

EL REZAGO SOCIAL EN ÁREAS METROPOLITANAS DE MÉXICO*

Francisco J. Benita Maldonado

ITESM, Campus Monterrey

Marco Vinicio Gómez Meza

Universidad Autónoma de Nuevo León

Resumen: En México existen grandes lagunas en la producción y análisis de datos a nivel de pequeña escala territorial. Este trabajo propone una metodología que permite estimar el grado de rezago social para las manzanas de las áreas metropolitanas de México. Se retoma la propuesta del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) y se utiliza el *Censo de Población y Vivienda 2010* para ejemplificar su uso para el área metropolitana de Monterrey. Se ofrecen dos versiones complementarias y se encuentra que, en ambos casos, la proporción de la población que sufre de un grado de rezago social muy alto es considerablemente mayor a la presentada por el Coneval.

Abstract: Large gaps exist in Mexico about production and analysis of small-scale territorial units. The paper addresses a new methodology for estimating the social backwardness for blocks of the metropolitan areas in Mexico. Based on The National Council for Evaluation of Social Development Policy (Coneval) methodology and using the 2010 Population and Housing Census, as an example, we estimate the index of social backwardness for blocks in the Monterrey metropolitan area. We have implemented two versions to compute the index and both of them suggest the rate of the population suffering from Very High social backwardness is substantially different rate from that presented by Coneval.

Clasificación JEL/JEL Classification: C43, I31, I32, R58.

Palabras clave/keywords: Rezago social, manzanas, área metropolitana, análisis factorial, focalización geográfica, social backwardness, small area statistics, metropolitan area, factor analysis, geographical targeting.

Fecha de recepción: 27 XI 2012

Fecha de aceptación: 17 IX 2013

* francisco_benita@hotmail.com, marco.gomezmez@uanl.edu.mx.

1. Introducción

Al tratar de medir el nivel de desarrollo de un país o una región, la primera dificultad surge en la definición misma de desarrollo. Una definición simple lo puede considerar como el avance en las condiciones de vida de la población y puede interpretarse como mayor bienestar. Así, se espera que el desarrollo y el bienestar alcancen a la mayoría de la población de un país y a todas sus zonas geográficas. Sin embargo, por situaciones estructurales, algunas áreas y grupos alcanzan con mayor facilidad los beneficios del desarrollo mientras que otros se mantienen rezagados en relación con sus contiguos. En este sentido, existen brechas tanto geográficas como entre grupos de población, y es precisamente la identificación de estas diferencias lo que se ha tratado de medir con la definición y construcción de indicadores sociales y económicos.

En México, no ha sido sino hasta la primera década del siglo XXI cuando se comenzó la producción de información estadística acerca de la multidimensionalidad de la pobreza, que complementan los anteriores estudios de pobreza por ingresos. Bajo esta lógica es que surge el índice de rezago social, propuesto por el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) en el año 2007, que parte de un enfoque relativo de pobreza y opera bajo el método de medición directo. Este indicador sintetiza en una sola variable cuatro dimensiones de la pobreza: educación, acceso a servicios de salud, calidad y espacios en la vivienda y activos en el hogar. Su objetivo es identificar la posición relativa que toma cada observación –ya sean estados, municipios, localidades o áreas geoestadísticas básicas, Agebs– respecto a sus semejantes. Sin embargo, una de sus debilidades es que se ha venido construyendo a niveles de agregación muy grandes, lo que asume homogeneidad entre las observaciones.

La manzana, como unidad de análisis, ha tomado un papel marginal en los estudios de dicha índole debido a la escasez o incluso inexistencia de información oficial. Hecho que constituye una de las principales aportaciones de este trabajo de investigación, ya que se detalla una metodología que hace posible la cuantificación del grado de rezago social para tal tipo de extensiones territoriales, las cuales resultan ser geográficamente las más pequeñas de analizar. La motivación del análisis radica en la importancia de la focalización de la población vulnerable –entendida como la más rezagada– ya que, tras el uso de los indicadores tradicionales, como el de rezago social del Coneval o el de marginación del Consejo Nacional de Población (Conapo), por entidad, por municipio, por localidad o por Ageb resultan en la práctica poco eficientes, ya que los resultados son generalidades derivadas

del conjunto de características de la población agrupada en unidades territoriales relativamente grandes. En otras palabras, establecer un ordenamiento de acuerdo con el grado de rezago social por manzana resulta ser sumamente práctico para los hacedores de política pública, pues la focalización, a esos niveles de desagregación, les permite alcanzar mayor eficiencia y equidad en las políticas sociales, al tener así un criterio decisivo para la asignación del gasto público social, ya que asegura que los grupos vulnerables tengan acceso a los bienes y servicios públicos sociales.

En este sentido, el objetivo principal de la investigación es el de presentar una metodología que permita estimar con mayor precisión el grado de rezago social, para su posterior focalización. Al retomar la propuesta del Coneval (2007) y utilizar el *Censo de población y vivienda 2010*, se ofrecen dos versiones complementarias para la estimación del índice para las áreas metropolitanas de México (AMMX), la primera, que va en la línea del Coneval, aplica la técnica de análisis de componentes principales (ACP) y la segunda utiliza el método de máxima verosimilitud (MMV). Se espera que al construir el índice de rezago social que considere por separado cada una de las 59 AMMX, el número de personas que sufran de un grado de rezago muy alto¹ se incremente, en contraste con los resultados expuestos por el Coneval (2011) e inclusive por el Conapo (2011 y 2012), debido a que, al trabajar con observaciones geográficamente más pequeñas, la mayor heterogeneidad entre observaciones es tomada en cuenta.

