

Nota Técnica

Estimación del ingreso a nivel municipal¹

Obtener indicadores de bienestar a nivel desagregado es difícil en general pues existe un conflicto entre cobertura, representatividad y calidad de las fuentes de datos. México no es la excepción. Por ejemplo, en términos de información para el año 2000, 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 (El Censo), por su parte, contiene información sociodemográfica para cada vivienda enumerada en el país; sin embargo, los datos de ingreso que contiene no poseen el grado de detalle y confiabilidad que la medición del desarrollo humano municipal requiere.

Así, la estimación del Índice de Desarrollo Humano (IDH) municipal utiliza ambas fuentes de datos para imputar ingresos combinando la ENIGH y el Censo, siguiendo la metodología desarrollada por Elbers, Lanjouw y Lanjouw (2002). La metodología consiste en elaborar modelos que caractericen el ingreso como función de las características del hogar, del tipo de vivienda, de la educación de los miembros del hogar, y variables de la localidad, obtenidas en las encuestas para ser aplicadas a los datos del censo con el fin de predecir el ingreso (o el consumo). En este trabajo, la variable de ingreso necesaria para medir el desarrollo humano en los municipios de México se ha añadido artificialmente a la información del censo de población 2000 usando las relaciones estructurales derivadas de la ENIGH 2000.

El proceso requiere la selección de variables que sean comunes al Censo y 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 (Ver López-Calva, et al. 2004a, b). 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. En el caso de México esta encuesta es la ENIGH 2000,
- ii) La base de datos del censo disponible a nivel de registro;
- iii) Un grupo de variables comunes y con distribuciones consistentes;

La metodología consiste básicamente en que si un indicador de bienestar W depende de la variable de interés, en este caso el ingreso del hogar, y_h , a partir de la encuesta y el censo se puede obtener la distribución conjunta de y_h y una serie de variables correlacionadas, x_h . En realidad, de manera intuitiva, consiste en tomar a encuesta de hogares como una muestra

¹La metodología aquí resumida se encuentra descrita con todo detalle en López-Calva, et al. (2004a, 2004b).

aleatoria de la población, representada en el censo, de manera que las variables correlacionadas, x_h , se encuentren disponibles en ambas fuentes. De esta manera se producen estimaciones puntuales y errores de predicción del vector y_h .

El primer punto metodológico consiste en obtener 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 linealmente 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 en este punto 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:

$$u_{ch} = \eta_c + \varepsilon_{ch} \quad (2)$$

Las especificaciones en (1) y (2) forman conjuntamente un “modelo lineal jerárquico”. Es necesario asumir que ambos componentes en (2), el específico 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 por 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 en el 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, de manera que ellos se minimicen. El modelo además permite corregir por heteroscedasticidad en los errores inherentes al hogar (idiosincrásicos).

Para este caso específico, la lista de variables comunes en la Encuesta Nacional de Ingresos y Gastos de los Hogares (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, tenencia;
- ii) Bienes en la vivienda: teléfono, radio, televisión, videocasetera, licuadora, refrigerador, lavadora, calentador de agua, automóvil o camioneta propios, computadora;
- iii) Demográficas: sexo, estado civil, parentesco;
- iv) Educación: asistencia escolar, alfabetismo, escolaridad;
- v) Laborales: condición de actividad, ocupación, horas trabajadas, posición en el trabajo;

A partir de las variables mencionadas se construye el conjunto de variables (originales y compuestas) del que se escogerán las que son adecuadas como variables explicativas para el modelo en cada estrato y en cada región.

Para no forzar los parámetros a un solo modelo de imputación, con el consiguiente costo en precisión, el país se dividió en cinco agrupaciones de entidades de acuerdo a sus índices de marginación, y cada agrupación a su vez se separó en áreas rural y urbana. Así, se estimaron 10 variantes del modelo (1) en total.

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.

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

Posteriormente, combinamos los parámetros estimados en la primera etapa con las características observables para 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 capita por municipio, se procede a ajustarlo a cuentas nacionales mediante la aplicación de un factor homogéneo a nivel nacional y, posteriormente, se ajusta mediante factor de conversión a términos de Paridad de Poder de Compra en dólares estadounidenses (PPC en USD) (*ver Informe sobre Desarrollo Humano México 2002*).

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

$$Ind.Ingreso = \frac{\log(Ing.Pr\ om.Mun.) - \log(Ing.\ min)}{\log(Ing.\ max) - \log(Ing.\ min)}$$

Referencias

Elbers, C., P. Lanjouw, and J. Lanjouw (2002), “Welfare in Towns and Villages. Micro-measurement of Poverty and Inequality”, *Econometrica*.

Hentschel, J., J. Lanjouw, P. Lanjouw and J. Poggi (1999), “Combining Survey Data with Census Data to Construct Spatially Disaggregated Poverty Maps: A Case Study of Ecuador”, *World Bank Economic Review*. Vol. 14, No. 1, pp. 147-66.

López-Calva, L.F., E. Rascón, L. Rodríguez-Chamussy y M. Székely (2004a), “Poniendo a los pobres en el mapa: Estimación de indicadores de bienestar a nivel municipal en México”, Documento de Investigación, SEDESOL, México.

López-Calva, L.F., L. Rodríguez-Chamussy y C. Rodríguez-García (2004b), “Indicadores de desarrollo humano a nivel municipal en México”, *Estudios sobre Desarrollo Humano, PNUD-México*.