



## **SUBDIRECCIÓN**

# **NOTA METODOLÓGICA METODOLOGIA PARA EL EMPALME DE LAS SERIES DE POBREZA MONETARIA, POBREZA MONETARIA EXTREMA Y GINI PARA LOS AÑOS 2012 A 2020**

**Diciembre de 2024**

## **CONTENIDO**

<b>1. Introducción</b>	<b>3</b>
<b>2. Metodología de empalme</b>	<b>6</b>
<b>2.1. Empalme de la serie de ingreso</b>	<b>6</b>
<b>2.2. Ajuste de los factores de expansión</b>	<b>10</b>
<b>2.3. Cálculo de la incidencia de pobreza, pobreza extrema y Gini</b>	<b>15</b>
<b>2.4. Resultados, gráficas</b>	<b>17</b>
<b>3. Conclusión</b>	<b>20</b>
<b>4. Anexo</b>	<b>21</b>
<b>5. Referencia</b>	<b>21</b>

## 1. Introducción

*Para la medición de la pobreza según Sen (1976) existen dos principales métodos, el primero de ellos es el método directo en el cual se identifica el conjunto de hogares (o personas) que no satisfacen un grupo específico de necesidades previamente establecidas (condiciones de la vivienda, educación, composición demográfica del hogar, tenencia de activos, etc.). El segundo método se denomina método indirecto, en este la identificación se lleva a cabo mediante el cálculo de un umbral mínimo, por lo general asociado a un nivel de ingreso o gasto, por debajo del cual se considera que una persona no puede satisfacer sus necesidades básicas. La diferencia está en que mientras el primero se basa en la satisfacción efectiva de un conjunto de necesidades, el segundo hace referencia a la posibilidad de satisfacerlas (considerando para ello un nivel mínimo de ingresos).*

*Para obtener la incidencia de pobreza y pobreza extrema para un año cualquiera se compara el ingreso per cápita mensual de la unidad de gasto -IPCG (calculado a partir de las Encuestas de Hogares del año respectivo) frente a los valores de las líneas de pobreza y pobreza extrema actualizados por la inflación, creando 3 categorías de clasificación de los hogares:*

- **No pobre:** si el ingreso es superior a la línea de pobreza.
- **Pobre:** si el ingreso es inferior a la línea de pobreza.
- **Pobre extremo:** si el ingreso per cápita es inferior a la línea de indigencia

*El coeficiente de Gini es el indicador que se utiliza con mayor frecuencia para medir el grado de desigualdad en la distribución del ingreso en un país. En términos gráficos, es la distancia entre la curva de Lorenz y la línea de equidistribución de los ingresos. Este coeficiente varía entre cero y uno, tomando el valor 0 en el escenario de perfecta igualdad, donde los ingresos de los hogares de un país son todos iguales, es decir el ingreso está equitativamente distribuido y toma el valor uno cuando la desigualdad es total, es decir, el ingreso se concentra en un hogar o un individuo.*

*La Gran Encuesta Integrada de Hogares (GEIH), además de proveer información sobre el mercado laboral, contiene preguntas sobre los ingresos de los hogares que permiten construir el ingreso per cápita de la unidad de gasto, en Colombia la medición de pobreza monetaria se realiza a través de la definición de unas líneas de pobreza e indigencia, esta técnica se enmarca en el*

*método indirecto, y fue producto de actualización y revisión metodológica por parte de la MESEP. La línea de indigencia corresponde al valor de una canasta básica de alimentos y la de pobreza al valor de una canasta que incluye además de alimentos otros bienes básicos.*

*El marco de muestreo para las encuestas a hogares corresponde a un marco de áreas donde se cubre todo el territorio nacional, con el propósito de que cada unidad del marco tenga una probabilidad conocida y mayor a cero. Con el Censo Nacional de Población y Vivienda (CNPV) de 2018, se generó un nuevo marco muestral para seleccionar muestras de hogares, por lo que se tiene un nuevo marco para la Gran Encuesta Integrada de Hogares (GEIH). En el año 2019 inició el proceso de rediseño de la encuesta con el fin de actualizar los diseños muestrales, operativos y conceptuales de la encuesta, con la información del CNPV 2018 y los actuales lineamientos de mercado laboral y pobreza monetaria. En el marco de este proceso, en 2021 se inició la recolección de la GEIH marco 2018 en paralelo a la GEIH marco 2005, teniendo para este periodo dos encuestas paralelas con diferentes marcos. En diciembre de 2021 se terminó la recolección en paralelo y a partir del año 2022 la información de pobreza es proporcionada por la GEIH marco 2018.*

*Dado lo anterior, es necesario generar series históricas comparables para la medición de los indicadores de pobreza según el cambio de marco, las proyecciones poblacionales del CNPV 2018 y la actualización de la Población en Edad de Trabajar (PET) de 15 años y más. De esta forma se espera la comparabilidad de los indicadores en el tiempo a nivel nacional y en cada uno de los 25 dominios: total cabeceras, total centros poblados y rural disperso y para cada una de las 23 ciudades y áreas metropolitanas.*

*Al interior del DANE se llevó a cabo un ejercicio inicial para empalmar las series en el año 2021, este consistió en realizar una partición geográfica de la muestra nacional en 31 subconjuntos y al interior de estos se dividió cada subconjunto en dos: hogares con ingreso igual a cero y hogares con ingreso mayor a cero. Posteriormente para los hogares con ingreso mayor a cero se calcula al interior de cada decil del ingreso per cápita (10 particiones adicionales) un factor de ajuste el*

*cual corresponde al cociente de la mediana del ingreso per cápita bajo el marco 2018 sobre la mediana del ingreso per cápita bajo el marco 2005. Para el empalme faltante (2012-2020) se buscó otra metodología más robusta técnicamente y se presenta en este documento.*

*Se plantea entonces el ejercicio de empalme de las series de pobreza monetaria, pobreza monetaria extrema y coeficiente de Gini explorando una alternativa metodológica en la cual se hace uso de los modelos estructurales, los cuales han demostrado ser de gran utilidad para el empalme de las series de tiempo cuando se tiene o no una encuesta paralela. Este proceso de empalme está conformado por tres puntos fundamentales, en primer lugar, dado que los tres indicadores a empalmar dependen de la variable ingreso per cápita, el objetivo es obtener un adecuado empalme del ingreso, acercándonos a la distribución de este empalmando los deciles, con esto se busca capturar el impacto en esta variable por el cambio de marco. De esta forma surge el segundo punto el cual consiste en calcular un factor de ajuste para el factor de expansión de los Microdatos de la encuesta, de tal forma que este factor replique las series empalmadas de los deciles del ingreso, los percentiles correspondientes a las incidencias de pobreza y pobreza extrema empalmadas según las líneas de pobreza e indigencia y el total poblacional según las retroproyecciones del CNPV 2018. Por último, el tercer punto consiste en obtener las series de pobreza, pobreza extrema y coeficiente de Gini empalmadas estimándolas de forma directa haciendo uso del factor de expansión con marco 2005 y el factor de ajuste para este resultante del punto 2.*