La estructura del documento se compone de cinco secciones, después de la introducción se presentan los antecedentes y el contexto dentro del cual se enmarca este trabajo de investigación, la tercera sección detalla la metodología para la estimación del índice de rezago social para las AMMX, en la cuarta se ejemplifica el uso de esta metodología al aplicarla en el área metropolitana de Monterrey (AMM) y para finalizar se exponen algunas conclusiones y recomendaciones.

2. Antecedentes teóricos y conceptuales

Son múltiples los aspectos que pueden relacionarse con el nivel de desarrollo social de una zona o área geográfica: culturales, ambientales, educativos, económicos, de salud, entre otros. Sin embargo, la definición de los aspectos y variables por considerar en la construcción de

¹ La clasificación del rezago social se lleva a cabo en cinco estratos de acuerdo con su intensidad: muy bajo, bajo, medio, alto y muy alto.

un indicador sintético depende de las posibilidades de medición que se tengan y la disponibilidad de la información. En la historia reciente de México tales diferencias de medición han variado de acuerdo con las dimensiones y características que se estén analizando, además del intento por considerar las condiciones macroeconómicas vigentes en diferentes momentos.

El creciente interés y preocupación por el tema llevó a la promulgación de la *Ley General de Desarrollo Social* el 20 de enero de 2004, y destaca como uno de sus objetivos centrales "...garantizar el pleno ejercicio de los derechos sociales consagrados en la Constitución Política de los Estados Unidos Mexicanos, asegurando el acceso de toda la población al desarrollo social" (LGDS, 2004:1). La incorporación de mecanismos institucionales de evaluación y seguimiento de las políticas de desarrollo social constituye una innovación fundamental de esta ley. Para ello, se creó el Coneval, que es un organismo público con autonomía técnica y de gestión, que tiene por misión, normar y coordinar la evaluación de las políticas y programas de desarrollo social, así como establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza.

Así, según las disposiciones del capítulo VI, artículos 36 y 37 de dicha ley, se enmarca la periodicidad mínima y el nivel de desagregación –estatal y municipal– con la que deben hacerse las mediciones de pobreza en México. Asimismo, señala que los lineamientos y criterios que establezca el Coneval para la definición, identificación y medición de la pobreza deben considerar, al menos, los siguientes ocho indicadores (LGDS, 2004:5): 1) Ingreso corriente *per cápita*, 2) Rezago educativo promedio en el hogar, 3) Acceso a los servicios de salud, 4) Acceso a la seguridad social, 5) Calidad y espacios de la vivienda, 6) Acceso a los servicios básicos en la vivienda 7) Acceso a la alimentación y 8) Grado de cohesión social.

Bajo esta dinámica es que se han venido propuesto y modificado distintos indicadores, tales como el índice de marginación en 1990 por el Conapo, el índice de desarrollo humano en 1990 por el PNUD y el índice de rezago social en 2007 por el Coneval. La utilidad de tales indicadores radica en que resumen las distintas condiciones sociales que existen en las entidades y que, para su construcción, se considera el acceso a los servicios de salud, la educación e ingresos monetarios; algunos incluyen las condiciones de la vivienda y otros toman en cuenta la distribución de la población. Empero, la debilidad de estos índices es que no permiten reconocer las heterogeneidades que existen al interior de dichos estados o municipios.

Entre las metodologías existentes para la medición del bienestar

social en el país, la medición del desarrollo humano es un elemento fundamental para el diseño de las políticas públicas. Aunque en México se han elaborado y aplicado metodologías para medir el desarrollo de manera intensiva, el índice de desarrollo humano –que combina el PIB por habitante, la salud y la educación– llena un hueco importante en el conjunto de medidas de desarrollo utilizadas sistemáticamente de manera institucional en el país. Debido a su simplicidad y requisición de información se ha convertido en un punto de referencia muy utilizado para realizar comparaciones internacionales e incluso muchos países han adoptado los índices de desarrollo humano como instrumento de política y como indicador del éxito o fracaso de sus políticas nacionales. Recientemente, el *Informe mundial sobre desarrollo humano 2010* introdujo una nueva metodología para el cálculo del indicador que refina las dimensiones de educación e ingreso y ajusta los referentes internacionales de todas sus variables y la manera en que las mismas se sintetizan. Sin embargo, existe gran controversia respecto a su utilización,² por ejemplo, la gran correlación existente entre el índice y sus componentes subyacentes, las posibles “paradojas” que podrían surgir en caso de que una sociedad intentara maximizar el índice de desarrollo humano, su inadecuación como una medida más amplia del bienestar e incluso se ha puesto en duda su capacidad para medir el potencial de la economía para proveer bienes y servicios a las personas. Para atender estas limitantes, en el caso de México se ha propuesto construir el indicador con distintas variables *proxys* a las originales y analizar la sensibilidad de los cambios en la posición relativa de las observaciones (López, Rodríguez y Velez, 2006). El índice de rezago social, por su parte, es una medida que busca superar dichas limitantes, ya que en su construcción considera otras dimensiones relativas a la pobreza y el desarrollo.

