

Análisis del ingreso corriente total per cápita en los hogares de la Ciudad de México mediante un modelo de regresión Bayesiana en áreas pequeñas.

Saul Arturo Ortiz Muñoz
Sergio Pérez Elizalde
Jesús Emanuel Paredes Romero
Beatriz Juárez Piña

Resumen

Se propone un modelo de regresión Bayesiana variacional con errores normales asimétricos en áreas pequeñas para estudiar el logaritmo del ingreso corriente total per cápita en los hogares de la Ciudad de México, con información proveniente de fuentes oficiales y siguiendo los criterios de procesamiento establecidos por el Coneval. El modelo señala que la variable respuesta presenta distintos grados de sesgo por alcaldía. Se identificaron las covariables más relevantes para los modelos de regresión. Se concluye que el método variacional es una alternativa viable para acelerar la inferencia con respecto a los algoritmos Bayesianos usuales basados en muestreo.

1 Introducción

La medición de la pobreza adquiere relevancia porque permite identificar con mayor precisión las carencias que enfrentan los grupos afectados y orientar intervenciones públicas más efectivas. En México, esta medición estuvo durante años a cargo del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval), cuya metodología retomaba elementos del enfoque propuesto por Alkire y Foster; actualmente, la responsabilidad recae en el Instituto Nacional de Estadística y Geografía (INEGI). Desde una perspectiva integral, la pobreza se relaciona con condiciones de vida que vulneran la dignidad humana, restringen el ejercicio de derechos y obstaculizan la satisfacción de necesidades básicas, dificultando así la plena participación social de las personas Coneval (2023).

Durante décadas, el producto interno bruto per cápita fue utilizado como indicador principal del bienestar económico de la población; sin embargo, su capacidad explicativa resulta limitada para capturar la complejidad de las condiciones de vida contemporáneas. La evaluación del bienestar requiere enfoques más amplios que reconozcan que la pobreza no puede reducirse únicamente al nivel de ingresos, pues intervienen dimensiones que no se reflejan de manera directa en la capacidad

monetaria de los hogares. Desde esta perspectiva multidimensional, surge la necesidad de integrar indicadores que permitan comprender de manera más completa las privaciones que experimentan los individuos. Contar con mediciones adecuadas no solo mejora el diagnóstico social, sino que constituye un elemento clave para diseñar políticas públicas que respondan de manera efectiva a las distintas manifestaciones de la pobreza Sáenz Vela (2020).

Con el propósito de explorar la relación entre el ingreso corriente total per cápita (ICTPC) de los hogares con las características de la población y de las viviendas que habitan, se propone un modelo de regresión que estudie y cuantifique la asociación con respecto a estas covariables, así mismo, se desea modelar también la naturaleza de este conjunto de datos: en términos generales, la distribución empírica del ingreso se caracteriza por estar sesgada a la derecha, es decir, es común observar valores grandes. Así pues, se plantea construir un modelo de regresión que sea capaz de no solo capturar este sesgo, si no también cuantificarlo y aprovecharlo para enriquecer y hacer más robusta la inferencia estadística ante la incertidumbre.

El objetivo de este trabajo es explorar una alternativa para medir el bienestar económico, el cuál es parte de la medición multidimensional de la pobreza (MMP) municipal para el año 2025, el siguiente periodo quinquenal con respecto a la medición previa. Para ello, se realizan predicciones del ICTPC para todos los hogares no muestreados de los cuáles se tiene información auxiliar proveniente del Censo 2020, empleando y procesando los datos de acuerdo a los criterios establecidos por el Coneval. La aportación principal es el uso de errores normales sesgados que aportan flexibilidad y un método variacional para acelerar la inferencia Bayesiana.

El resto del documento está estructurado de la siguiente forma: en la Sección 2 se revisa la distribución normal sesgada, los modelos en áreas pequeñas y la MMP a nivel municipal desarrollada por el Coneval. La Sección 3 se centra en los detalles del modelo propuesto y la obtención del conjunto de datos de acuerdo a los criterios oficiales que El Consejo estableció. Posteriormente, en la Sección 4 se presentan algunos estadísticos descriptivos del ICTPC y los principales resultados obtenidos con la muestra a posteriori aproximada. Finalmente, en la Sección 5 se discuten brevemente los resultados y se presentan las conclusiones.

2 Revisión de literatura

2.1 Distribución normal asimétrica

La distribución normal asimétrica es una densidad continua de tres parámetros con soporte en \mathbb{R} , generaliza a la distribución normal a través de un parámetro de forma que controla el sesgo, lo que le proporciona mayor flexibilidad, además de que hereda algunas de sus propiedades. Estas características la hacen atractiva para incluirla en modelos estadísticos como alternativa a la distribución normal. La variable aleatoria X tiene densidad normal asimétrica o normal sesgada con parámetros de localidad, escala y forma $(\mu, \sigma^2, \lambda) \in \mathbb{R} \times \mathbb{R}_+ \times \mathbb{R}$ si su densidad de probabilidad está

dada por

$$f_X(x|\mu, \sigma^2, \lambda) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\lambda \frac{x-\mu}{\sigma}\right) I_{(-\infty, \infty)}(x), \quad (1)$$

aquí ϕ y Φ denotan la función de densidad y distribución normal estándar, además, escribimos $X \sim SN(\mu, \sigma^2, \lambda)$.¹ Esta representación se conoce como parametrización directa. Sea $\rho \in (-1, 1)$ definido como $\rho = \lambda/\sqrt{1+\lambda^2}$, Azzalini (2013) muestra las siguientes propiedades básicas de X :

- Para $a \in \mathbb{R}$ y $b > 0$, $a + bX \sim SN(a + b\mu, b^2\sigma^2, \lambda)$, es decir, es cerrada bajo transformaciones de localidad y escala.
- $\mathbb{E}[X] = \mu + \sigma\sqrt{\frac{2}{\pi}}\rho^2$. $\text{Var}[X] = \sigma^2(1 - \frac{2}{\pi}\rho^2)$
- El coeficiente de asimetría esta dado por

$$\frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{(\text{Var}[X])^{3/2}} = \sqrt{\frac{2}{\pi}}(4 - \pi)\lambda^3/\pi(1 - \frac{2}{\pi}\lambda^2)^{3/2}.$$

Usualmente, esta cantidad se denota con γ_1 . Bayes y Branco (2007) sugieren que como $\gamma_1 \in (-0.9953, 0.9953)$, esta densidad es adecuada solo para modelar sesgo leve o moderado.

- Si $\lambda = 0$, entonces $SN(x | \mu, \sigma^2, \lambda = 0) \equiv N(x|\mu, \sigma^2)$. Además, si $\lambda \rightarrow \infty$ ($-\infty$), entonces $SN(x | \mu, \sigma^2, \lambda) \rightarrow N(x | \mu, \sigma^2)$ truncada a la derecha (izquierda) en cero. Este es el caso límite.

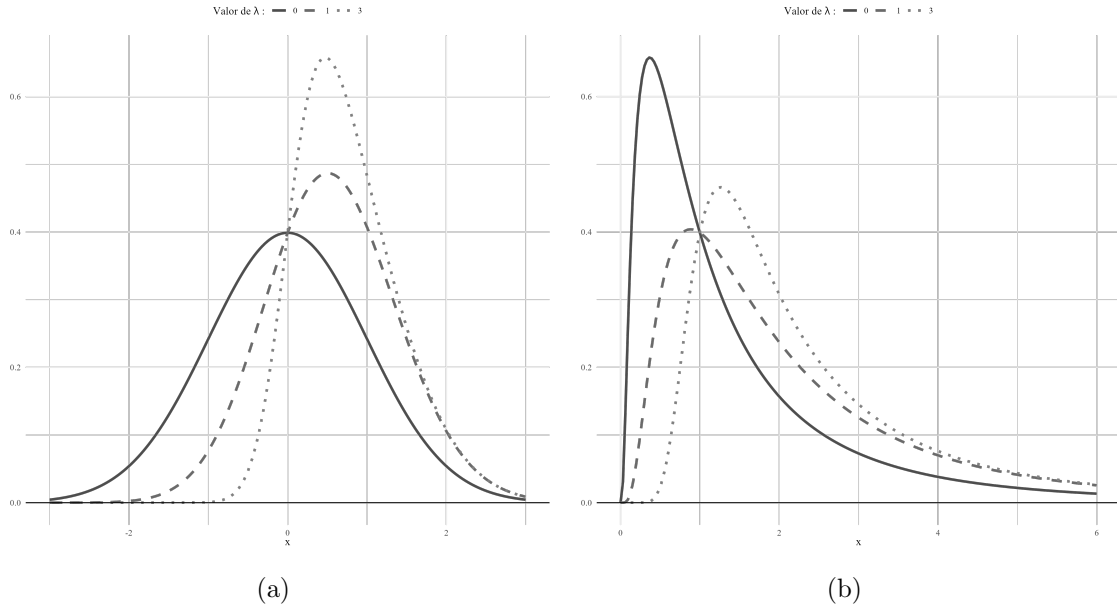


Figura 1: Densidad (log-)normal asimétrica estándar para algunos valores de $\lambda > 0$. Fuente: elaboración propia.

¹SN: skew normal.