## 2. Metodología de empalme

### 2.1. Empalme de la serie de ingreso

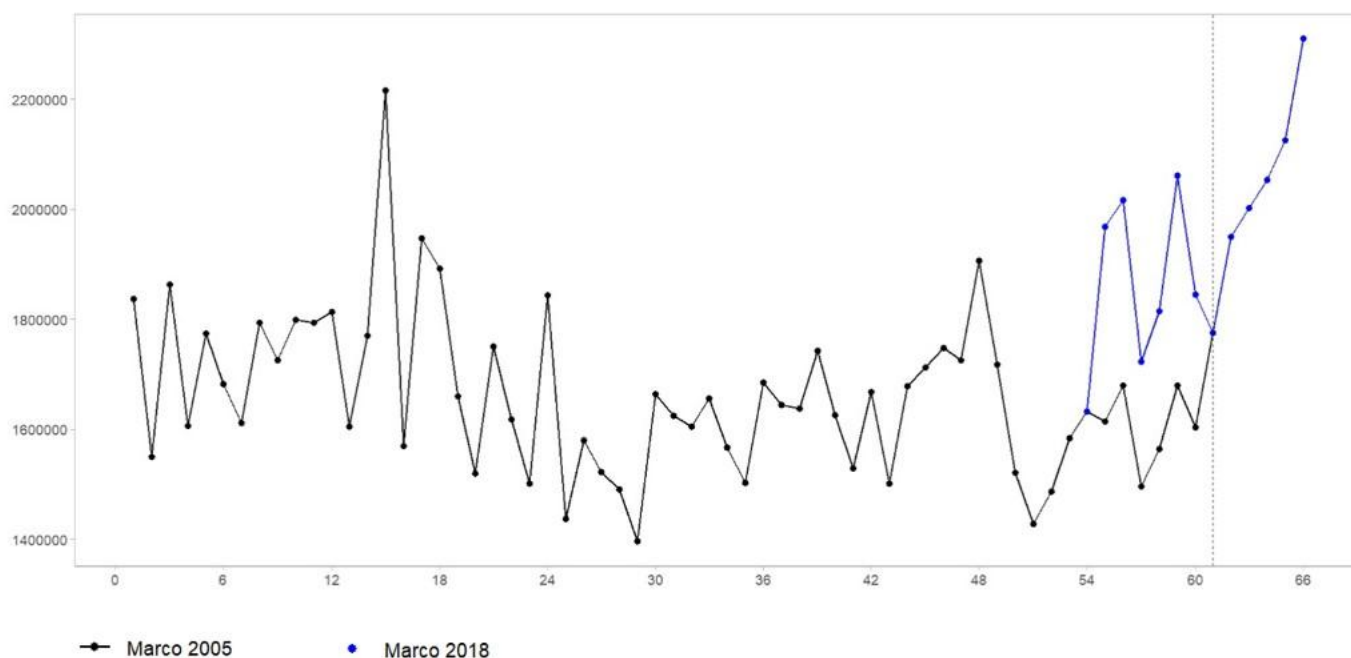
Para la estimación de las incidencias de pobreza monetaria, pobreza monetaria extrema y el coeficiente de Gini es necesario contar con los ingresos a nivel de hogar o persona, para esto se define:

- **Ingreso corriente:** El ingreso corriente incluye los ingresos monetarios y en especie que son devengados de manera constante por el hogar y excluye aquellos ingresos de carácter ocasional. Adicionalmente, la unidad de gasto excluye pensionistas y empleados(as) domésticos(os) del hogar.
- **Ingreso per cápita:** el ingreso per cápita de la unidad de gasto es el resultado de dividir el ingreso corriente disponible de la unidad de gasto entre el número de personas que la conforman.

Para realizar el empalme de las series es necesario contar con un conjunto de puntos ordenados en el tiempo para la construcción del modelo. La GEIH es una encuesta que reporta los indicadores de Mercado Laboral mensualmente y los indicadores de pobreza anual, por lo que con los microdatos anuales se pueden realizar la estimación de los indicadores de pobreza en diferentes periodos en el año para tener un mayor número de puntos en la serie de tiempo.

En primer lugar se realizó la estimación de la pobreza y pobreza extrema mensualmente a nivel de dominio (total cabeceras, total centros poblados y rural disperso y para cada una de las 23 ciudades y áreas metropolitanas), obteniendo en algunos dominios como Chocó y Riohacha márgenes de error superiores al 9 % con un tamaño de muestra promedio por dominio mensual de 780 hogares, por lo cual se toma la decisión de bimensualizar la serie de tal forma que se obtengan márgenes de error inferiores a 6 % y tamaño de muestra promedio de 1.560 hogares por dominio. De esta forma para cada dominio se obtienen once series bivariadas (9 deciles del ingreso, pobreza monetaria y pobreza monetaria extrema), teniendo una serie con el marco 2005 de 60 puntos (desde el año 2012 al año 2021) y 12 puntos con el marco 2018 (desde el año 2021 al 2022), teniendo un traslape de 6 puntos en el año 2021, por la encuesta paralela que se corrió, en el figura 1 se presenta la serie bivariada del decil 9 del ingreso per cápita en el dominio Armenia, la línea negra corresponde a la serie original, correspondiente al marco 2005 y la línea azul representa la serie nueva con marco 2018.

Figura 1: Decil 9 del ingreso per cápita en Armenia



*Fuente: elaboración propia.*

En este caso donde se dispone de una encuesta paralela, es posible ajustar un modelo estructural de series de tiempo para estimar el impacto por el cambio de marco. Los modelos de series de tiempo tienen en cuenta que las series no son determinísticas, sino que tiene un proceso dinámico interno; adicionalmente los modelos estructurales consideran los componentes latentes inherentes en las series de tiempo, tales como la tendencia, la estacionalidad, la dependencia de otra serie de tiempo exógena, y permite la extracción de estas características. De esta manera nos permite estimar y aislar el impacto del cambio de marco sin que esté afectado por otros componentes aleatorios de la serie. Brakel (2007), explica que este tipo de modelos son flexibles para usar en empalmes de series de tiempo teniendo en cuenta que se puede adaptar para series con estacionalidad y series que cuenten con o sin paralela en el punto de empalme.

La estimación de un modelo estructural de series de tiempo suele ser llevada a cabo por medio de la representación como un modelo de estado espacio, o en este caso, un modelo lineal dinámico, y posteriormente se logra la estimación y predicción por medio de la aplicación del filtro de Kalman (para más detalles consultar Petris et al. (2009)).