El índice de marginación resulta de especial interés pues la metodología para su construcción es casi en su totalidad idéntica a la que utiliza el Coneval para el rezago social. Este indicador es una medida-resumen que permite diferenciar entidades federativas, municipios, localidades y Agebs según el impacto global de las carencias que padece la población, al considerar cuatro dimensiones estructurales: falta de acceso a la educación, residencia en viviendas inadecuadas, percepción de ingresos monetarios insuficientes y vivir en localidades con menos de cinco mil habitantes. A partir de tales dimensiones se generan nueve indicadores socioeconómicos de marginación, que son pos-

² Véase, por ejemplo, de la Torre (2004), Székely y López (2006) y Cahill (2007).

teriormente sintetizados con el ACP –técnica de análisis factorial que también es utilizada en el índice de rezago social–. Sin embargo, diversos autores han estudiado las limitantes de este indicador, Cárdenas (2010) argumenta que no es posible emplearlo para medir la efectividad de programas gubernamentales orientados a disminuir el rezago social, como alternativa el autor genera un índice reparametrizado que permite superar la limitante referida. Camberos y Bracamontes (2007) utilizan métodos de regresión para relacionar el índice de año base con los indicadores de otras fechas, con objeto de atender la limitación temporal del indicador. Gutierrez y Gama (2010) critican la forma en que se ponderan los indicadores socioeconómicos y proponen la normalización de las variables de manera que tengan un impacto promedio similar en la construcción del índice. Finalmente, Bistrain (2010) propone aplicar un método complementario que reduzca los errores de estratificación y coadyuve a alcanzar los objetivos del índice, ya que la técnica de estratificación utilizada por el Conapo –y el Coneval para el caso del índice de rezago social– no establece límites en cuanto al número de intervalos de clase en que se puede fraccionar el rango del índice y, por ende, el número de observaciones asignadas a cada estrato resulta sensible a dicho criterio, pues depende del número de intervalos de clase en que se divida el rango.

El rezago social, por su parte, hace referencia a un indicador que busca establecer las diferencias que existen entre distintas zonas geográficas de una misma región en función del acceso a los beneficios del desarrollo social. Para ello se hace un ordenamiento de cada una de estas zonas, se identifica así su posición relativa en la región y logra identificar aquéllas que se han visto menos beneficiadas por el desarrollo social. Debemos aclarar que el rezago social no es considerado propiamente un indicador de pobreza, pues en su construcción no se consideran indicadores de ingresos monetarios. Sin embargo, al considerar cuatro dimensiones relativas a la dinámica de este fenómeno –educación, acceso a servicios de salud, calidad y espacios en la vivienda y activos en el hogar– el indicador puede enmarcarse dentro del enfoque relativo de la pobreza³ –igual que el índice de desarrollo

³ *Grosso modo* existen dos grandes enfoques para concebir a la pobreza: el absoluto y relativo. De acuerdo con Spicker (1999) la diferencia entre absoluto y relativo no estaría en la definición de pobreza, sino que son, más bien, interpretaciones de la manera en la que se forman socialmente las necesidades. Mientras el primer enfoque sostiene que las necesidades son independientes de la riqueza de los demás, y no satisfacerlas revela una condición de pobreza en cualquier contexto, el segundo plantea que las necesidades surgen a partir de la comparación con los demás, y la condición de pobreza depende del nivel general de riqueza.

humano y el de marginación— que concibe a este evento en términos de su relación con estándares existentes en la sociedad.

De la misma manera en la que los enfoques buscan definir a la pobreza existen métodos específicos para su medición. No obstante, no existe consenso a nivel internacional que permita obtener un método de medición único. El enfoque y el consecuente método a utilizar dependerán del tipo de pobreza que el investigador quiera cuantificar, así los resultados obtenidos en diferentes investigaciones pueden variar sensiblemente de acuerdo con la definición del término de pobreza; concretamente la pobreza puede medirse de manera indirecta, directa o combinada. Se dice que la medición es directa cuando el investigador observa directamente las condiciones de vida de la población para así, en función de lo observado, poder determinar si un hogar o individuo es pobre o no. Bajo este enfoque una persona sería considerada pobre si no satisface una o varias necesidades básicas definidas con anterioridad, como por ejemplo, cierto nivel educativo, condiciones adecuadas para vivir o cumplir con un nivel mínimo de alimentación. Es el método que utiliza el rezago social para la identificación de la población vulnerable por su elevado grado de rezago.⁴

De manera comparativa, el índice de desarrollo humano es una herramienta útil para explorar las características locales de las capacidades básicas de los individuos en una sociedad, pone énfasis en crear las condiciones propicias para que la población disfrute de una vida prolongada, saludable y creativa. Sin embargo es calculado a nivel municipal, en el mejor de los casos, conserva su estructura básica de cálculo, pero se observa como algunas variables de medición son remplazadas por otras más propias para poblaciones de tamaño reducido. Por otro lado, de modo similar al índice de rezago social, el índice de marginación considera que la privación de capacidades se puede medir a través del acceso a bienes, servicios e infraestructura, y tanto la marginación como el rezago se circunscriben al enfoque de necesidades básicas insatisfechas. Empero, una de las principales diferencias entre ambos es que el rezago social no toma en cuenta los ingresos monetarios, conserva su carácter directo de medición, mientras que el de marginación sí incorpora esta variable, intenta conseguir un indicador de medición combinado, pero imposibilita distinguir el

⁴ A saber, la diferencia principal con el método indirecto es que éste asocia el bienestar con el consumo efectivamente realizado, mientras que el método directo lo relaciona con la posibilidad de realizar consumo (Ringén, 1988: 355). Por su parte, el método combinado busca mezclar ambas metodologías y refleja la multiplicidad de las fuentes de bienestar de los hogares.

tipo de pobreza –por ingresos o por carencias– según cada uno de los métodos. Por su parte, la manzana como unidad de análisis es posiblemente la que mejor describe las condiciones de rezago en una zona geográfica que se encuentra aglomerada, pues comparten vínculos económicos, sociales y culturales. Bajo esta lógica, la manzana es la unidad mínima de análisis para estimar el índice de rezago social de acuerdo con la disponibilidad de la información. Hay que señalar aquí que lo idóneo sería elaborar el indicador por hogares, sin embargo, los criterios de confidencialidad del Instituto Nacional de Estadística y Geografía (INEGI) no permiten revelar la identidad de los informantes que proporcionan información para fines estadísticos.