A diferencia del caso normal, la media y varianza de esta distribución no coinciden con los parámetros de localidad y escala. Así mismo, la inferencia sobre estos parámetros presenta dificultades prácticas y por ello se propone emplear la parametrización centrada (Azzalini 1985; Bayes y Branco 2007). Sea $Z \sim SN(0, 1, \lambda)$, definimos a X como

$$X = \mu + \sigma \left((Z - \mathbb{E}[Z]) / \sqrt{\text{Var}[Z]} \right), \quad (2)$$

además, si reemplazamos λ por el coeficiente de asimetría γ_1 , entonces X tiene distribución normal asimétrica centrada, y se denota como $X \sim SN^C(\mu, \sigma^2, \gamma_1)$. La función de densidad de X es más extensa, aunque reensambla la Ecuación 1.² En la discusión posterior, únicamente centramos la media y varianza, dejando el parámetro de forma sin modificación. De este modo, es posible mantener sólo un tipo de notación.

2.2 Truncamiento oculto

Una representación estocástica de la densidad normal asimétrica que es útil para la exposición siguiente, es mediante el proceso de truncamiento oculto.³ La utilidad práctica es evitar el uso de la densidad normal sesgada de forma explícita. Ahora, este proceso puede surgir en diversos contextos, quizás incluso de forma inconsciente, suponga que se observan dos variables aleatorias normales correlacionadas X_1 y X_2 , si registramos alguna de estas variables siempre que la otra exceda cierto umbral μ_2 (es decir $Y = X_1$ siempre que $X_2 > \mu_2$), entonces inducimos sesgo en las observaciones retenidas (Y), y este sesgo es proporcional a su correlación ρ , de este modo, las observaciones que filtramos (Y) de acuerdo a la otra variable (X_2) tendrán distribución normal asimétrica; además, este mecanismo puede extenderse a más dimensiones. (Azzalini 2013; Arellano-Valle y Azzalini 2008; Arnold y Beaver 2000).

2.3 Modelos en áreas pequeñas

Cuando se trabaja en el contexto de áreas pequeñas se obtienen dos tipos de estimadores: directos e indirectos. Los estimadores directos llevan a cabo la inferencia con información exclusiva de cada área pequeña, y casi siempre se construyen a partir de un diseño muestral, por ejemplo, encuestas. Por otro lado, los estimadores indirectos toman información de otras áreas pequeñas, de tal modo que pueden hacer inferencias más robustas. De este modo, los estimadores indirectos emplean modelos estadísticos con el fin de hacer predicciones de alguna variable objetivo, empleando información auxiliar (Rao y Molina 2015).

En la estimación de áreas pequeñas, small-area estimation (SAE), no es el área la que es pequeña, sino la muestra tomada del dominio específico: un dominio se considera pequeño si su tamaño de muestra no es lo suficientemente grande como para obtener estimaciones directas con la

²Para más detalles, vea Pérez-Rodríguez et al. (2018) y Azevedo, Bolfarine y Andrade (2011).

³Otra representación estocástica puede consultarse en Bayes y Branco (2007).

precisión adecuada. En caso contrario, se dice que el dominio es grande.⁴ Por lo tanto, el término área pequeña no se limita a una región en el sentido geográfico, sino que se refiere a cualquier subpoblación para la que no se pueden producir estimaciones directas con una precisión adecuada.

Dentro de la corriente de SAE, existen dos niveles principales de estudio, el más general son los modelos a nivel de área, mientras que los más específicos son a nivel de unidad, es decir, la unidad de observación son objetos muestreados dentro de cada dominio.

2.4 Estudios previos

Parte de la literatura que ha estudiado los ingresos desde una perspectiva estadística, se concentra en bondad de ajuste de modelos probabilísticos a datos observados. Por ejemplo: Arnold (2015) propone modelos Pareto generalizados para tal propósito, Montes Rivera, Pérez Rodríguez y Pérez Elizalde (2012) comparan el ajuste del ingreso mediante la distribución Pareto, Lognormal y Dagum, entre otros autores.

Por otro lado, dentro del contexto de áreas pequeñas, se persigue otro objetivo: realizar predicciones y estimaciones para los parámetros del modelo asumido. Battese, Harter y Fuller (1988) propusieron un modelo a nivel unidad con errores anidados, el cuál identificamos como un modelo de regresión lineal con un efecto aleatorio para cada región de estudio: sea $(y_{ij}, \mathbf{x}_{ij})$ el atributo de interés y la información auxiliar asociada, para las regiones $i = 1, 2, \dots, M$ y observación $j = 1, 2, \dots, N_i$, se asume que

$$y_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}, \quad (3)$$

donde $u_i \sim N(0, \sigma_u^2)$, $e_{ij} \sim N(0, \sigma_e^2)$ y además u_i es independiente de e_{ij} , para todo par (i, j) . Adicionalmente, se asume que no existe sesgo, de modo que la inferencia puede realizarse únicamente con los datos observados.⁵ Al momento de calcular los estimadores, el modelo toma información de las demás regiones de estudio, de ahí a que se consideren estimadores indirectos. En el mismo sentido, Diallo y Rao (2018) proponen generalizar este modelo al relajar el supuesto de normalidad: ahora ambos errores pertenecen a la familia normal asimétrica, dotando así de mayor flexibilidad; específicamente, los autores propusieron

$$\begin{aligned} u_i &\sim SN(-\rho_u \sigma_u \sqrt{2/\pi}, \sigma_u^2, \lambda_u) \\ e_{ij} &\sim SN(-\rho_e \sigma_e \sqrt{2/\pi}, \sigma_e^2, \lambda_e) \end{aligned} \quad (4)$$

donde $\rho_u = \lambda_u / \sqrt{1 + \lambda_u^2}$, $\rho_e = \lambda_e / \sqrt{1 + \lambda_e^2}$ y nuevamente u_i es independiente de e_{ij} , para todo par (i, j) . u_i es un efecto aleatorio dentro de cada área pequeña, sin embargo, los parámetros de forma λ_u y λ_e son los mismos para todas las M regiones. Para hacer inferencia sobre el modelo, los autores recurren a una generalización de la distribución normal sesgada llamada Closed Skew Normal (CSN), la cuál viene acompañada de mayor complejidad computacional.

⁴Hedlin (2008) menciona que una expresión mejor podría ser: ‘estimación de muestra pequeña’.

⁵Para más detalle sobre los aspectos de inferencia, puede consultar Rao y Molina (2015).

La densidad log-normal sesgada no es la única alternativa para modelar datos de ingreso, y en general, respuestas estrictamente positivas; sin embargo, esta distribución hereda algunas propiedades que resultan útiles desde el punto de vista práctico: por ejemplo, es cerrada bajo transformaciones de cambio de localidad y escala, además, en el contexto de regresión, conserva la interpretación de los coeficientes β como el cambio promedio al observar un incremento unitario -o bien, si las variables están estandarizadas, en desviaciones estándar-, de las covariables, una a la vez.

2.5 Metodología para la medición multidimensional de la pobreza

El Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) fue un organismo autónomo y descentralizado cuya tarea principal consistía en la generación e implementación de diversas técnicas estadísticas para cuantificar la pobreza desde varios ámbitos, abordando así la medición multidimensional de la pobreza (MMP) en México. La última edición sobre la MMP que realizó el Coneval fue en 2022, siendo ahora el Instituto Nacional de Estadística y Geografía (INEGI) el organismo encargado de darle continuidad a esta medición. En 2024 el INEGI publicó su primera edición de la MMP tomando como base los mismos lineamientos propuestos por el Coneval. Estos lineamientos tienen sustento en los artículos 36 y 37 de la Ley General de Desarrollo Social (LGDS) y determinan que la MMP se construye a partir de dos fuentes de datos oficiales y productos generados de estas: (1) el Censo de Población y Vivienda (Censo) y (2) la Encuesta Nacional de Ingresos y Gastos en los Hogares (ENIGH).

Los lineamientos establecen que la MMP debe tener una periodicidad mínima bienal en los niveles estatal y nacional, y quinquenal a nivel municipal. Para este último caso resulta complejo realizar la MMP dado que la ENIGH solo es representativa a nivel estatal y nacional, y por tanto, se hace uso de otros métodos de estimación. En 2020, Coneval construyó la MMP a nivel municipal a partir del ajuste de modelos en áreas pequeñas para realizar predicciones sobre atributos de interés, entre ellos, el ingreso corriente total per cápita, mediante la técnica del mejor predictor heterocedástico empírico, empirical best predictor heterokedastic, (EBPH), la cual está basada en la teoría de modelos lineales mixtos y es equivalente al planteamiento en la Ecuación 3, pero permitiendo que la varianza de los hogares sea distinta entre sí, de ahí su nombre.