El modelo lineal dinámico consiste en las dos siguientes ecuaciones:

- **Ecuación de observación** busca expresar la serie observada  $y_t$  (univariada o multivariada) en función de los componentes no observables  $\alpha_t$

$$y_t = F_t \alpha_t + \varepsilon_t$$

donde  $\varepsilon_t \sim N(0, V_t)$ ,  $\alpha_t$  es el vector de estado, que contiene los componentes de la serie que queremos extraer (tendencia, estacionalidad, regresión dinámica, entre otros) y  $F_t$  es una matriz de tamaño apropiado.

- **Ecuación de transición** describe el comportamiento dinámico del vector de estado  $\alpha_t$

$$\alpha_t = G_t \alpha_{t-1} + w_t$$

con  $w_t \sim N(0, W_t)$  y  $G_t$  una matriz cuadrada.

El análisis del anterior modelo lineal dinámico consiste en la estimación de las matrices  $V_t$ ,  $W_t$ , así como los mejores predictores del vector de estado  $\alpha_t$  condicionado en los datos observados. Este objetivo se logra con la implementación del filtro de Kalman.

Ahora, tal como se describió anteriormente, el primer paso en el ejercicio de empalme consiste en empalmar la serie de ingreso, para esto se cuenta con una encuesta paralela, y se tiene un traslape en 6 puntos en la serie se establece el modelo para una serie bivariada ( $y_{1t}$ ,  $y_{2t}$ ) donde  $\{y_{1t}\}$  es la serie de ingreso con el marco 2005 y  $\{y_{2t}\}$  es la serie de ingreso con el marco 2018. Para establecer cuál es el modelo estructural que se ajustaría para los datos se consideran los siguientes aspectos:

- No se considera el componente estacional para la serie de ingreso ya que la prueba de raíz unitaria estacional para dicha serie arroja la ausencia del componente estacional para la gran mayoría de los dominios
- Sí se incluye el componente de la tendencia que describe la fluctuación inherente en el comportamiento del ingreso
- Durante el traslape, entre las dos series  $\{y_{1t}\}$  y  $\{y_{2t}\}$  hay una diferencia debido al cambio del marco, y esta diferencia es mayor a medida que se acerca al final del traslape

Dado lo anterior, el modelo estructural para la serie bivariada del ingreso está dado por la siguiente expresión:

$$y_{1t} = L_t + \varepsilon_{1t}$$

$$y_{2t} = L_t + \beta \delta_t + \varepsilon_{2t}$$

donde  $L_t$  es la tendencia estocástica de la serie, dada por:

$$L_t = L_{t-1} + R_{t-1} + \varepsilon_{1t}$$



$$R_t = R_{t-1} + \varepsilon_{2t}$$

$\beta$  corresponde a la magnitud del efecto del cambio del marco sobre la serie  $y_t$ . Suponiendo que  $y_{1t}$  y  $y_{2t}$  tienen un translope entre el tiempo  $t_1$  y  $t_2$ , la magnitud del efecto  $\delta_t$  se define:

$$\delta_t = \begin{cases} 0, & t \leq t_1 \\ 1 - \frac{t-t_1}{t_2-t_1}, & t_1 < t < t_2 \\ 0, & t \geq t_2 \end{cases}$$

Con el fin de estimar el anterior modelo estructural, se hace la observación de que éste se puede representar por un modelo lineal dinámico definiendo las siguientes matrices:

$$F_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & \delta_t \end{bmatrix}$$

y

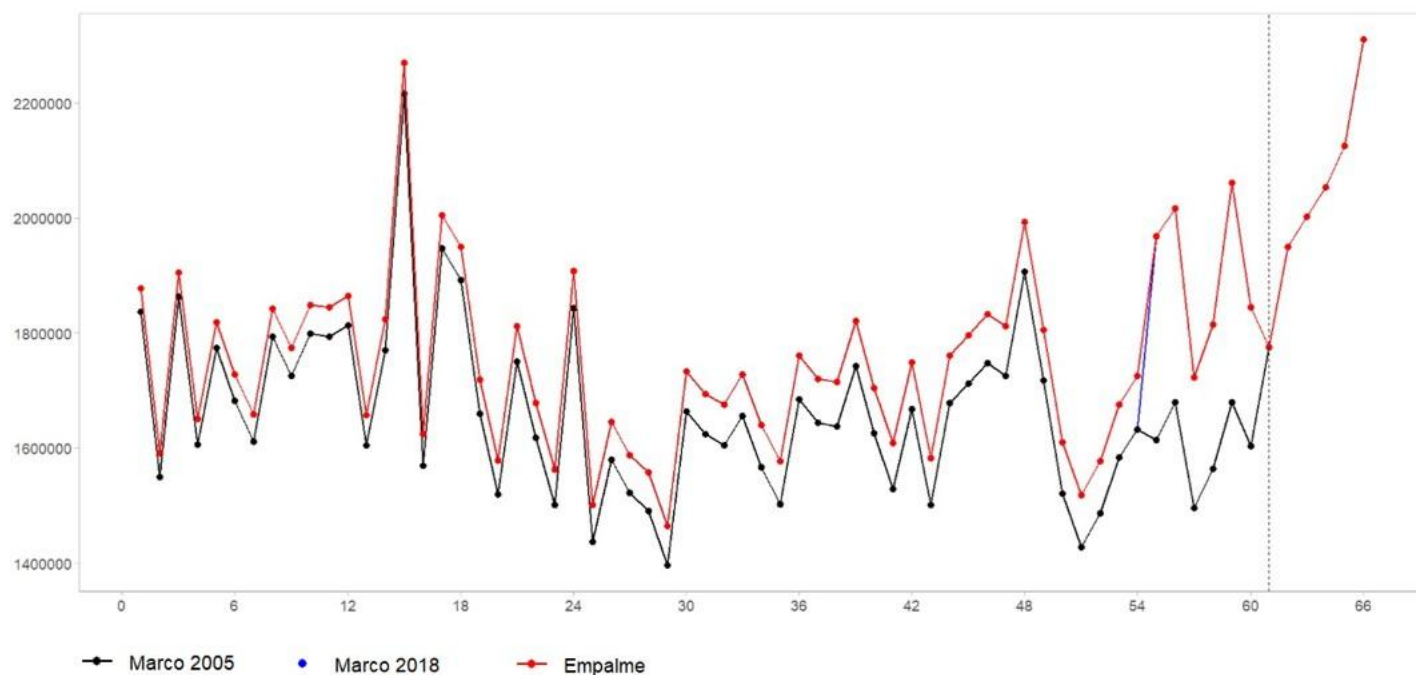
$$G_t = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

La aplicación del filtro de Kalman y la estimación del impacto  $\beta$  por el modelo lineal dinámico se implementa por medio del paquete *dln* en el software estadístico *R* (ver Petris (2010)). El siguiente paso es obtener la serie ajustada empalmada, aplicando el impacto de tal forma que entre más cercano se esté al punto del cambio el impacto sea mayor, entre más lejos se esté de este el impacto debe ser cercano a cero:

$$y_t^* = \begin{cases} y_{1t} + \hat{\beta} * \frac{t-t_1}{t_2-t_1}, & t \leq t_2 \\ y_{2t}, & t > t_2 \end{cases}$$

Aplicando esta metodología para cada serie en los 25 dominios se obtiene una serie empalmada para cada decil, para la incidencia de pobreza y la incidencia de pobreza extrema en cada dominio. La figura 2 presenta la serie empalmada del decil 9 para el dominio Armenia, representando en color negro la serie inicial con el marco 2005, en azul la serie con el marco 2018 la cual se traslapa con la serie empalmada, dado que después del tiempo 54 la serie empalmada coincide con la serie con el marco 2018.