En el polo opuesto se ubican los estudios elaborados por entidad federativa, en los que, en teoría, están asumiendo homogeneidad entre individuos a lo largo de toda la extensión territorial, lo que eventualmente conllevaría a sobre/subestimaciones en los resultados. Las cifras oficiales publicadas sobre rezago social y marginación por el Coneval y el Conapo, respectivamente, han mostrado que al trabajar con extensiones territoriales más pequeñas el número de personas con mayor grado de rezago o marginación se incrementa de forma considerable, al pasar de estados a municipios, de municipios a localidades y de localidades a Agebs.⁵

Por otro lado, han sido pocos los estudios que han utilizado a la manzana como unidad de análisis, en Martínez, Treviño y Gómez (2009) se construye por primera vez en México el índice de rezago social para manzanas y tiene como población objetivo el AMM. Con base en información del *Conteo de población y vivienda 2005*, el estudio permitió al gobierno del estado diseñar estrategias focalizadas de abatimiento a la pobreza. Además, el Consejo de Evaluación del Desarrollo Social del Distrito Federal ha elaborado recientemente un índice de desarrollo social para el Distrito Federal, diseñado específicamente para unidades territoriales, es decir, delegaciones, colonias y manzanas. El estudio permitió contar con una herramienta valiosa para la planeación de políticas y programas sociales al permitir el análisis desagregado en el territorio, de los avances que se han logrado (Evalúa DF, 2011).

La construcción del indicador por áreas metropolitanas a partir de las manzanas permite capturar el comportamiento del rezago para zonas urbanas cuya estructura es relativamente homogénea al interior. Es decir, a diferencia de las metodologías tradicionales, en las que se computa el grado de rezago, desarrollo o marginación para todas las

⁵ Véase, por ejemplo, Conapo (2006, 2011, 2012) y Coneval (2007, 2011).

observaciones que componen la totalidad de la organización territorial en México; el hacerlo de manera independiente permite obtener resultados claros y comparables al interior de la zona en cuestión. En términos de política social, estos resultados tienen un mayor sentido y sientan bases para criterios de planeación y política urbana.

Finalmente, y derivado del uso de la manzana como unidad de análisis, es posible focalizar geográficamente el rezago social. Con el uso de instrumentos de focalización se pueden identificar áreas geográficas en donde se concentra la población más rezagada y se establecen ordenamientos geográficos que facilitan la priorización de la asignación del gasto público. Este tipo de focalización es útil a niveles reducidos de agregación geográfica, que se caracterizarían por la relativa homogeneidad de su población; la principal ventaja de estos instrumentos, comúnmente conocidos como mapas de pobreza –de rezago para el caso que nos ocupa–, es su bajo costo de instrumentación y monitoreo. Además, Bigman y Fofack (2000:134) señalan que proveen criterios claros de identificación de la población objetivo, son fáciles de monitorear y administrar, no inciden en el comportamiento de la población, entre otras características.

Para el caso particular de los mapas de rezago del AMMX es necesario profundizar en el entendimiento de los factores que operan en ellos, cómo se encadenan, articulan, acumulan, se refuerzan entre sí y se convierten en trampas, muchas veces insuperables, para una gran parte de sus residentes. En este sentido, los mapas de rezago pueden tener alguna o varias de las siguientes condiciones:

- a) Concentración de población que acumula factores de vulnerabilidad, como bajo nivel educativo, de formación profesional obsoleta para el mercado laboral, desempleo o subempleo.
- b) Viviendas inadecuadas, cuando no se trata directamente de infra-vivienda, especialmente las que están ubicadas en áreas marginales o en suelos irregulares.
- c) Falta de expectativas y estímulos para la juventud, común en estos barrios, lo que facilita la deserción y el fracaso escolar y, en ciertos casos, la vinculación posterior a conductas irregulares, delictivas, consumo de droga y/o algún otro tipo de marginalidad social.
- d) En algunos casos, concentración de población inmigrante con distinta cultura, lengua y etnia, que dificulta su integración en la sociedad local.

- e) Normalmente estas áreas presentan degradación ambiental, servicios deficientes, difícil accesibilidad y nula iniciativa económica local.

3. Metodología para áreas metropolitanas

El índice de rezago social propuesto por el Coneval (2007) se construye a partir de 12 indicadores. La propuesta del índice de rezago social de las AMMX consiste en utilizar diez de los 12 indicadores propuestos originalmente, así como uno adicional que sustituye al indicador de analfabetismo, dada la naturaleza de la disponibilidad de información.

El cuadro 1 muestra los indicadores considerados, el Coneval para la construcción del índice para estados, municipios y localidades utiliza del dos al once, así como el porcentaje de la población de 15 años y más analfabeta y el porcentaje de la población de 6 a 14 años que no asiste a la escuela. Sin embargo, hay que aclarar, que al trabajar con manzanas y dados los criterios de confidencialidad del INEGI, las variables que son necesarias para la construcción de los indicadores aparecen como perdidas, lo que provocaría una pérdida generalizada de información.