El objetivo principal de la MMP es estimar el número y porcentaje de la población en situación de pobreza. Para realizar esta estimación, se establece que la pobreza se compone de dos espacios analíticos principales: la dimensión de derechos sociales o carencias, que consta de seis indicadores dicotómicos: rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda y acceso a la alimentación; y el espacio de bienestar económico, cuyo indicador es el ICTPC. Al incorporar estos dos rubros, se clasifica a la población en uno de cuatro cuadrantes de pobreza (INEGI 2025; Coneval 2021):

- I. Población en situación de pobreza multidimensional: población con ingreso inferior al valor

monetario de las líneas de pobreza por ingresos⁶ (LPI) respectivas y con al menos una carencia social, en este cuadrante, la población se desagrega en dos grupos:

- I'. Población en situación de pobreza extrema: su ingreso es inferior al valor monetario de las líneas de pobreza extrema por ingresos⁷ (LPEI) y presenta tres o más carencias sociales.
- I'. Población en situación de pobreza moderada: percibe un ingreso inferior a las LPI y presenta entre una y dos carencias.
- II. Población vulnerable por carencias sociales: población con una o más carencias sociales, pero cuyo ingreso es igual o superior a las LPI respectivas,
- III. Población vulnerable por ingresos: población sin carencias sociales y con ingreso inferior a las LPI respectivas,
- IV. Población no pobre multidimensional y no vulnerable: población con ingreso igual o superior a las LPI respectivas y sin ninguna carencia social.

Es relevante señalar que los indicadores de seguridad social, alimentación e ingreso no pueden obtenerse directamente a partir de información del Censo, por lo que necesitan ser estimados a través de modelos de regresión en áreas pequeñas (Coneval 2021). En este sentido, el modelo propuesto permite abordar la dimensión de la pobreza que corresponde a la vulnerabilidad por ingresos.

3 Metodología

A continuación, se resume el modelo propuesto de acuerdo a los siguientes puntos

- Se emplean errores log-normales sesgados para la variable objetivo, es decir, se asume que el logaritmo del ICTPC sigue una distribución normal asimétrica.
- Se considera un contexto de áreas pequeñas: se desea realizar predicciones para los objetos de estudio a lo largo de todos los dominios, a partir de una muestra de solo algunas regiones.
- Se emplea un enfoque Bayesiano objetivo: las distribuciones a priori que se consideran son no informativas, lo cuál maximiza la información proveniente de la muestra, a la vez que se pretende realizar un análisis reproducible y neutral.
- Se utiliza un algoritmo automático de inferencia variacional. Este método es fácil de implementar y permite acelerar la inferencia Bayesiana.

⁶Valor monetario de la canasta alimentaria más el valor monetario de la canasta no alimentaria: 3296.92 y 4564.97 pesos mexicanos para los ámbitos rural y urbano.

⁷Valor de la canasta alimentaria: 1800.55 y 2354.65 pesos mexicanos para los ámbitos rural y urbano.

3.1 Planteamiento del modelo

El modelo propuesto pertenece a la corriente de estimación indirecta a nivel unidad: se supone que la variable objetivo sigue una distribución log-normal sesgada, lo cual intenta mimetizar dos aspectos inherentes al conjunto de datos de ingresos: siempre son positivos y su distribución empírica es asimétrica. Para la observación (i, j) se plantea:

$$y_{ij} \sim SN(\mu_i + \mathbf{x}_{ij}^T \boldsymbol{\beta} - \sigma \rho_i \sqrt{2/\pi}, \sigma^2/(1 - 2\rho_i^2/\pi), \lambda_i), \quad (5)$$

y de forma equivalente:

$$y_{ij} = \mu_i + \mathbf{x}_{ij}^T \boldsymbol{\beta} + \sigma \frac{e_{ij} - \mathbb{E}[e_{ij}]}{\sqrt{\text{Var}[e_{ij}]}} \quad (6)$$

donde $e_{ij} \sim SN(0, \sigma^2, \lambda_i)$. Es decir, si \tilde{y}_{ij} es la respuesta en escala original, se modela $\log(\tilde{y}_{ij}) \equiv y_{ij}$. Ahora, ya que centramos los parámetros de localidad y escala, garantizamos que $\mathbb{E}[y_{ij}] = \mu_i + \mathbf{x}_{ij}^T \boldsymbol{\beta}$ y $\text{Var}[y_{ij}] = \sigma^2$. A diferencia del modelo anterior, aquí se estima un parámetro de forma para cada área pequeña, de este modo se obtiene un resultado similar a incorporar un efecto aleatorio para cada región, sin embargo, se prioriza la estimación del parámetro de asimetría para cada región; μ_i es un intercepto para cada área. Esta representación es más simple que el modelo de Diallo y Rao (2018), ya que no incorporamos la estructura de errores anidados, es decir, efectos aleatorios, y no es necesario recurrir a la familia CSN.

Finalmente, el último ingrediente que se considera es el paradigma Bayesiano. Para esta aplicación en particular, permite generar resultados más ricos en contenido que el tratamiento frecuentista, ya que es posible estimar fácilmente funciones de interés a partir de la muestra a posteriori. Sin embargo, el costo de oportunidad es la complejidad computacional⁸ y la necesidad de especificar distribuciones a priori, y que de acuerdo a la elección de estas, conduce a seguir un enfoque Bayesiano objetivo o no objetivo. Aquí se propone un enfoque objetivo con el fin de generar resultados reproducibles e imparciales. El vector de parámetros de interés es $\boldsymbol{\theta} = (\rho_{1:M}, \sigma^2, \mu_{1:M}, \boldsymbol{\beta})$, el modelo propuesto admite la siguiente estructura jerárquica

⁸No obstante, no siempre es el caso, por ejemplo, con modelos conjugados o simples.

$$\begin{aligned}
y_{ij} \mid \rho_i, \sigma^2, w_i, \mu_i, \boldsymbol{\beta} &\sim N\left(\mu_i - \mathbf{x}_{ij}^T \boldsymbol{\beta} - w_i \rho_i + \frac{\sigma \sqrt{\frac{2}{\pi} \rho_i^2}}{\sqrt{1 - \frac{2}{\pi} \rho_i^2}}, \frac{\sigma^2(1 - \rho_i^2)}{\sqrt{1 - \frac{2}{\pi} \rho_i^2}}\right) \\
w_i \mid \rho_i, \sigma^2 &\sim NT\left(0, \frac{\sigma^2}{\sqrt{1 - \frac{2}{\pi} \rho_i^2}}\right) \\
p(\sigma^2, \rho_1, \dots, \rho_M) &\propto \frac{1}{\sigma^2} \prod_{i=1}^M \frac{\sqrt{1 + \rho_i^2}}{1 - \rho_i^2} \\
p(\mu_1, \dots, \mu_M) &\propto 1 \\
p(\boldsymbol{\beta} \mid \gamma_{1:p}) &\propto \prod_{k=1}^p \text{MixNorm}(\beta_k \mid 0, \tau^2 c^2, 0, c^2, \gamma_k) \\
p(\gamma_{1:p}) &= \prod_{k=1}^p \text{Be}(\gamma_k \mid 0.5, 0.5),
\end{aligned} \tag{7}$$

Además de centrar la media y varianza, ya se empleó el método de truncamiento oculto para expresar la distribución de y_{ij} . La a priori en $\rho_{1:M}$ consiste en la distribución de referencia o mínimamente informativa que obtuvo López (2024).⁹ De acuerdo con (George y McCulloch 1993), la a priori para $\boldsymbol{\beta}$ permite realizar selección de variables con búsqueda estocástica, stochastic search variable selection (SSVS): a partir de la muestra a posteriori de γ_k , la probabilidad de inclusión de la covariable k , se genera una muestra de las covariables seleccionadas mediante la simulación de una variable aleatoria indicadora o Bernoulli. Así, los modelos más prominentes tendrán mayor frecuencia de aparición. Los hiperparámetros c y τ se establecen de acuerdo con algún criterio práctico, aquí se fijó $(c^2, \tau^2) = (1/300^2, 3000^2)$. Lo que significa a priori que una estimación de β_k igual a cero será generada por $N(0, 1/300^2)$, mientras que una estimación diferente de cero será generada por $N(0, 100)$.

⁹Para más información, consulte Berger y Bernardo (1992).

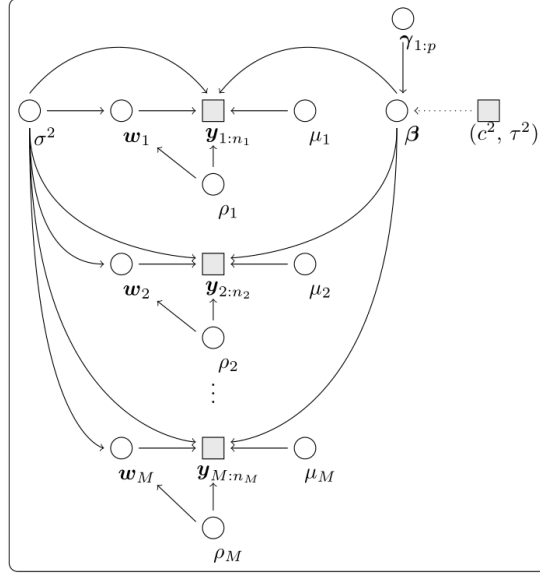


Figura 2: Grafo dirigido acíclico que corresponde a la representación jerárquica del modelo en la Ecuación 7. Las líneas punteadas relacionan parámetros e hiperparámetros. Fuente: elaboración propia.