Figura 2: Empalme del decil 9 del ingreso per cápita en Armenia



*Fuente: elaboración propia.*

El resultado final del proceso es obtener una serie empalmada anual a partir del empalme de las series bimensuales, por lo que estas se agregan ponderando cada punto bimensual por la suma del total de la población para obtener la serie anualizada:

$$y_{t,m} = \frac{\sum_{j=1}^6 y_{t,m,j} w_{tj}}{\sum_{j=1}^6 w_{tj}}$$

donde  $y_{t,m}$  denota el valor de la serie del dominio  $m$  en el año  $t$ ,  $y_{t,m,j}$  el valor de la serie del dominio  $m$  en el bimes  $j$  del año  $t$ ,  $w_{tj}$  denota el número de personas de la GEIH del bimes  $j$  del año  $t$ .

## 2.2. Ajuste de los factores de expansión

Dentro de los objetivos del ejercicio de empalme está el proveer a la ciudadanía los microdatos con los cuales puedan replicar el cálculo de los indicadores, es muy importante garantizar que este ejercicio sea

lo más parsimonioso posible. Si se realizara un modelo de espacio estado para el empalme de cada una de las series se obtendría un total de 3 factores de ajuste, uno para cada indicador (pobreza, pobreza extrema y Gini), lo cual no es viable por la dificultad que tendría la ciudadanía al momento de replicar los cálculos, es por esto que se plantea una metodología de calibración, en la cual a partir de un conjunto de restricciones se obtiene un único factor de ajuste con el cual se calculan las series empalmadas de los principales indicadores de pobreza; dado que el único cambio en la encuesta fue el cambio de marco el enfoque propuesto es coherente con este principio ya que se obtiene un único factor de ajuste el cual se aplica al factor de expansión.

La calibración es un ajuste que se realiza a los pesos de muestreo para que las estimaciones de algunas variables de control reproduzcan de forma perfecta normalmente los totales poblacionales de estas variables, sin embargo, dentro de las restricciones a incluir en el ejercicio de calibración no solo está el total poblacional, se encuentran diferentes indicadores por lo que a continuación se presenta la metodología teórica para la obtención del factor de ajuste en cada restricción:

- **Calibración sobre total**

Se tiene una población finita  $U = \{1, 2, \dots, N\}$  sobre la cual se extrae una muestra  $s$  para la cual se observa la variable de interés  $y_k$  y un vector de variables auxiliares  $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{pk})$ , adicional se conoce el total poblacional  $t_x = \sum_U x_k$ , normalmente el objetivo es estimar el total o la media de la variable de interés  $y$  por medio del estimador de Horvitz-Thompson:

$$\hat{t}_{y,\pi} = \sum_{k \in s} \frac{y_k}{\pi_k} = \sum_{k \in s} y_k d_k$$

El objetivo de la calibración es encontrar unos nuevos pesos  $w$  que reproduzcan el total poblacional:

$$\sum_{k \in s} y_k w_k = t_x$$

Dewille & Särndal (1992) definen los pesos de calibración de tal forma que sean lo más cercanos a los pesos de muestreo iniciales  $d_k = 1/\pi_k$  y que minimicen la pseudodistancia  $\Phi_s$  dados por:

$$w_k = d_k F_k(\mathbf{x}_k \lambda)$$

Para la distancia chi-cuadrado  $\Phi_s = \sum_{k \in s} c_k (w_k - d_k)^2 / d_k$  los pesos de calibración se definen:

$$w_k = d_k + (t_x - \hat{t}_{x\pi})' \left( \sum_{k \in s} d_k c_k \mathbf{x}_k \mathbf{x}_k' \right)^{-1} c_k d_k \mathbf{x}_k$$

- **Calibración sobre razón**

Suponga que para cada elemento de la muestra se tienen los vectores  $x_k$  y  $y_k$ , se definen las razones poblacionales  $R_q = \frac{t_{yq}}{t_{xq}}$ ,  $q = 1, \dots, Q$ . El objetivo es estimar los totales  $\hat{t}_{yq} = \sum_{k \in S} w_k y_{qk}$  y  $\hat{t}_{xq} = \sum_{k \in S} w_k x_{qk}$  con los nuevos pesos  $w_k$  que satisfacen la condición

$$\frac{\sum_{k \in S} w_k y_{qk}}{\sum_{k \in S} w_k x_{qk}} = R_q, \quad q = 1, \dots, Q$$

Rojas et al. (2016) demuestran que calibrar sobre  $R$  es equivalente a calibrar sobre el vector  $t_z = 0$ :

$$t_z = (t_{z1}, \dots, t_{zQ})'$$

donde  $t_z = 0$  se calibra por medio de la metodología anterior dado que se está calibrando un total.

- **Calibración sobre percentil**

Beręsewicz y Szymkowiak (2023) exponen la metodología de calibración sobre percentiles, suponga que se desea estimar el percentil  $Q_{y,\alpha}$  de la variable de interés  $y$ ,  $\alpha \in (0,1)$ , el cual se puede expresar como  $Q_{y,\alpha} = \inf\{t \mid F_y(t) \geq \alpha\}$ , donde  $F_y(t) = N^{-1} \sum_{k \in U} H(t - y_k)$  y la función Heavyside dada por:

$$H(t - y_k) = \begin{cases} 1, & \text{si } t \geq y_k \\ 0, & \text{si } y < y_k \end{cases}$$