Cuadro 1
Descripción de los indicadores

<i>Dimensión</i>	<i>Indicador</i>	<i>Descripción</i>
Educación	1 Iesc*	Diferencia respecto al valor máximo del grado promedio de escolaridad
	2 Iedbasinc	Porcentaje de la población de 15 años o más con educación básica incompleta
Acceso a servicios de salud	3 Isdersalud	Porcentaje de la población sin derechohabiencia a servicios de salud
Calidad y espacios en la vivienda	4 Iptierra	Porcentaje de las viviendas particulares habitadas con piso de tierra
	5 Ihacin	Promedio de ocupantes por cuarto
	6 Inosan	Porcentaje de las viviendas particulares habitadas que no tienen excusado o sanitario

Cuadro 1
(continuación)

<i>Dimensión</i>	<i>Indicador</i>	<i>Descripción</i>
	7 Inoagua	Porcentaje de las viviendas particulares habitadas sin agua entubada de la red pública
	8 Inodren	Porcentaje de las viviendas particulares habitadas que no disponen de drenaje
	9 Inoelec	Porcentaje de las viviendas particulares habitadas que no tienen energía eléctrica
Activos en el hogar	10 Inolav	Porcentaje de las viviendas particulares habitadas que no tienen lavadora
	11 Inoref	Porcentaje de las viviendas particulares habitadas que no cuentan con refrigerador

Nota: *Este indicador sustituye al considerado originalmente: porcentaje de la población de 15 años y más analfabeta. Se toma la diferencia respecto al valor máximo para que la correlación entre el grado promedio de escolaridad y la población de 15 años y más analfabeta sea positiva.

Fuente: Elaboración propia.

Adicionalmente, se hace necesario realizar imputaciones de valores para poder construir algunos de los indicadores de las manzanas. En la base de datos cualquier variable a nivel manzana, cuyo valor sea menor a tres unidades, aparece como perdida, a excepción de la población total y el total de hogares. Sin embargo, es posible recuperar información por diferencias, por ejemplo, para el caso del indicador cuatro, su construcción surge a partir de $(vph_pisoti_i / tothog_i) \cdot 100$, donde en el primer término el numerador corresponde a las viviendas particulares habitadas con piso de tierra dentro de la manzana i y el denominador al total de hogares de la manzana i . Para cada manzana la igualdad $tothog_i = vph_pisoti_i + vph_pisodt_i$ debe cumplirse, en donde vph_pisodt_i son las viviendas particulares habitadas con piso de material distinto de tierra de la manzana i . Lo mismo ocurre con la construcción de los indicadores tres, siete, ocho y nueve.

El algoritmo consiste en imputar para los cinco indicadores anteriormente descritos, valores de uno y dos cuando la diferencia de las igualdades es menor a tres, pero, cuando la diferencia es mayor o igual a tres, se sugiere imputar el valor del promedio del indicador de

acuerdo con el Ageb al que pertenece. No obstante, hay que señalar que este procedimiento debe ser instrumentado con cautela, ya que podría alterar significativamente los resultados que arroje el índice, especialmente debido a que en el *Censo de población y vivienda 2010* las igualdades que se supondría deberían cumplirse para aquellas manzanas cuya información existe en su totalidad, no necesariamente son cumplidas.

De tal suerte, el criterio sugerido consiste en: *a)* Imputar por diferencia los valores uno y dos para cualquiera de los indicadores tres, cuatro, siete, ocho y nueve; *b)* En caso de que, antes de imputación, la igualdad no se cumpla para, al menos, 75% de las manzanas cuya información existe en su totalidad, descartar la imputación del promedio por Ageb y *c)* No instrumentar el algoritmo si existe información para más de 95% del total de manzanas. Así, estas restricciones buscan minimizar las posibles sobre/subestimaciones del grado de rezago social derivadas por la instrumentación del algoritmo. Una vez construidos los indicadores el procedimiento para la construcción del índice de rezago social para las AMMX se estima a partir de dos técnicas de análisis factorial: el ACO y el MMV.

3.1. *Análisis de componentes principales*

El primer paso consiste en la estandarización de las once variables indicador para las n manzanas de la zona metropolitana elegida, que se lleva a cabo restándole la media y al dividir por la desviación estándar para cada una de las n observaciones en cada uno de los indicadores, de forma que las variables indicador se transforman en $I_j \sim X_j$. Seguido del proceso de estandarización, donde estas nuevas variables tienen media cero, varianza y desviación estándar unitaria, el segundo nivel de análisis implica transformar el espacio de los once vectores estandarizados $X_j (j = 1, 2, \dots, 11)$ en uno nuevo, $Y_k (k = 1, 2, \dots, p; \text{ con } p = 11)$, tal que, este espacio sean combinaciones lineales de los indicadores estandarizados –transformación $X_j \sim Y_k$ –, y se cumple:

$$\begin{aligned} Y_1 &= \mathbf{a}'_1 X = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= \mathbf{a}'_2 X = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ Y_p &= \mathbf{a}'_p X = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned} \quad (1)$$