El teorema o regla de Bayes proporciona una regla formal para actualizar el conocimiento acerca de las cantidades de interés desconocidas, es decir, los parámetros del modelo. Esta regla dice que la distribución a posteriori es proporcional a la verosimilitud multiplicada por la información a priori, simbólicamente

$$p(\boldsymbol{\theta} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \quad (8)$$

A partir de la representación jerárquica del modelo en la Ecuación 7 podemos obtener, salvo una constante, la distribución a posteriori. Para completar la especificación del modelo, señalamos que que se restringe el parámetro de correlación ρ_i al intervalo $(0, 1)$, es decir que $\lambda_i \in (0, \infty)$; esto debido a que la distribución empírica del ingreso sugiere que está sesgado a la derecha, tal restricción también alivia costo computacional, ya que se evita explorar la mitad del soporte.

3.2 Fuentes de información y procesamiento

Las estimaciones con modelos de regresión en áreas pequeñas se realizaron empleando datos oficiales: el Censo 2020 y la ENIGH 2025. Esta última consta de diecisiete productos o tablas: viviendas, hogares, población, gastos monetarios y no monetarios en los hogares, entre otras. Así mismo, la desagregación más pequeña en estas tablas es a nivel vivienda, seguido por los hogares y finalmente se encuentra la población. Por su parte, se dispone de datos del Censo hasta nivel de hogar.

Para actualizar el conjunto de datos de hogares con la información proveniente de la ENIGH 2024, fue necesario realizar ligeros ajustes al código Stata que genera las tablas para realizar la

MMP municipal de la pobreza¹⁰¹¹¹². De este modo, el conjunto de datos conformado por la variable objetivo e información de covariables auxiliares (conteos, indicadores, categorías) fue generado de acuerdo a los mismos criterios que el Coneval usó en el año 2020 para la medición municipal de la pobreza.

Tras integrar las fuentes de información, el conjunto de datos de la Ciudad de México se compone de 79,881 observaciones a nivel hogar, de las cuáles únicamente se tiene registro del ICTPC en 2,329 hogares adicionales. La base de datos está compuesta de 52 covariables continuas (conteos, porcentajes, categorías ordenadas), 81 covariables binarias (indicadores de carencias) y 7 covariables categóricas (más de dos niveles no ordenados). Un análisis de componentes principales sobre las 52 covariables continuas -no reportado aquí- señala que es posible reducir la dimensión a 26 componentes recuperando hasta el 95.27% de la estructura de covarianzas, por tal motivo, se empleó esta técnica. En el Cuadro 8 se listan las covariables incluidas. De igual modo, se omitió el uso de las siete covariables categóricas por simplicidad en el análisis.

El ICTPC se obtiene a partir de la MMP 2024 elaborada por el INEGI, para su construcción, se considera el cociente entre ingreso corriente total (ICT) del hogar y el tamaño del hogar ajustado. El ingreso corriente se compone del ingreso monetario: salarios, transferencias, rentas, entre otros; y el ingreso no monetario: pagos y regalos en especie. La ENIGH levanta registro de los ingresos percibidos por los individuos encuestados hasta seis meses previos a la entrevista. El ICT se obtiene sumando los ingresos monetarios y no monetarios promedio percibidos durante este periodo de tiempo por cada integrante del hogar, analizados a precios constantes del 2018. Luego, para obtener el ICTPC, se reescala el ICT de acuerdo al tamaño del hogar ajustado, esto es, a cada integrante del hogar se le asigna un peso entre (0, 1) de acuerdo al grupo etario al que pertenece, estos pesos están distribuidos como se indica en el Cuadro 1

Cuadro 1: Fuente: elaboración propia con información del programa de cómputo de la MMP 2024.

Tamaño ajustado	Condición
0.7031	Menor de seis años
0.7382	Mayor de seis y menor de trece años
0.7057	Mayor de trece años y menor de diecinueve años
0.9945	Mayor de veinte años y datos perdidos
1.000	Único integrante del hogar

¹⁰Pobreza a nivel municipal: <https://www.coneval.org.mx/Medicion/Paginas/Pobreza-municipio-2010-2020.aspx>

¹¹MMP estatal 2020 (Coneval): https://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza_2022.aspx

¹²MMP 2024 estatal (INEGI): <https://www.inegi.org.mx/desarrollosocial/pm/>

3.3 Ruta de trabajo

Como se mencionó previamente, el objetivo de los modelos en áreas pequeñas es producir estimaciones para aquellas regiones con tamaños de muestra insuficientes para realizar estimación directa, e incluso donde no se dispone de observaciones sobre la variable objetivo. No obstante, en este caso se tienen observaciones para todos los $M = 16$ dominios. Por tanto, en este escenario se propuso la siguiente ruta de trabajo:

- Entrenar al modelo con información de catorce regiones y predecir las dos restantes. De forma concreta, se propone predecir el log-ICTPC para Miguel Hidalgo y Milpa Alta. De acuerdo con el Cuadro 2, en estas dos demarcaciones se registra los ICTPC promedio más grande y pequeño.
- Repetir el paso anterior con el método Hamiltoniano Monte Carlo (HMC) a fin de comparar las métricas de ajuste y el tiempo de ejecución promedio.¹³
- Ajustar el modelo con todas las observaciones disponibles a fin de calcular y reportar estimaciones a posteriori; además, con esta muestra aproximada es posible generar otros productos relevantes como medidas de desigualdad y estimación del porcentaje de la población con ICTPC por debajo de la línea de pobreza.

Cuando se tiene registro de observaciones en la región i ($n_i > 0$), el pronóstico de $h(y_{ij}^*)$ se genera a partir de la muestra a posteriori e integración Monte Carlo como:

$$p(h(y_{ij}^*) | \mathbf{y}^s) = \int p(h(y_{ij}^*) | \boldsymbol{\theta}) p(\boldsymbol{\theta} | \mathbf{y}^s) d\boldsymbol{\theta} \approx \frac{1}{S} \sum_{k=1}^S h(p(y_{ij}^* | \boldsymbol{\theta}^{(k)})), \quad (9)$$

donde $p(y_{ij}^* | \boldsymbol{\theta})$ es el modelo muestral, es decir, el mismo mecanismo estocástico propuesto para generar a los datos \mathbf{y}^s , y S denota el número de muestras a posteriori. No obstante, si $n_i = 0$, el pronóstico de y_{ij}^* presenta un desafío mayor, ya que si bien σ^2 , y $\boldsymbol{\beta}$ toman información directamente de otras áreas pequeñas, los efectos a nivel de área ρ_i únicamente son informados de manera indirecta, lo cuál puede ocasionar que los parámetros se encojan o bien colapsen hacia los límites de su soporte. En este escenario, la alternativa usual es realizar agrupamiento o pooling por medio de una estructura jerárquica en los parámetros propios de cada región.

Por otro lado, para calcular los porcentajes de la población por debajo de las líneas de pobreza, dado el pronóstico de los y_{ij}^* provenientes del censo, se identifica a los hogares con ingresos inferiores a las líneas de pobreza de acuerdo al ámbito, urbano o rural, y a continuación se multiplica el tamaño del hogar por su factor de expansión asociado. Al sumar este atributo, se obtiene la medición municipal del total de personas con esta carencia. Sólo se tomó en cuenta los pronósticos generados a partir de la información del Censo 2020, es decir que se excluyó a los valores del ICTPC que brinda

¹³HMC es un método de muestreo de la a posteriori basado en Markov Chain Monte Carlo (MCMC), la alternativa Bayesiana usual.

la MMP 2024, ya que los factores de expansión de ambas fuentes de información representan a toda la ciudad: en otras palabras, no se duplicó la información.

4 Principales hallazgos

4.1 Análisis descriptivo

En la Figura 3 se muestra la densidad estimada del log-ICTPC, en la izquierda se agrupan los datos de toda la Ciudad y en la derecha se agrupan de acuerdo al ámbito (urbano y rural). Los datos de toda la ciudad indican que la variable respuesta está sesgada hacia la derecha. Por otro lado, cuando clasificamos de acuerdo al tipo de ámbito, el conjunto de hogares urbanos exhiben mayor sesgo, es decir, se observan valores más grandes, mientras que los ingresos en los hogares rurales están más concentrados.