Se asume que se tiene un vector  $Q_{x,\alpha} = (Q_{x_1,\alpha}, \dots, Q_{x_{J_2},\alpha})'$  de cuantiles poblacionales conocidos para  $J_2$  variables auxiliares, para encontrar los nuevos pesos  $w_k$  que reproduzcan los percentiles poblacionales se estiman a partir de la función  $\hat{F}_{y,cal}(t) = \frac{\sum_{k \in S} w_k H_{y,s}(t, y_k)}{\sum_{k \in S} w_k}$ , de esta forma la función Heavyside es remplazada por  $H_{y,s}(t, y_k)$  dada por:

$$H_{y,s}(t, y_k) = \begin{cases} 1, & y_k \leq L_{y,s}(t), \\ \beta_{y,s}(t), & y_k = U_{y,s}(t), \\ 0, & y_k > U_{y,s}(t) \end{cases}$$

donde  $L_{y,s}(t) = \max\{y_k, k \in s | y_k \geq t\} \cup \{-\infty\}$ ,  $U_{y,s}(t) = \min\{y_k, k \in s | y_k > t\} \cup \{\infty\}$  y  $\beta_{y,s}(t) = \frac{t - L_{y,s}(t)}{U_{y,s}(t) - L_{y,s}(t)}$  para  $k = 1, \dots, n$ ,  $t \in \mathbb{R}$

El estimador de calibración para  $Q_{y,\alpha}$  es  $\widehat{Q_{y,cal,\alpha}} = \widehat{F_{y,cal}^{-1}}(\alpha)$  en el que se obtiene el vector de los nuevos pesos  $\mathbf{w} = (w_1, \dots, w_n)'$  es la solución del problema de optimización  $D(\mathbf{d}, \mathbf{v}) = \sum_{k \in s} d_k G\left(\frac{v_k}{d_k}\right) \rightarrow \min$  condicionado en la calibración a  $\sum_{k \in s} w_k = N$  y  $\widehat{Q_{x,cal,\alpha}} = (\widehat{Q_{x1,cal,\alpha}}, \dots, \widehat{Q_{J_2x,cal,\alpha}}) = Q_{x,\alpha}$  o equivalente a  $\widehat{F_{xj,cal}}(Q_{xj,\alpha}) = \alpha$  para  $j = 1, \dots, J_2$ .

Con el fin de obtener el empalme de la serie del ingreso en términos de diferentes parámetros distribucionales como los percentiles de la distribución, se llevó a cabo diferentes escenarios de calibración con el objetivo de acercarnos al mejor empalme de la distribución del ingreso per cápita y por ende el empalme de los indicadores de pobreza. Esta calibración se realiza de forma independiente a nivel anual en cada uno de los 25 dominios, obteniendo en cada caso un único factor de ajuste para el factor de expansión de los microdatos:

- **Escenario 1:** Se calibra por la media del ingreso per cápita, percentiles correspondientes a la incidencia de pobreza y pobreza extrema empalmadas y total poblacional según las retroproyecciones del Censo Nacional de Población y Vivienda (CNPV) 2018.
- **Escenario 2:** Se calibra por la media del ingreso per cápita del percentil 95%, percentiles correspondientes a la incidencia de pobreza y pobreza extrema empalmadas y total poblacional según las retroproyecciones del Censo Nacional de Población y Vivienda (CNPV) 2018.
- **Escenario 3:** Se calibra por los percentiles 5%, 25%, 50%, 75%, 95% y los correspondientes a la incidencia de pobreza y pobreza extrema empalmadas y total poblacional según las retroproyecciones del Censo Nacional de Población y Vivienda (CNPV) 2018.
- **Escenario 4 - Escenario final:** Se calibra por los deciles del ingreso y los percentiles correspondientes a la incidencia de pobreza y pobreza extrema empalmadas y total poblacional según las retroproyecciones del Censo Nacional de Población y Vivienda (CNPV) 2018.

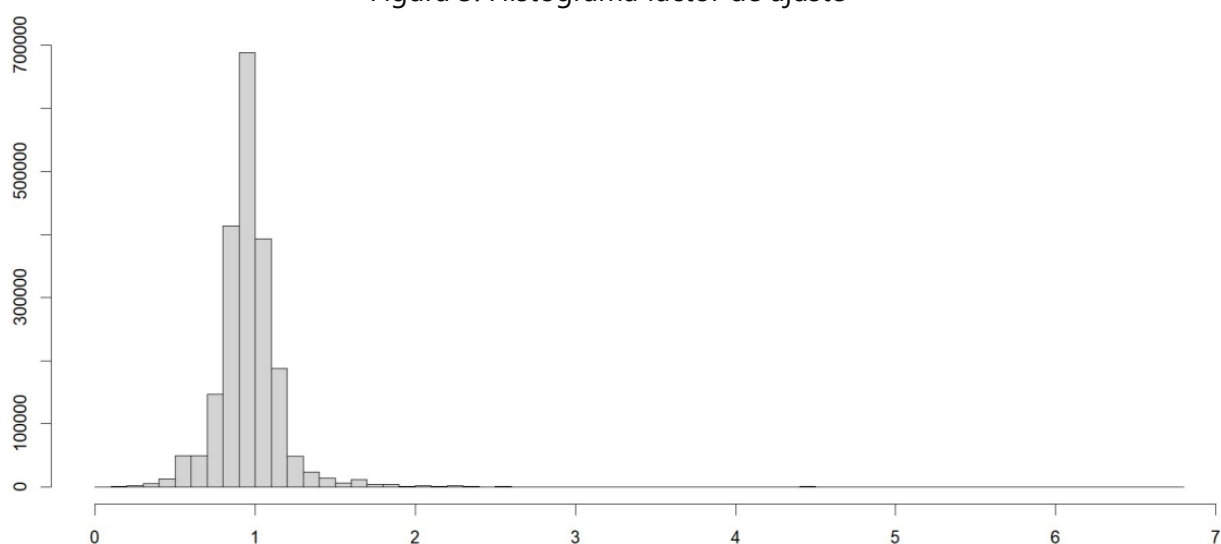
Cada escenario de calibración se evalúa teniendo en cuenta diferentes criterios como la cantidad de factores de ajuste mayor a 2, la distribución de estos, en términos de los valores donde se concentra la mayoría de los datos y el número de cruces que se generan con las series finales, es decir al aplicar el factor de ajuste al factor de expansión y obtener las series de pobreza monetaria, pobreza monetaria extrema y Gini empalmadas se comparan con las series bajo el marco 2005 para ver la tendencia de ambas y el comportamiento que tienen respecto al valor de la encuesta paralela.

Finalmente se escoge el escenario 4 el cual satisface los criterios definidos y provee el mejor escenario de empalme para los tres indicadores. Como se mencionó anteriormente la calibración se hace para cada año y cada dominio, en cada caso se verifica la distribución de los factores de ajuste y en caso de tener valores mayores a 7 o negativos se elimina la restricción de los deciles en la calibración y se conserva la incidencia de pobreza, pobreza extrema y el total poblacional, el cuadro 1 presenta las restricciones que aplican en cada dominio, el símbolo **x** indica que la calibración garantiza los deciles de ingreso, incidencia de pobreza, pobreza extrema y techos poblacionales, el símbolo **o** indica que la calibración garantiza la incidencia de pobreza, pobreza extrema y techos poblacionales.