donde \mathbf{a}'_k contiene al vector de coeficientes a_{jk} . La forma en la que se calcula cada componente es a través de un sistema de ecuaciones propuesto en (1), maximizar la varianza para cada uno de ellos, sujeto a que el vector de la suma del cuadrado de sus ponderadores tenga longitud unitaria. Sin embargo, este problema equivale a encontrar los eigenvalores y eigenvectores asociados a la matriz de covarianzas de los datos estandarizados, Σ . Vale la pena recordar que Σ es una matriz cuadrada semidefinida positiva; entonces si a Σ le asociamos el par de eigenvalores y eigenvectores $(\lambda_1, e_1), (\lambda_2, e_2), \dots, (\lambda_p, e_p)$ donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, el k -ésimo componente principal está dado por:

$$Y_k = e'_k X = e_{k1}X_1 + e_{k2}X_2 + \dots + e_{kp}X_p \quad (2)$$

El tercer paso implica calcular los coeficientes de correlación entre los componentes Y_k y los indicadores X_k , y obtener así la matriz de cargas factoriales Λ :

$$\begin{aligned} X_1 &= r_{11}Y_1 + r_{12}Y_2 + \dots + r_{1p}Y_p \\ X_2 &= r_{21}Y_1 + r_{22}Y_2 + \dots + r_{2p}Y_p \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ X_p &= r_{p1}Y_1 + r_{p2}Y_2 + \dots + r_{pp}Y_p \end{aligned} \quad (3)$$

De esta manera, a la suma horizontal de los cuadrados de las cargas factoriales se le denomina comunalidad (parte de la dispersión total de la variable explicada por el conjunto de p componentes). Una vez obtenidos los componentes es preciso realizar algunas pruebas estadísticas para descartar la inoperatividad del ACP, las pruebas más comunes son la de Kaiser-Mayer-Olkin (KMO) y la de esfericidad de Bartlett (EB), basadas en la matriz de correlaciones, R , de los indicadores originales y dados por:

$$KMO = \frac{\sum_{k \neq j} \sum_{k \neq j} \rho_{kj}^2}{\sum_{k \neq j} \sum_{k \neq j} \rho_{kj}^2 + \sum_{k \neq j} \sum_{k \neq j} \rho_{kj(k)}^2} \quad (4)$$

donde $\rho_{kj(k)}$ es el coeficiente de correlación parcial entre los indicadores I_k e I_j y elimina la influencia del resto de las variables. El indicador KMO está acotado en el intervalo $[0,1]$ y, según Kaiser (1970), si el valor de $KMO \geq 0.75$ es recomendable utilizar ACP; si $0.75 > KMO \geq 0.5$, ACP es aceptable, y si $KMO < 0.5$ la idea de realizar un análisis factorial es inaceptable. Por otro lado, el EB está dado por:

$$EB = - \left[n - 1 - \frac{1}{6}(2p + 5) \right] \ln |R| = - \left(n - \frac{2p + 11}{6} \right) \sum_{j=1}^p \ln \lambda_j \quad (5)$$

donde n es el número de manzanas en la muestra y λ_j son eigenvalores de la matriz Σ anteriormente descritos. El estadístico se distribuye asintóticamente χ^2 con $[p(p-1)]/2$ grados de libertad, esta prueba contrasta, bajo la hipótesis de normalidad multivariada, si la matriz de correlaciones, R , es identidad. Así pues, si el estadístico de la prueba toma valores grandes, se rechaza la hipótesis nula de matriz identidad con un cierto grado de significancia. En caso de no rechazarse, significaría que las variables no están correlacionadas, bajo este supuesto debería reconsiderarse la aplicación del análisis factorial.

Una vez descartada la inoperatividad del ACP se procede a la rotación *varimax* de los factores –transformación $Y_k \sim Z_k$ – dada por:

$$G = \sum_{o=1}^q \sum_{o \neq k}^q \left[\sum_{j=1}^p a_{jk}^2 a_{jo}^2 - \frac{\gamma}{p} \sum_{j=1}^p a_{jk}^2 \sum_{j=1}^p a_{jo}^2 \right] \quad (6)$$

donde γ es un parámetro tal que $0 \leq \gamma \leq 1$. Esta rotación es necesaria pues su finalidad es conseguir que unas correlaciones sean altas a costa de otras que serán bajas, para así destacar la influencia de los componentes comunes sobre las variables observables.

La penúltima etapa consiste en la construcción propiamente del índice de rezago social. Para decidir cuántos de los once componentes rotados (Z_k) serán considerados para la construcción del índice, se propone el uso del criterio de Kaiser, en el que se señala que obtener los componentes principales a partir de la matriz de correlaciones originales equivale a suponer que las variables observables tienen varianza unitaria. Por lo tanto, un componente principal con varianza

inferior a uno explica menos variabilidad que una observable. El criterio pues es retener los m primeros componentes tales que $\lambda_l \geq 1$ para $l = 1, 2, \dots, m$. En realidad, una diferencia sustancial esperada respecto a la construcción del índice del Coneval es la decisión de cuántos componentes utilizar, mientras que para esta institución el indicador para cada observación se compone únicamente del valor que tome el primer componente –dado que la regla de Kaiser se cumple para $m = 1$ –, al trabajar con unidades territoriales pequeñas la varianza explicada por el primer componente será baja respecto al total.⁶

