

Programa de las Naciones Unidas para el Desarrollo/México
Oficina Nacional de Desarrollo Humano

Índice de Desarrollo Humano Municipal en México 2000-2005

Nota técnica 4
Estimación del ingreso en el ámbito municipal¹

Al obtener indicadores de bienestar desagregados debe tomarse en cuenta el conflicto entre cobertura, representatividad y calidad de la información. En México, por ejemplo, en términos de información para los años 2000 y 2005, la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) del Instituto Nacional de Geografía, Estadística e Informática (INEGI), provee una amplia gama de información útil para la medición del ingreso. Sin embargo, el tamaño y la representatividad de la encuesta sólo permiten crear medidas representativas a nivel nacional y para los estratos urbano y rural. El Censo General de Población y Vivienda (*Censo*) y el Conteo de Población y Vivienda (*Conteo*), por su parte, contienen información sociodemográfica para cada vivienda enumerada en el país; sin embargo, los datos de ingreso que contienen no poseen el grado de detalle y confiabilidad que la medición del desarrollo humano municipal requiere.

Así, el ingreso municipal utilizado en el cálculo del índice de desarrollo humano municipal en México, es una estimación con base en la metodología desarrollada por Elbers, Lanjouw y Lanjouw (2003) que combina el uso de las dos fuentes de información antes mencionadas en un proceso de *imputación de ingresos* (la ENIGH y el Censo en el caso del año 2000 y la ENIGH y el Conteo para el 2005). La metodología consiste en elaborar modelos que caractericen al ingreso como función de las características del hogar, del tipo de vivienda, de la educación de los miembros del hogar, y de variables de la localidad, todas obtenidas en las encuestas, para ser aplicadas a los datos del Censo o Conteo con el fin de predecir el ingreso.

El proceso de imputación de ingresos requiere la selección de variables que sean comunes al Censo o Conteo y a la ENIGH, de manera que se obtengan los *mejores* modelos de predicción de ingreso en la encuesta de hogares, para ser aplicados a la población en su conjunto. El proceso requiere que existan:

- i) Una encuesta de hogares que incluya ingreso y otras características sociodemográficas, y que esté disponible idealmente para el mismo periodo de un Censo o Conteo. En el caso de México esta encuesta es la ENIGH en sus versiones 2000 y 2005;
- ii) La base de datos del Censo o Conteo disponible a nivel de registro, y
- iii) Un grupo de variables comunes y con distribuciones consistentes;

La metodología consiste básicamente en que si un indicador de bienestar, W , depende de alguna variable de interés, en este caso el ingreso del hogar, y_h , es posible obtener, a partir de una encuesta y un censo, la distribución conjunta de y_h empleando una serie de variables correlacionadas, x_h . De manera intuitiva se trata de tomar la encuesta de hogares como una muestra aleatoria de la población, representada en el censo, de manera que las variables correlacionadas, x_h , se encuentren disponibles en ambas fuentes. De esta forma se producen estimaciones puntuales y errores de predicción del vector y_h .

¹ La metodología aquí resumida se encuentra descrita con detalle en Székely, López-Calva et al. (2007), y utilizada para la actualización de la información por Izaguirre, López-Calva et al.

La primera consideración se relaciona con la obtención de un modelo adecuado de predicción de ingresos. Así, la variable y_{ch} , el ingreso del hogar h que reside en la localidad o el grupo c , se estima de manera lineal como:

$$\ln y_{ch} = E[\ln y_{ch} | \mathbf{x}_{ch}] + u_{ch} = \mathbf{x}_{ch}\beta + u_{ch} \quad (1)$$

donde el vector de errores $\mathbf{u} \sim \Gamma(0, \Sigma)$. Es fundamental señalar que el vector β en este caso no representa los efectos marginales teóricos de una característica sobre el ingreso. El vector β podría estar sesgado, pero si ello contribuye a reducir el error de predicción, es aceptable dentro de este esquema. El error se considera de la siguiente manera para permitir correlación espacial:

$$\mathbf{u}_{ch} = \eta_c + \varepsilon_{ch} \quad (2)$$

Las especificaciones en (1) y (2) forman conjuntamente un “modelo lineal jerárquico”. Es necesario asumir que ambos componentes en (2), el de la localidad y el del hogar, no están correlacionados entre sí. Si una mayor proporción de dicho error corresponde al factor de localidad, las estimaciones serán de menor precisión y no se ganará demasiado al agregar más hogares a la localidad, dada una matriz de varianza-covarianza de los errores. La forma de reducir el componente inherente a la localidad consiste en agregar al vector de características el mayor número posible de variables que son específicas a la misma. En realidad, gran parte de la metodología se centra en el procedimiento para tratar los dos componentes del error total de predicción, con el propósito de minimizarlos. El modelo permite además corregir por heteroscedasticidad en los errores inherentes al hogar (idiosincrásicos).