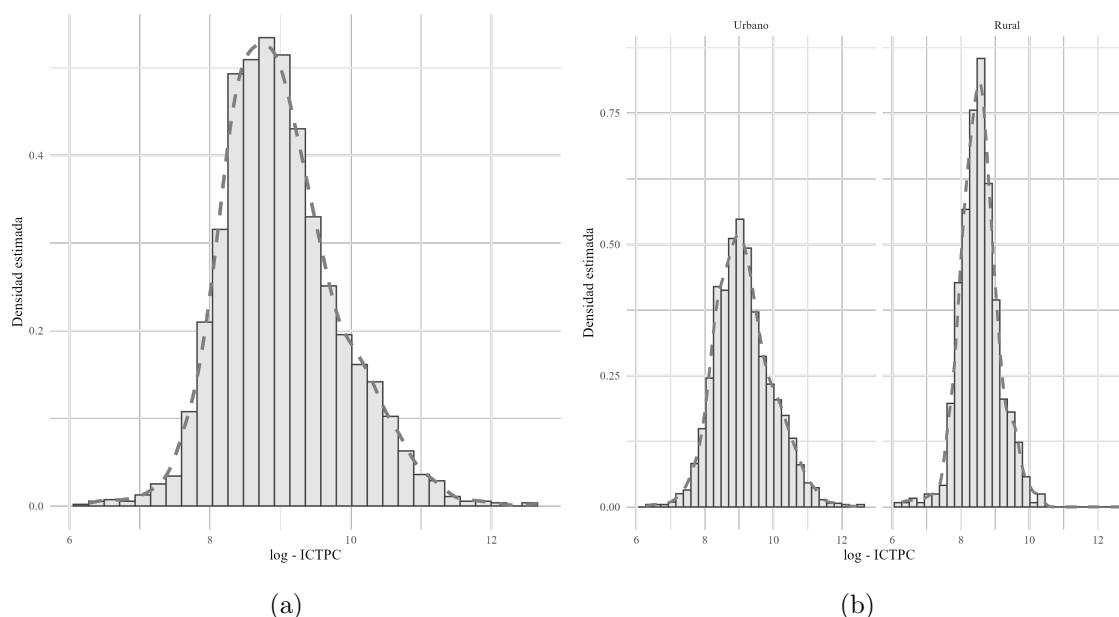


Figura 3: Estimación por kernel (suavizado) de la densidad empírica del logaritmo natural de la respuesta (log-ICTPC) e histograma de probabilidad. Se omitieron tres observaciones asociadas a ingresos pequeños (52.66, 187.71 y 367.30 pesos mexicanos). Fuente: elaboración propia.

En el Cuadro 2 se muestran estadísticos resumen básicos sobre el ICTPC por alcaldía y por ámbito, notamos que sólo en las demarcaciones Milpa Alta, Tláhuac, Tlalpan y Xochimilco, se tiene registro de hogares que pertenecen al ámbito rural, sumando a un total de 553 hogares, el 21.8% con respecto al total de hogares encuestados (2,329). En la Cuadro 3, se muestran los estadísticos resumen agrupados de acuerdo al tipo de ámbito; por ejemplo, la desviación estándar en el ámbito rural es de aproximadamente 17,800 pesos mexicanos, mientras que en el ámbito urbano es de 4,000. Así, pese a que el ICTPC en este ámbito sea en promedio menor, tienen menos dispersión. Sin

embargo, de acuerdo con el índice de Gini, mayor dispersión no necesariamente implica que exista un mayor grado de desigualdad.

Cuadro 2: Estadísticos resumen del ICTPC por alcaldía y ámbito. Fuente: elaboración propia a partir de información de la ENIGH 2024.

Alcaldía	Ámbito	Mínimo	Mediana	Media	Máximo	D. est.	n_i
Azcapotzalco	Urbano	2102.456	12090.792	16910.051	114668.00	16878.831	111
Azcapotzalco	Rural						
Coyoacán	Urbano	1551.186	11696.878	14677.289	83211.33	12202.847	105
Coyoacán	Rural						
Cuajimalpa de Morelos	Urbano	3177.718	8461.970	14948.206	118785.32	19527.767	53
Cuajimalpa de Morelos	Rural						
Gustavo A. Madero	Urbano	781.046	7787.525	10774.117	68570.14	9824.513	252
Gustavo A. Madero	Rural						
Iztacalco	Urbano	631.123	8913.196	13545.988	61666.93	12877.532	105
Iztacalco	Rural						
Iztapalapa	Urbano	187.717	6073.853	8171.228	48248.66	7074.950	349
Iztapalapa	Rural						
La Magdalena Contreras	Urbano	1504.451	5995.553	10735.292	72000.85	12073.719	68
La Magdalena Contreras	Rural						
Milpa Alta	Urbano	1277.369	6171.656	7005.241	21981.60	4342.122	31
Milpa Alta	Rural	529.397	4949.803	5483.564	17037.63	2809.900	227
Álvaro Obregón	Urbano	1564.368	9434.721	13817.397	118931.95	14118.201	207
Álvaro Obregón	Rural						
Tláhuac	Urbano	1245.704	7960.860	8400.618	21017.26	4244.966	52
Tláhuac	Rural	2531.590	5178.865	6949.424	29950.02	5877.352	23
Tlalpan	Urbano	2163.907	9171.878	13563.610	65103.11	12886.281	117
Tlalpan	Rural	720.764	5180.475	6142.782	21935.70	3720.346	162
Xochimilco	Urbano	1430.432	6970.185	10052.865	56342.32	9781.255	87
Xochimilco	Rural	1069.667	4786.025	6545.860	29559.91	5240.275	141
Benito Juárez	Urbano	3741.365	20065.175	25167.655	176719.22	22441.338	135
Benito Juárez	Rural						
Cuauhtémoc	Urbano	921.870	8436.753	17636.413	313206.74	35445.246	115
Cuauhtémoc	Rural						
Miguel Hidalgo	Urbano	2290.712	17156.079	25668.338	132651.56	23268.993	96
Miguel Hidalgo	Rural						
Venustiano Carranza	Urbano	52.668	7880.444	15085.567	297123.05	32880.344	101
Venustiano Carranza	Rural						

Cuadro 3: Estadísticos resumen del ingreso corriente total per cápita por alcaldía y ámbito. Fuente: Elaboración propia con información de la ENIGH 2024.

Ámbito	Minimo	Mediana	Media	Maximo	D. est.	n_i
Urbano	52.668	8520.740	13769.613	313206.74	17793.618	1984
Rural	529.397	5032.302	6008.504	29950.02	3979.968	553

4.2 Precisión y tiempo de ejecución

El ajuste HMC se ejecutó con relativamente pocas iteraciones MCMC (10,000 de burn-in y 2,000 iteraciones de muestreo con thin de 2), generando 1,000 muestras de la a posteriori. Para el método BV también se generó una muestra de tamaño 1,000. En la Figura 4 se grafican los valores observados contra los ajustados obtenidos con el método BV y HMC en escala logaritmo, además de una recta con su tendencia general junto a la identidad: si la predicción fuera perfecta, todos los puntos caerían sobre esta línea.

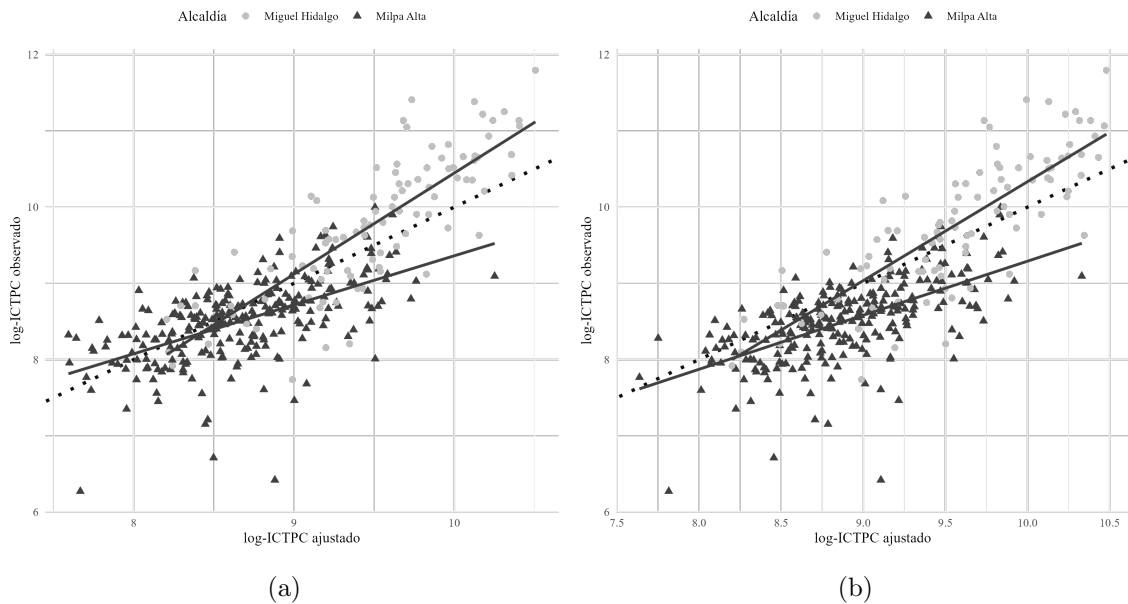


Figura 4: Valores observados vs valores ajustados con el modelo propuesto y usando el método BV (izquierda) y HMC (derecha). El valor de la correlación de Pearson es de 0.7935 para el método BV y 0.7876 para el método HMC. Fuente: elaboración propia.

En el Cuadro 4 se calculan algunas métricas básicas sobre el desempeño de ambos modelos. El error absoluto medio (MAE) señala las unidades en que las predicciones están erradas, la raíz del error cuadrado medio (RMSE) se interpreta de forma similar al MAE, pero añade más penalización a los errores grandes. El error porcentual absoluto medio (MAPE) indica el porcentaje promedio en que los valores ajustados se alejan de los valores reales. El método HMC muestra valores grandes

de estas métricas debido a que el muestreo no ha convergido. Esto ilustra la eficiencia del método BV en términos del tiempo de ejecución y precisión.