<b>Dominio</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
Armenia	x	x	x	x	x	x	x	x	x
Barranquilla	x	x	x	x	x	x	x	x	x
Bogotá	x	x	x	x	x	x	x	x	x
Bucaramanga	x	x	x	x	x	x	x	x	x
Cali	x	x	x	x	x	x	x	x	x
Cartagena	x	x	x	o	x	x	x	x	x
Cúcuta	x	x	x	x	x	x	x	x	x
Florencia	x	x	x	x	x	x	x	x	x
Ibagué	x	o	o	x	o	o	x	o	x
Manizales	x	x	x	x	x	x	x	x	x
Medellín	x	x	o	x	o	x	o	o	x
Montería	x	x	x	x	x	x	x	x	x
Neiva	x	o	o	x	x	x	o	o	x
Pasto	x	x	x	x	x	x	x	x	x
Pereira	x	x	x	x	x	x	x	x	x
Popayán	x	x	x	x	x	x	x	x	x
Quibdó	x	x	x	x	x	x	x	x	x
Resto urbano	x	x	x	x	x	x	x	x	x
Riohacha	x	x	x	x	x	x	x	x	x
Rural	x	x	x	x	x	x	x	x	x
Santa Marta	x	o	o	x	o	o	x	o	x
Sincelejo	x	x	x	x	x	x	x	x	x
Tunja	x	x	x	x	x	x	o	o	x
Valledupar	x	o	o	x	o	o	x	o	x
Villavicencio	x	x	x	x	x	x	x	x	x

La figura 3 presenta el histograma de los factores de ajuste obtenidos a partir de la calibración de los 25 dominios en los años 2012-2020, observando que la gran mayoría de los factores se concentran alrededor de 1, indicando que no se realizan grandes cambios a los datos muestrales y tan solo el 0.5% de los factores son mayores a 2.

Figura 3: Histograma factor de ajuste



*Fuente: elaboración propia.*

### 2.3. Cálculo de la incidencia de pobreza, pobreza extrema y Gini

Con la calibración se obtiene un factor de ajuste  $g_k$  para el factor de expansión  $d_k$  en cada hogar de la muestra de la GEIH para los años 2012 a 2022, con este nuevo factor de expansión  $g_k d_k$  se obtienen los valores de la serie empalmada anual de cada indicador usando los siguientes estimadores:

- **Incidencia de pobreza**

$$\hat{p} = \frac{\sum_{k \in S} y_k * d_k * n_{per_k} * g_k}{\sum_{k \in S} d_k * n_{per_k} * g_k}$$

donde  $y_k$  es la variable indicadora (1 si el hogar es pobre, 0 en caso contrario),  $d_k$  es el factor de expansión,  $n_{per_k}$  es el total de personas en el hogar y  $g_k$  el factor de ajuste.

- **Incidencia de pobreza extrema**

$$\hat{p}_{ext} = \frac{\sum_{k \in S} y_k * d_k * n_{per_k} * g_k}{\sum_{k \in S} d_k * n_{per_k} * g_k}$$

donde  $y_k$  es la variable indicadora (1 si el hogar es pobre extremo, 0 en caso contrario),  $d_k$  es el factor de expansión,  $n_{per_k}$  es el total de personas en el hogar y  $g_k$  el factor de ajuste.

- **Total de personas en pobreza**

$$\hat{t}_p = \sum_{k \in S} y_k * d_k * n_{per_k} * g_k$$

donde  $y_k$  es la variable indicadora (1 si el hogar es pobre, 0 en caso contrario),  $d_k$  es el factor de expansión,  $n_{per_k}$  es el total de personas en el hogar y  $g_k$  el factor de ajuste.

- **Total de personas en pobreza extrema**

$$\hat{t}_{p_{ext}} = \sum_{k \in S} y_k * d_k * n_{per_k} * g_k$$

donde  $y_k$  es la variable indicadora (1 si el hogar es pobre extremo, 0 en caso contrario),  $d_k$  es el factor de expansión,  $n_{per_k}$  es el total de personas en el hogar y  $g_k$  el factor de ajuste.

- **Coeficiente de Gini**

Para calcular el coeficiente de Gini se deben seguir los siguientes pasos:

- I. Calcular el total poblacional y el ingreso total
  - a. Total poblacional:  $NPT = \sum_{k \in S} d_k * n_{per_k} * g_k$
  - b. Total ingreso:  $TX = \sum_{k \in S} x_k * d_k * n_{per_k} * g_k$



donde  $x_k$  es el ingreso per cápita de la unidad de gasto,  $d_k$  es el factor de expansión,  $n_{per_k}$  es el total de personas en el hogar y  $g_k$  el factor de ajuste.

- II. Se calcula la variable  $t_{per_k}$  correspondiente al total poblacional por hogar

$$t_{per_k} = n_{per_k} * d_k * g_k$$

- III. Se ordena la base de menor a mayor según el ingreso per cápita  $x_k$ .  
IV. Se calcula la variable  $fex_{acum}$  que corresponde a la suma acumulada de la población por hogar

$$fex_{acum_k} = \sum_{l=1}^k t_{per_l}$$

- V. Se calcula el aporte de cada observación al coeficiente de Gini total

$$gini_k = \frac{t_{per_k}}{NPT} * \frac{2NPT}{TX} * \frac{2fex_{acum_k} - t_{per_k} + 1}{2 * NPT} * \left( x_k - \frac{TX}{NPT} \right)$$

