730.288	119 978.997	123 730.982
98	175 332.940	178 492.345	161 774.033	168 167.788	193 120.251
99	226 429.530	237 723.877	198 248.318	210 624.507	268 848.551