El método HMC se considera como un algoritmo MCMC eficiente, ya que incorpora información del gradiente de la función de log-verosimilitud (Stan Development Team 2025); sin embargo, en este ajuste es aproximadamente 80 veces más lento que la aproximación BV, sin una mejora sustancial de las métricas de ajuste. En las estimaciones del parámetro de forma, se obtuvo $\tilde{\rho}_8 = 0.998$ y $\tilde{\rho}_{15} = 0.790$ para el método HMC, frente a $\tilde{\rho}_8 = 0.125$ y $\tilde{\rho}_{15} = 0.172$ obtenidas con el método variacional. Sin presencia de información adicional, el método BV encoge este parámetro, lo que genera predicciones más simétricas para estas alcaldías.

Cuadro 4: Métricas de ajuste en la escala original. Fuente: elaboración propia.

Métrica	Definición	Método VB	Método HMC
Corr.	$\frac{\sum_{i,j} (y_{ij} - \bar{y}_{ij}) (\tilde{y}_{ij} - \bar{\tilde{y}}_{ij})}{\sqrt{\sum_{i,j} (y_{ij} - \bar{y}_{ij})^2 \sum_{i,j} (\tilde{y}_{ij} - \bar{\tilde{y}}_{ij})^2}}$	0.7928	0.7540
MAE	$\sum_{i=1}^M \sum_{j=1}^{N_i} y_{ij} - \tilde{y}_{ij} $	5,411.4858	6297.2901
RMSE	$\left(\sum_{i=1}^M \sum_{j=1}^{N_i} (y_{ij} - \tilde{y}_{ij})^2 \right)^{1/2}$	9,981.7878	10,732.6546
MAPE	$\frac{1}{\sum_i N_i} \sum_{i=1}^M \sum_{j=1}^{N_i} \frac{ y_{ij} - \tilde{y}_{ij} }{y_{ij}}$	76.88%	94.55%
Tiempo		66 s	2890 s

4.3 Estimaciones de los parámetros

A continuación se presentan las medias de las estimaciones, los errores estándar y los intervalos creíbles basados en la aproximación Bayesiana variacional (BV) de la distribución a posteriori. En el Cuadro 5 se muestran las estimaciones de los parámetros $\rho_{1:16}$, σ^2 y $\mu_{1:16}$. De igual modo, en el Cuadro 8 se muestran las estimaciones de los parámetros β , junto a su porcentaje de aparición en la muestra a posteriori. Por cuestiones prácticas, únicamente se listan las covariables cuya frecuencia de aparición es mayor al 75%, en cuyo caso, corresponde a 77 covariables.

Cuadro 5: Elaboración propia basada en la muestra a posteriori.

Alcaldía	Par.	Media (Error est.)	Intervalo c. (2.5 %, 97.5 %)
Azcapotzalco	ρ_1	0.5987 (0.1384)	(0.3112, 0.8350)
Coyoacán	ρ_2	0.6655 (0.0996)	(0.4605, 0.8363)
Cuajimalpa de Morelos	ρ_3	0.7016 (0.1171)	(0.4325, 0.8913)
Gustavo A. Madero	ρ_4	0.7399 (0.0432)	(0.6496, 0.8143)
Iztacalco	ρ_5	0.3121 (0.1904)	(0.0470, 0.7545)
Iztapalapa	ρ_6	0.7099 (0.0391)	(0.6339, 0.7805)
La Magdalena Contreras	ρ_7	0.8276 (0.0495)	(0.7132, 0.9063)
Milpa Alta	ρ_8	0.8547 (0.0185)	(0.8179, 0.8897)
Álvaro Obregón	ρ_9	0.7589 (0.0454)	(0.6647, 0.8399)
Tláhuac	ρ_{10}	0.9213 (0.0175)	(0.8798, 0.9505)
Tlalpan	ρ_{11}	0.7854 (0.0329)	(0.7157, 0.8446)
Xochimilco	ρ_{12}	0.7271 (0.0512)	(0.6196, 0.8157)
Benito Juárez	ρ_{13}	0.6530 (0.0921)	(0.4633, 0.8098)
Cuauhtémoc	ρ_{14}	0.1763 (0.1427)	(0.0193, 0.5465)
Miguel Hidalgo	ρ_{15}	0.6480 (0.0984)	(0.4549, 0.8277)
Venustiano Carranza	ρ_{16}	0.1626 (0.1131)	(0.0275, 0.4492)
	σ	0.6141 (0.0080)	(0.5989, 0.6299)
	μ	8.6619 (0.0106)	(8.6410, 8.6823)

4.4 Medidas de desigualdad

Con la muestra a posteriori, se genera un pronóstico de los ingresos con datos del Censo 2020, y a partir de esta, se calculan medidas de desigualdad como la curva de Lorenz y el índice de Gini. Estas dos métricas están relacionadas entre sí, ya que el índice es función del área bajo la recta de igualdad perfecta (identidad) y la curva de desigualdad. Se estiman estas medidas para los hogares de acuerdo al tipo de corte al que pertenecen: rural o urbano. Se encontró que el índice de Gini para el ámbito urbano es de 0.3679, mientras que para el ámbito rural es de 0.3113. Así mismo, el índice para ambos ámbitos es de 0.3684.

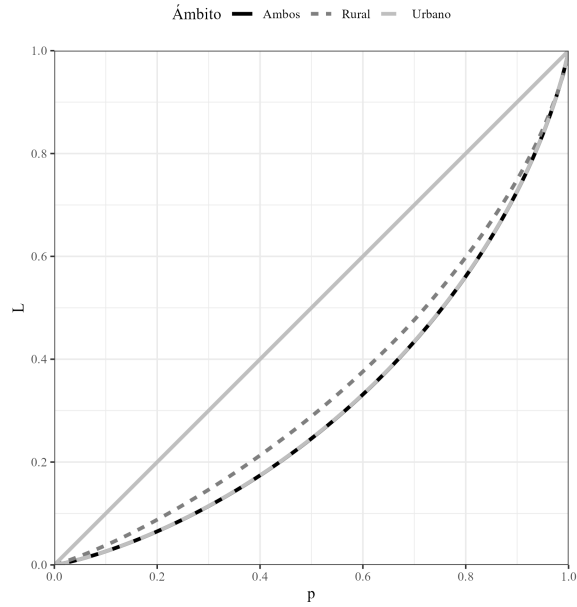


Figura 5: Curva de Lorenz para los valores del ICTPC ajustado de acuerdo al ámbito y ponderado por los factores de expansión provenientes del Censo 2020. Fuente: elaboración propia a partir de la muestra a posteriori.

4.5 Población con ingresos por debajo de las líneas de pobreza

En el Cuadro 6 se reportan los porcentajes de la población con ingresos por debajo de las líneas de pobreza, agrupados por alcaldía. Los porcentajes estatales estimados son de 24.38% y 3.73%, mientras que los porcentajes estatales oficiales registrados en la MMP 2024 de la Ciudad de México son de 24.5% y 4.5%.

Cuadro 6: Elaboración propia basada en la muestra a posteriori.

Alcaldía	Población bajo la LPI (%)	Población bajo la LPEI (%)
Azcapotzalco	12.32	0.74
Coyoacán	13.38	2.64
Cuajimalpa de Morelos	16.97	2.14
Gustavo A. Madero	25.50	3.35
Iztacalco	16.96	2.30
Iztapalapa	37.28	6.88
La Magdalena Contreras	30.22	4.28
Milpa Alta	53.22	9.22
Álvaro Obregón	22.74	2.73
Tláhuac	44.46	7.92
Tlalpan	24.89	2.88
Xochimilco	35.17	5.33
Benito Juárez	1.51	0.27
Cuauhtémoc	9.93	0.78
Miguel Hidalgo	5.18	0.28
Venustiano Carranza	18.86	3.04

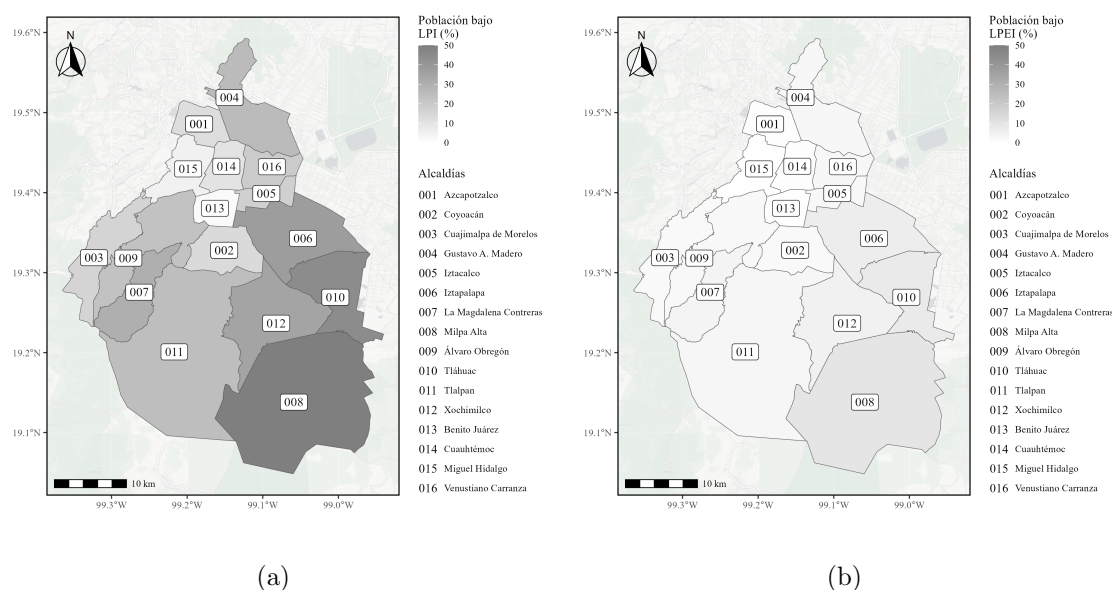


Figura 6: Mapa con el porcentaje de la población con ICTPC por debajo de la línea de pobreza por ingresos (LPI) en la derecha y pobreza extrema por ingresos (LPEI) en la izquierda, en las 16 alcaldías de la Ciudad. Fuente: elaboración propia basada en la muestra a posteriori.