A manera de ejemplo considérese el caso específico de la imputación de ingresos para el año 2000. La lista de variables comunes en la ENIGH 2000 y el Censo 2000 es la siguiente:

- i) Vivienda: disponibilidad de agua; disponibilidad de electricidad; combustible para cocinar; material en pisos; material en muros; material en techos; cuarto para cocinar; disponibilidad de drenaje, y tenencia;
- ii) Bienes en la vivienda: teléfono; radio; televisión; videocasetera; licuadora; refrigerador; lavadora; calentador de agua; automóvil o camioneta propia, y computadora;
- iii) Demográficas: sexo; estado civil, y parentesco;
- iv) Educación: asistencia escolar; alfabetismo, y escolaridad, y
- v) Laborales: condición de actividad; ocupación; horas trabajadas, y posición en el trabajo.

A partir de estas variables se construye el conjunto de variables (originales y compuestas) del que se elegirán a las que sean adecuadas como variables explicativas en el modelo para cada estrato y en cada región.

Para no forzar los parámetros a un solo modelo de imputación, lo cual disminuiría la precisión del procedimiento, el país se dividió en cinco grupos de entidades federativas de acuerdo con sus índices de marginación y cada grupo a su vez fue separado en áreas rural y urbana. Así, en total se estimaron diez variantes del modelo (1).

La agrupación utilizada, con base en niveles de marginación, es la siguiente:

Agrupación 1 (marginación muy baja): Aguascalientes; Baja California; Baja California Sur; Coahuila; Chihuahua; Distrito Federal, y Nuevo León;

Agrupación 2 (marginación baja): Colima; Jalisco; estado de México; Sonora, y Tamaulipas;

Agrupación 3 (marginación media): Durango; Guanajuato; Morelos; Nayarit; Querétaro; Quintana Roo; Sinaloa; Tlaxcala, y Zacatecas;

Agrupación 4 (marginación alta): Campeche; Hidalgo; Michoacán; Puebla; San Luis Potosí; Tabasco, y Yucatán, y

Agrupación 5 (marginación muy alta): Chiapas; Guerrero; Oaxaca, y Veracruz.

Posteriormente se combinaron los parámetros estimados en la primera etapa con las características observables en cada hogar en el censo, para generar un gasto logarítmico estimado, así como la simulación de los errores. Cabe señalar que los ingresos promedio estimados se obtienen con coeficientes de variación promedio menores al 10%.

Una vez obtenido el ingreso promedio per cápita por municipio, éste se ajusta a cuentas nacionales mediante la aplicación de un factor homogéneo a nivel nacional. El ingreso también es ajustado mediante un factor de conversión a términos de Paridad de Poder de Compra en dólares estadounidenses (dólares PPC).

Así, para cada municipio, el índice se construye como:

$$\text{Índice de ingreso} = \frac{\log(\text{ingreso promedio per cápita municipal}) - \log(\text{ingreso mínimo})}{\log(\text{ingreso máximo}) - \log(\text{ingreso mínimo})}$$

donde el ingreso promedio per cápita municipal es el estimado mediante el ejercicio de imputación (expresado en dólares estadounidenses PPC). Los valores máximo y mínimo de referencia, son los establecidos oficialmente por el PNUD de 40,000 y 100 dólares PPC, respectivamente.

Referencias:

Elbers, C., J. Lanjouw y P. Lanjouw. 2003. "Micro-Level Estimation of Poverty and Inequality", *Econometrica*, vol. 71, núm. 1, pp. 355-64.

Izaguirre, C., L. F. López-Calva y L. Rodríguez-Chamussy. "Actualización del mapa de pobreza de México, 2005", mimeo, Programa de las Naciones Unidas para el Desarrollo/México.

Székely, M., L. F. López-Calva, Á. Meléndez, E. Rascón y L. Rodríguez-Chamussy. 2007. "Poniendo a la pobreza de ingresos y a la desigualdad en el mapa de México", *Economía mexicana, nueva época*, vol. X I, núm. 2, pp. 239-303, segundo semestre.