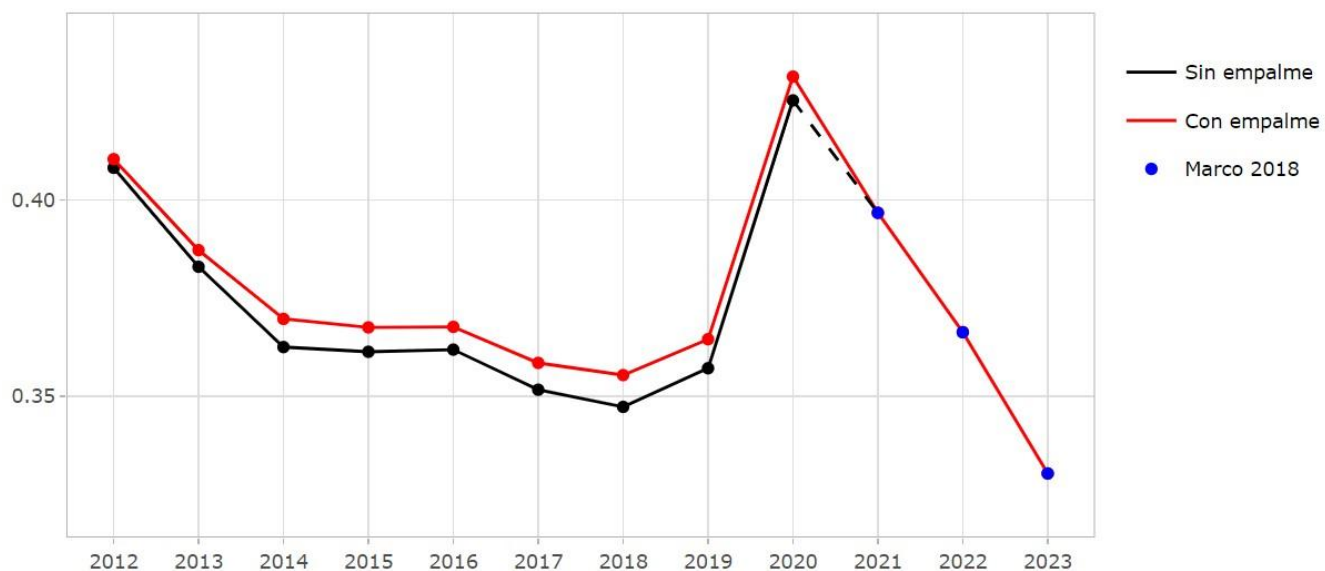
- VI. Finalmente, el coeficiente de Gini se obtiene sumando los valores  $gini_x$  en toda la muestra de hogares

$$\widehat{Gini} = \sum_{k \in S} gini_k$$

## 2.4. Resultados, gráficas

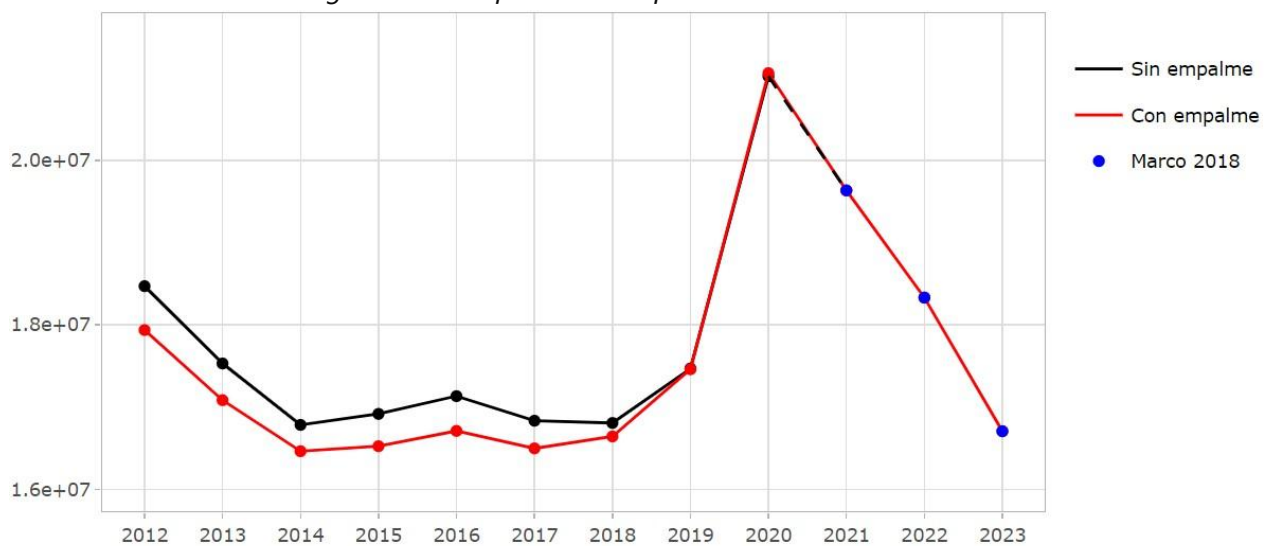
A continuación, se presentan las series empalmadas Nacionales para la incidencia de pobreza, incidencia de pobreza extrema, total de personas en pobreza, total de personas en pobreza extrema y el coeficiente de Gini. Para las series de los totales cabe resaltar que las proyecciones poblacionales con el marco 2018 son menores a las proyecciones poblacionales con el marco 2005, así que, aunque la incidencia de pobreza es mayor el total de pobres es menor dado el cambio en los techos poblacionales.