4.6 Sesgo promedio

Una forma adicional de comparar el ajuste, es mediante el cálculo del sesgo en la estimación del ICTPC entre los valores observados y los ajustados por el modelo, agrupados de acuerdo a la alcaldía. El promedio en la diferencia absoluta de los sesgos es de 895.11 pesos mexicanos. Se resaltan los valores más grandes y pequeños tanto para los valores observados, ajustados y el promedio de la diferencia absoluta.

Cuadro 7: Sesgo promedio agrupado por alcaldía. Elaboración propia basada en la muestra a posteriori.

Alcaldía	Media ICTPC observado	Media ICTPC ajustado	Diferencia absoluta
Azcapotzalco	16910.051	16274.642	635.409
Coyoacán	14677.289	15059.292	382.003
Cuajimalpa de Morelos	14948.206	14014.203	934.003
Gustavo A. Madero	10774.117	10558.893	215.224
Iztacalco	13545.988	13693.790	147.802
Iztapalapa	8171.228	8221.276	50.048
La Magdalena Contreras	10735.292	10436.516	298.776
Milpa Alta	5666.401	6123.678	457.277
Álvaro Obregón	13817.397	13849.550	32.154
Tláhuac	7955.585	8259.454	303.869
Tlalpan	9254.742	8971.716	283.026
Xochimilco	7884.059	8023.998	139.939
Benito Juárez	25167.655	24282.706	884.950
Cuauhtémoc	17636.413	13460.428	4175.985
Miguel Hidalgo	25668.338	23096.471	2571.867
Venustiano Carranza	15085.567	12276.001	2809.566

5 Conclusiones

Un aspecto clave sobre el modelo propuesto, es la interpretación del parámetro de forma. En este escenario, puede entenderse como una medida de la heterogeneidad o desigualdad en los ingresos dentro de una alcaldía, y con ello, de la población en esta. Debido a la dualidad de este parámetro en el modelo propuesto, puede estudiarse como correlación, es decir, de forma aproximada, valores menores a 0.5 son considerados sesgo leve, de 0.5 a 0.8 sesgo moderado, y mayores a 0.8 como sesgo grande.

Las estimaciones del parámetro de correlación en el Cuadro 5, indican que en las alcaldías Tláhuac (0.9204), Milpa Alta (0.8606), La Magdalena Contreras (0.8458) y Tlalpan (0.7889), se

observan los valores de sesgo más grandes. Estas alcaldías se ubican al sureste, sur y suroeste de la Ciudad, y constituyen zonas periféricas. Además, tres de estas cuatro alcaldías contienen poblaciones de contextos urbanos y rurales. Así mismo, el modelo encontró que la distribución del ingreso es más simétrica en las alcaldías Venustiano Carranza (0.1638) y Cuahutémoc (0.1775), localizadas en el centro y poniente de la Ciudad.

Los resultados muestran que las alcaldías con los niveles más bajos del ICTPC presentan también sesgos grandes en las estimaciones del ingreso. Este hallazgo sugiere que las estimaciones directas resultan menos confiables precisamente en los territorios más vulnerables, lo que evidencia una limitación metodológica relevante en la medición de la pobreza a escalas territoriales pequeñas. En este sentido, la aplicación de modelos en áreas pequeñas permite reducir dichos sesgos y mejorar la precisión de las estimaciones, contribuyendo a una representación más adecuada de las desigualdades territoriales del ingreso en la Ciudad de México.

El hecho de que las alcaldías con mayores niveles de pobreza presentan también mayores sesgos en las estimaciones del ingreso, es consistente con la existencia de una mayor brecha salarial al interior de estos territorios. Una mayor heterogeneidad de ingresos incrementa la varianza y reduce la representatividad de las estimaciones directas. En este contexto, los modelos en áreas pequeñas contribuyen a capturar de manera más adecuada dicha heterogeneidad, mejorando la precisión de la medición del ingreso a nivel municipal.

(Interpretación sobre los β)

Por otra parte, una desventaja del método de estimación propuesto, es que puede llegar a ser sobreconfidente, es decir, reportar errores estándar pequeños e intervalos creíbles más angostos de como son en realidad. Para las regiones sin observaciones, no se dispone de información para aprender directamente el efecto que tiene el parámetro de forma, por lo que en este escenario, las estimaciones del log-ICTPC tienden a ser más simétricas. Una área de oportunidad para esta metodología, consiste en proponer una distribución a priori que permitan compartir información acerca del parámetro de forma.

Referencias

- Arellano-Valle, R. B. y A. Azzalini (2008). “The centred parametrization for the multivariate skew-normal distribution”. En: *Journal of Multivariate Analysis* 99.7. Special Issue: Multivariate Distributions, Inference and Applications in Memory of Norman L. Johnson, págs. 1362-1382. ISSN: 0047-259X. DOI: <https://doi.org/10.1016/j.jmva.2008.01.020>.
- Arnold, B. C. (2015). *Pareto Distributions*. 2nd. Boca Raton, FL: CRC Press (Taylor & Francis).
- Arnold, B. C. y R. J. Beaver (2000). “Hidden Truncation Models”. En: *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* 62.1, págs. 23-35. ISSN: 0581572X.

- Azevedo, C. L., H. Bolfarine y D. F. Andrade (2011). “Bayesian inference for a skew-normal IRT model under the centred parameterization”. En: *Computational Statistics & Data Analysis* 55.1, págs. 353-365. ISSN: 0167-9473. DOI: <https://doi.org/10.1016/j.csda.2010.05.003>.
- Azzalini, A. (1985). “A Class of Distributions Which Includes the Normal Ones”. En: *Scandinavian Journal of Statistics* 12.2, págs. 171-178. ISSN: 03036898, 14679469. URL: <http://www.jstor.org/stable/4615982>.
- Azzalini, A. (2013). *The Skew-Normal and Related Families*. Institute of Mathematical Statistics Monographs. Cambridge University Press.
- Battese, G. E., R. M. Harter y W. A. Fuller (1988). “An Error-Components Model for Prediction of County Crop Areas Using Survey and Satellite Data”. En: *Journal of the American Statistical Association* 83.401, págs. 28-36. ISSN: 01621459, 1537274X.
- Bayes, C. L. y M. D. Branco (2007). “Bayesian inference for the skewness parameter of the scalar skew-normal distribution”. En: *Brazilian Journal of Probability and Statistics* 21.2, págs. 141-163. ISSN: 01030752, 23176199. URL: <http://www.jstor.org/stable/43601095>.
- Berger, J. O. y J. M. Bernardo (ago. de 1992). “On the Development of Reference Priors*”. En: *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting, Dedicated to the memory of Morris H. DeGroot, 1931–1989*. Oxford University Press. ISBN: 9780198522669. DOI: 10.1093/oso/9780198522669.003.0003.
- Coneval (2021). *Metodología para la medición de la pobreza en los municipios de México, 2020*. México: Coneval.
- (2023). *Metodología para la medición multidimensional de la pobreza en México*. 3ª ed. México: Coneval.
- Diallo, M. S. y J. N. K. Rao (2018). “Small area estimation of complex parameters under unit-level models with skew-normal errors”. En: *Scandinavian Journal of Statistics* 45.4, págs. 1092-1116. DOI: <https://doi.org/10.1111/sjos.12336>.
- George, E. I. y R. E. McCulloch (1993). “Variable Selection Via Gibbs Sampling”. En: *Journal of the American Statistical Association* 88.423, págs. 881-889. ISSN: 01621459, 1537274X.
- Hedlin, D. (2008). “Small Area Estimation: a Practitioner’s Appraisal”. En: *Rivista Internazionale di Scienze Sociali* 116.4, págs. 407-417. ISSN: 0035676X, 18277918.
- INEGI (ago. de 2025). *Análisis de los resultados de la medición de la pobreza multidimensional, 2024*. Inf. téc. Reporte de Resultados 27/25. Instituto Nacional de Estadística y Geografía. URL: https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2025/pm/pm2025_RR_08.pdf.
- López, A. R. (2024). “Estimación Bayesiana de un modelo de áreas pequeñas con distribución asimétrica en el error”. Tesis doct. Colegio de Postgraduados.
- Montes Rivera, F. Y., P. Pérez Rodríguez y S. Pérez Elizalde (jul. de 2012). “Income adjustment in Mexico with a Bayesian approach”. En: *Estudios Económicos de El Colegio de México* 27.2, págs. 273-293. DOI: 10.24201/ee.v27i2.90.