Una vez decidido los m de componentes a utilizar, el índice de rezago social para cada manzana estará dado por la distancia euclidiana que hay respecto a la manzana menos rezagada. En teoría es posible identificar a esta manzana a partir de la matriz de correlación de los componentes rotados, Λ' , sin embargo, dada la posibilidad de encontrar varias manzanas que puedan cumplir con las mismas características, el criterio sugerido es calcular el valor que tomarían los m componentes si existiera una “manzana ideal”,⁷ dichos valores se encuentran a partir de:

$$\begin{aligned} \hat{Z}_1 &= \mathbf{a}'_1 \hat{X} = a_{11}\hat{X}_1 + a_{12}\hat{X}_2 + \dots + a_{1p}\hat{X}_p \\ \hat{Z}_2 &= \mathbf{a}'_2 \hat{X} = a_{21}\hat{X}_1 + a_{22}\hat{X}_2 + \dots + a_{2p}\hat{X}_p \\ &\vdots \qquad \qquad \qquad \vdots \qquad \qquad \qquad \vdots \\ \hat{Z}_m &= \mathbf{a}'_m \hat{X} = a_{m1}\hat{X}_1 + a_{m2}\hat{X}_2 + \dots + a_{mp}\hat{X}_p \end{aligned} \tag{7}$$

donde \hat{X} es el vector que contiene los valores estandarizados de la manzana i en el indicador j , I_{ij} . Una vez obtenidos los valores de los \hat{Z}_l componentes de la manzana ideal, el índice de rezago social para cada manzana de áreas metropolitanas estará dado por:

⁶ Esta situación permitiría superar una de las limitantes anteriormente señaladas, ya que en el cómputo oficial del rezago social sólo se ha seleccionado el primer componente, dándole menor peso a ciertos indicadores, pues aquéllos con menor peso en el primero tienen un mayor peso en los siguientes.

⁷ Aquella en la que $I_j=0 \forall j=1,2,\dots,p$.

$$IRS_i = \sqrt{\sum_{l=1}^m (\hat{Z}_l - Z_{il})^2} \quad (8)$$

Después de haber computado el IRS_i para las n manzanas que constituyen el área metropolitana de interés es posible ordenar las observaciones. De acuerdo con la matriz de correlaciones rotada Λ' , es sencillo verificar que el valor que toma la “manzana ideal” es igual a cero ($IRS_{ideal} = 0$) y que, entre mayor sea el valor que arroje este indicador para cada manzana, implicará necesariamente mayor rezago. De este modo el espacio muestral del índice estará dado por $\Omega = \{IRS_1, IRS_2, \dots, IRS_n | IRS_i \in (0, \omega] \text{ e } IRS_i \leq IRS_{i+1}\} \forall i$, para cualquier $\omega \in \mathbb{R}^+$ debido a que no se propone ninguna estandarización del índice.

Finalmente el último paso consiste en la estratificación, aunque en general es posible utilizar cualquier técnica para formar grupos relativamente homogéneos, se siguió el método de estratificación óptima desarrollado por Dalenius y Hodges (1959), el cual crea grupos de puntajes, minimizando la variabilidad interna y maximizando la externa, además es usado por el Coneval y el Conapo para la elaboración de sus índices.⁸

3.2. Método de máxima verosimilitud

El MMV es un método de extracción factorial que proporciona las estimaciones de los parámetros que con mayor probabilidad han producido la matriz de correlaciones observada, si la muestra procede de una distribución normal multivariada. Resulta importante tener clara la diferencia entre ambos métodos, ya que, mientras el ACP analiza toda la varianza común y no común, el MMV sólo analiza la varianza compartida –estimaciones de la comunalidad en la diagonal de Σ –. En este método los factores van explicando sucesivamente la máxima

⁸ Aunque en el apartado anterior se hizo énfasis en las críticas hacia este método y a pesar de que hay alternativas multivariadas más eficientes desde el punto de vista estadístico –como el análisis de conglomerados–, lo que realmente resulta sustancial es el hecho de que $IRS_i \leq IRS_{i+1}$. Entonces, al tener un ordenamiento de las manzanas acorde a su grado de rezago, los hacedores de políticas sociales podrán tomar decisiones en función del presupuesto disponible, sobre la cantidad de manzanas que será necesario atender.

proporción de varianza en la población, no en la muestra. La estructura factorial que resulta en ambos tipos de análisis es más o menos la misma, sin embargo, en algunas ocasiones se suele preferir el MMV, ya que ofrece unas estimaciones de los pesos más conservadoras, al tener en cuenta el error o varianza no compartida de las variables.

Igual que ACP, para iniciar el método es necesario estandarizar los indicadores originales de manera que $I_j \sim X_j$. Posteriormente, para probar que los datos provienen de una distribución aproximadamente normal, es necesario realizar el test de Kolmogorov-Smirnov para cada X_j . Una vez probada la distribución aproximadamente normal de cada uno de los indicadores estandarizados, podemos plantear la obtención de la matriz de factores como un problema de estimación de la matriz de covarianzas, Σ , con la restricción de que Σ se descomponga de la forma $\Sigma = AA' + \Psi$ donde Ψ es una matriz diagonal. Si asumimos que las n observaciones de los once indicadores provienen de una distribución normal con media cero ($\mu = 0$), entonces el logaritmo natural de la función de verosimilitud es $\ln L(X, \mu, \Sigma) = -\frac{n}{2} (\ln |2\pi\Sigma|) - tr(\Sigma^{-1}S)$ donde $S = (X_i - \bar{X})(X_i - \bar{X})'$. Al cambiar el signo y modificar algunas constantes el MMV trata de estimar A y Ψ , de manera que la función $F_p(A, \Psi) = \ln |\Sigma| + tr(\Sigma^{-1}S) - \ln |S| - p$ sea mínima, por lo tanto es necesario derivar F_p respecto a A y Ψ , lo que nos daría como resultado la maximización de la función de verosimilitud.