Figura 4: Pobreza monetaria Nacional



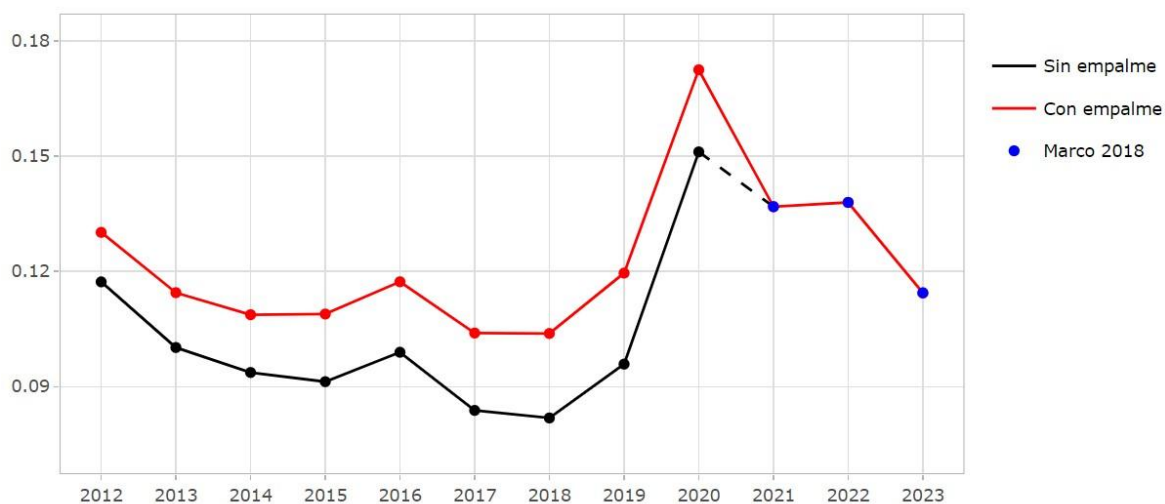
*Fuente: elaboración propia.*

Figura 5: Total personas en pobreza nivel Nacional



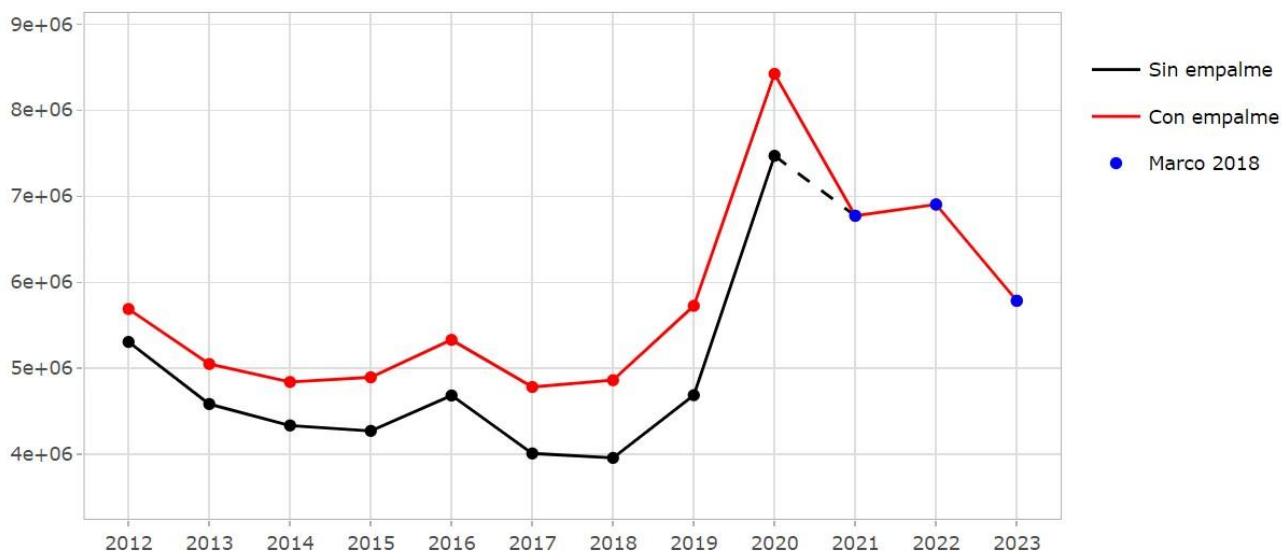
*Fuente: elaboración propia.*

*Figura 6: Pobreza monetaria extrema Nacional*



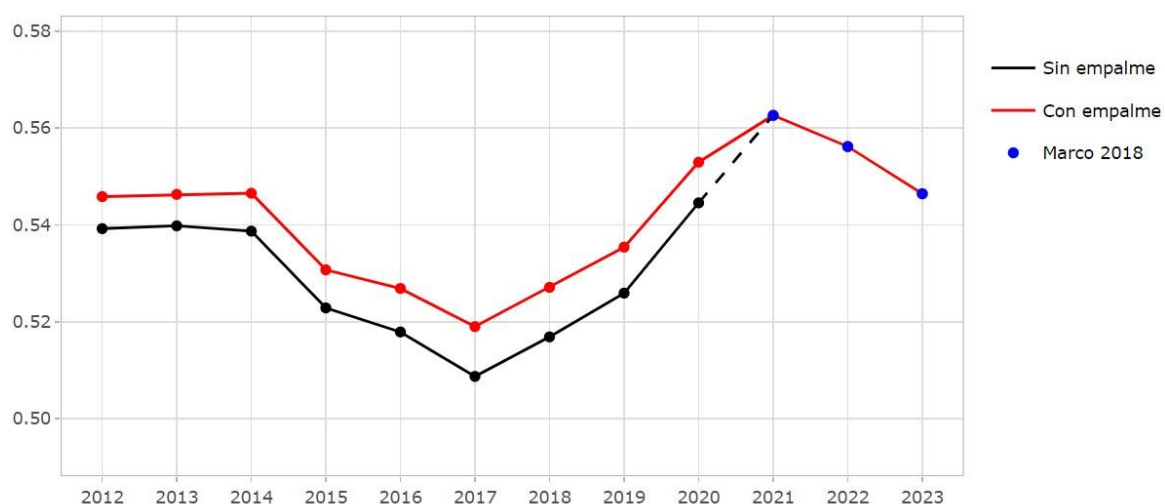
*Fuente: elaboración propia.*

*Figura 7: Total personas en pobreza extrema a nivel Nacional*



*Fuente: elaboración propia.*

*Figura 8: Coeficiente de Gini Nacional*



*Fuente: elaboración propia.*

### 3. Conclusión

En este documento se describe los procedimientos para el ejercicio de empalme para las series de incidencia de pobreza, pobreza extrema y el coeficiente de Gini: consiste en empalmar la serie de ingreso per cápita por medio de un modelo estructural de series de tiempo, posteriormente se calcula un factor de ajuste para los factores de expansión, y finalmente se calculan las series de interés empalmadas utilizando los microdatos en conjunto con los factores de expansión ajustados.

Es importante resaltar que el factor de ajuste  $g_k$  solo aplica para la estimación de los indicadores de pobreza monetaria, pobreza monetaria extrema y el coeficiente de Gini, dado que son los indicadores que dependen del ingreso y sobre los cuales se llevaron a cabo los ejercicios y pruebas en la calibración.

## **4. Anexo**

Los resultados gráficos de las series empalmadas a nivel nacional y en cada uno de los 25 dominios se encuentran en [https://dane-sub.shinyapps.io/Series\\_empalme\\_v2/](https://dane-sub.shinyapps.io/Series_empalme_v2/)

## **5. Referencia**

Beręsewicz, M., & Szymkowiak, M. (2023). A note on joint calibration estimators for totals and quantiles. arXiv preprint arXiv:2308.13281.

Brakel, Jan & Smith, Paul. (2007). Quality procedures for survey transitions - Experiments, time series and discontinuities. Survey Research Methods. Vol 2. No. 3, pp. 123-141

Departamento Administrativo Nacional de Estadística (DANE). Medición de Pobreza Monetaria y Desigualdad 2023.

Fuente: [www.dane.gov.co](http://www.dane.gov.co)

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. Journal of the American statistical Association, 87(418), 376-382.

Gutiérrez, H. A. (2009). Estrategias de muestreo: Diseño de encuestas y estimación de parámetros. Facultad de Estadística, Universidad Santo Tomás.

Petris G (2010). "An R Package for Dynamic Linear Models." Journal of Statistical Software, 36(12), 1–16. <https://www.jstatsoft.org/v36/i12/>.

Petris G, Petrone S, Campagnoli P (2009). Dynamic Linear Models with R, series user! Springer-Verlag, New York.

Rojas, H. A. G., Zhang, H., & Rodríguez, N. A. (2016). The performance of multivariate calibration on ratios, means and proportions. Revista Colombiana de Estadística, 39(2), 281-305.

Särndal, C. E. (2007), 'The calibration approach in survey theory and practice', *Survey Methodology* 33(2), 99–119.

Särndal, C.-E., Swensson, B. & Wretman, J. (2003), *Model assisted survey sampling*, Springer.

Sen, A. (1976). Poverty: an ordinal approach to measurement. *Econometrica: Journal of the Econometric Society*, 219-231.