- Pérez-Rodríguez, P. et al. (mayo de 2018). “A Bayesian Genomic Regression Model with Skew Normal Random Errors”. En: *G3 Genes—Genomes—Genetics* 8.5, págs. 1771-1785. ISSN: 2160-1836. DOI: 10.1534/g3.117.300406.
- Rao, J. N. K. e I. Molina (2015). *Small Area Estimation*. 2ª ed. Hoboken, NJ: John Wiley & Sons, pág. 480. ISBN: 9781118735787. DOI: 10.1002/9781118735855.
- Sáenz Vela, H. M. (2020). “Metodología AF: Pobreza multidimensional en México, 2008 y 2018”. En: *Economía Informa* 420.enero-febrero, págs. 48-62. URL: <https://www.economia.unam.mx/assets/pdfs/econinfo/420/05MetodologiaAF.pdf>.
- Stan Development Team (2025). *Stan Reference Manual, Version 2.37*. en. URL: <https://mc-stan.org/docs/reference-manual/>.

Anexo

Lista de covariables empleadas

Cuadro 8: Estimaciones y lista de covariables incluidas, únicamente se muestran aquellas covariables con frecuencias de aparición mayores al 75%. Fuente: elaboración propia basada en la muestra a posteriori.

Par. (Frecuencia)	Media (Error est.)	Intervalo c. (2.5 %, 97.5 %)	Descripción
β_6 (100.00%)	0.0755 (0.0110)	(0.0541, 0.0962)	Número de cuartos del hogar
β_7 (100.00%)	-0.1138 (0.0086)	(-0.1301, -0.0970)	Cantidad de equipamientos (tinaco boiler)
β_{11} (100.00%)	-0.0896 (0.0115)	(-0.1128, -0.0683)	Número de hombres en el hogar
β_{14} (100.00%)	-0.0584 (0.0109)	(-0.0793, -0.0373)	Número de bienes de equipamiento (regadera)
β_{15} (100.00%)	0.0445 (0.0077)	(0.0302, 0.0594)	Escolaridad del jefe
β_{16} (100.00%)	-0.1351 (0.0114)	(-0.1585, -0.1138)	Escolaridad relativa del jefe del hogar
β_{17} (100.00%)	0.1237 (0.0109)	(0.1026, 0.1439)	Edad del jefe del hogar
β_{18} (100.00%)	-0.1520 (0.0107)	(-0.1736, -0.1316)	Escolaridad relativa estandarizada del jefe del hogar
β_{25} (100.00%)	0.2237 (0.0111)	(0.2035, 0.2454)	Número de personas con 60 años o más
β_{30} (100.00%)	0.1178 (0.0110)	(0.0967, 0.1388)	Número de hijos(as) nacidos (as) vivos en el hogar
β_{33} (100.00%)	-0.0488 (0.0109)	(-0.0691, -0.0271)	Número de menores de 12 años en el hogar
β_{36} (100.00%)	-0.0954 (0.0259)	(-0.1459, -0.0448)	Número de perceptores de ingresos ocupados
β_{39} (92.20%)	-0.0510 (0.0416)	(-0.1369, 0.0292)	Personas con carencia en servicios de salud
β_{41} (90.40%)	-0.0467 (0.0431)	(-0.1301, 0.0388)	Porcentaje de hombres en el hogar
β_{42} (100.00%)	0.3439 (0.0921)	(0.1656, 0.5267)	Porcentaje de personas indígenas en el hogar
β_{43} (100.00%)	0.2234 (0.0152)	(0.1927, 0.2538)	Porcentaje de menores de 12 años en el hogar
β_{44} (100.00%)	0.0938 (0.0214)	(0.0528, 0.1370)	Porcentaje de mujeres en el hogar
β_{46} (100.00%)	0.1022 (0.0127)	(0.0790, 0.1290)	Porcentaje de perceptores ocupados
β_{47} (100.00%)	0.0591 (0.0135)	(0.0333, 0.0854)	Porcentaje de personas mayores de 16 años
β_{48} (91.20%)	-0.0493 (0.0346)	(-0.1181, 0.0159)	Porcentaje de personas menores de 16 años
β_{52} (100.00%)	0.2021 (0.0579)	(0.0872, 0.3262)	Proxy de mortalidad infantil
β_{54} (100.00%)	0.1571 (0.0150)	(0.1270, 0.1852)	Indicador de posesión de automóvil en el hogar
β_{58} (95.90%)	0.0594 (0.0292)	(0.0018, 0.1178)	Dispone de bomba de agua el hogar
β_{59} (100.00%)	-0.1162 (0.0141)	(-0.1429, -0.0883)	Dispone de calentador solar de agua el hogar
β_{60} (90.50%)	-0.0393 (0.0670)	(-0.1641, 0.0895)	Dispone de teléfono móvil o celular el hogar
β_{61} (87.00%)	0.0325 (0.0416)	(-0.0514, 0.1096)	Hogar con todos sus integrantes mayores de 16 años
β_{62} (78.50%)	0.0197 (0.0328)	(-0.0438, 0.0840)	Hogar nuclear con algún hijo en la PEA
β_{63} (100.00%)	0.0498 (0.0108)	(0.0297, 0.0709)	Hogar nuclear sin hijos con sus integrantes
β_{64} (97.30%)	0.1308 (0.0848)	(-0.0419, 0.2934)	Hogar nuclear con hijos menores, sin hijos mayores
β_{65} (99.70%)	-0.0716 (0.0204)	(-0.1117, -0.0328)	Dispone de cisterna el hogar
β_{66} (95.80%)	0.0848 (0.1602)	(-0.2274, 0.3976)	Indicador de disponibilidad de cuarto para dormir
β_{67} (93.50%)	0.0687 (0.1134)	(-0.1488, 0.2967)	Tipo de combustible para cocinar
β_{69} (100.00%)	0.2791 (0.0256)	(0.2312, 0.3283)	Dispone de al menos una computadora en el hogar

Método Bayesiano variacional e implementación

Para obtener estimaciones de los parámetros de interés, se empleó el algoritmo inferencia variacional con diferenciación automática, automatic differentiation variational inference, (ADVI). Este método se considera automático en el sentido de que el usuario sólo debe especificar un modelo y los datos, sin preocuparse por más aspectos inferenciales o de programación. Este se implementa en el lenguaje de programación probabilística Stan a través de la interfaz cmdstanr del lenguaje R. Una característica atractiva de Stan es que permite implementar una amplia variedad de modelos Bayesianos de forma sencilla.

El enfoque Bayesiano Variacional (BV) aproxima alguna densidad objetivo p , por ejemplo, la a posteriori $p(\boldsymbol{\theta} \mid \mathbf{y})$, mediante una densidad de probabilidad $q(\boldsymbol{\theta})$ que pertenece a la familia de distribuciones manejable \mathcal{Q} . La mejor aproximación Bayesiana variacional $q^* \in \mathcal{Q}$, se encuentra minimizando la divergencia Kullback-Leibler (KL) de $q(\boldsymbol{\theta})$ a $p(\boldsymbol{\theta} \mid \mathbf{y})$. Simbólicamente

$$q^*(\boldsymbol{\theta}) = \arg \min_{q \in \mathcal{Q}} \text{KL} (q(\boldsymbol{\theta}) \parallel p(\mathbf{y} \mid \boldsymbol{\theta})) . \quad (10)$$

Esta divergencia es una medida de teoría de la información sobre la proximidad entre dos densidades: para q y p esta se define como

$$\text{KL} (q \parallel p) \equiv \int q(\theta) \cdot \log \frac{q(\theta)}{p(\theta)} d\theta = \mathbb{E}_q[\log q(\theta)] - \mathbb{E}_q[\log p(\theta)], \quad (11)$$

donde el valor esperado se toma con respecto a $q(\theta)$. Esta medida es asimétrica, no negativa y se minimiza cuando $q = p$. En general, la divergencia KL entre la aproximación propuesta y la verdadera a posteriori es intratable, ya implica conocer la forma analítica o cerrada de la distribución a posteriori y concretamente la evidencia del modelo $p(\mathbf{y})$. Por tal motivo, en la práctica se minimiza una función que es proporcional a la divergencia KL salvo por esta constante: el límite inferior de la evidencia, evidence lower bound (ELBO). Así, el nuevo objetivo variacional es

$$q^*(\boldsymbol{\theta}) = \arg \min_{q \in \mathcal{Q}} \mathbb{E}_q[p(\mathbf{y} \mid \boldsymbol{\theta})] - \text{KL} (q(\boldsymbol{\theta}) \parallel p(\boldsymbol{\theta})) . \quad (12)$$

El tipo de restricción impuesta sobre la familia variacional \mathcal{Q} determina el tipo y la calidad de la aproximación. A diferencia de las técnicas MCMC, el resultado de este método variacional no son muestras aleatorias de la densidad objetivo, sino la combinación de parámetros variacionales óptimos que minimizan la divergencia Kullback-Leibler (KL) entre la aproximación y la verdadera a posteriori.

El código fuente de R, Stan y el conjunto de datos empleado, están disponibles a través del siguiente repositorio de GitHub

- https://github.com/Demian-33/Articulo_log_ictpc_cdmx