El siguiente paso consiste en realizar las pruebas de *KMO* (4) y *EB* (5) para descartar la inoperatividad del MMV. Posteriormente, los factores se rotan utilizando la expresión (6). Además, hay que hacer notar una ventaja del MMV frente al ACP, ya que éste permite formular una prueba de hipótesis sobre la estructura factorial de Σ y el número v de factores comunes. Johnson y Wichern (2001: 498) plantean la prueba de la forma $H_0 : \Sigma = AA' + \Psi$ vs. $H_1 : \Sigma$ es definida positiva, donde A es de rango v . Si $\hat{S} = \hat{A}\hat{A}' + \hat{\Psi}$, donde \hat{A} y $\hat{\Psi}$ son las estimaciones de máxima verosimilitud, los máximos del logaritmo de la razón de verosimilitud son:

$$H_0 : -\frac{n}{2} \left[\left(\ln |\hat{\Sigma}| \right) + tr \left(\hat{\Sigma}^{-1}S \right) \right] \quad vs \quad H_1 : -\frac{n}{2} (\ln |S| + p)$$

En Cuadras (2012: 105) se demuestra que es posible obtener el estadístico $C_c = n \left[\ln |\hat{\Sigma}| - (\ln |S| + tr(\hat{\Sigma}^{-1}S) - p) \right] = nF_p(\hat{A}, \hat{\Psi})$

que sigue asintóticamente la distribución χ^2 con

$$c = \frac{p(p-1)}{2} - \left[pv + p - \frac{v(v-1)}{2} \right] = \frac{1}{2} \left[(p-v)^2 - p - v \right]$$

grados de libertad. No obstante, Johnson y Wichern (2001: 499) advierten que la evidencia empírica ha mostrado que si n es grande y v es pequeña relativa a p –como es el caso que nos ocupa–, H_0 será rechazada sugiriendo que sea necesario utilizar un mayor número de factores.

El penúltimo paso consiste en construir el índice de rezago social a partir de los factores rotados. Con base en el criterio de Kaiser se espera $m = v$ y que haya concordancia con los resultados del ACP. Posteriormente, una vez identificada la “manzana ideal” se estima (8) y se sigue el criterio de estratificación de Dalenius y Hodges (1959).

4. El área metropolitana de Monterrey como ejemplo

Para ejemplificar la operatividad de la metodología se estimó el índice para las manzanas del AMM. El caso de esta zona es particularmente interesante ya que para los años 2000, 2005 y 2010 el estado de Nuevo León se ubicó, con respecto a las 32 entidades, en el último lugar de acuerdo con su grado de rezago social, es decir, ha sido la entidad menos rezagada del país en la primera década del siglo XXI. Sin embargo, la intención del ejercicio es mostrar la sensibilidad de los resultados –se considera que en el AMM habita 84.5% de la población de la entidad– si se parte del análisis de unidades geográficas más pequeñas.

La zona metropolitana se conforma por nueve municipios,⁹ y de acuerdo con el *Censo de población y vivienda 2010* existían en ella 341 localidades, mil 525 Agebs urbanos, 38 mil 899 manzanas y 994 mil 529 hogares que albergaban 3 millones 930 mil 338 habitantes, de los cuales 49.8% son hombres. En general para el AMM el promedio de hogares por manzana es de 25.6 y aproximadamente de 101 habitantes por manzana.

En el cuadro 2 se aprecia la relevancia del algoritmo de imputaciones para rescatar observaciones, ya que sin el proceso únicamente hubiera sido posible estimar el nivel de rezago para 70.4% de las

⁹ Apodaca, García, San Pedro Garza García, General Escobedo, Guadalupe, Juárez, Monterrey, San Nicolás de los Garza y Santa Catarina.

manzanas, mientras que con las imputaciones el porcentaje aumentó a 92.2. Por su parte, el cuadro 3 concentra la información pertinente a los indicadores antes de su estandarización.

Cuadro 2
Resultados del proceso de imputación

<i>Indicador</i>	<i>Sin imputación</i>		<i>Imputación: 1 y 2</i>		<i>Imputación: 1 y 2 y promedio por Ageb</i>	
	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>	<i>n</i>	<i>%</i>
1 I'esc	37 527	96.5	37 527	96.5	37 527	96.5
2 Iedbasinc	37 244	95.8	37 244	95.8	37 244	95.8
3 Isdersalud	35 497	91.3	36 673	94.3	36 673	94.3
4 Iptierra	31 349	80.6	36 455	93.8	37 345	96.0
5 Ihacin	37 527	96.5	37 527	96.5	37 527	96.5
6 Inosan	37 316	95.9	37 316	95.9	37 316	95.9
7 Inoagua	35 732	91.9	37 174	95.6	37 360	96.0
8 Inodren	36 285	93.3	37 246	95.8	37 378	96.1
9 Inoelec	37 333	96.0	37 334	96.0	37 407	96.2
10 Inolav	36 958	95.0	36 958	95.0	36 958	95.0
11 Inoref	37 220	95.7	37 220	95.7	37 220	95.7
<i>n válido</i>	27 388	70.4	34 765	89.4	35 852	92.2
<i>N</i>	38 899					

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Cuadro 3
Estadísticos descriptivos de los indicadores

<i>Indicador</i>	<i>Máximo</i>	<i>Mediana</i>	<i>Media</i>	<i>Des. est.</i>
1 I'esc	18	8.2	7.7	2.5
2 Iedbasinc	100	18.8	21.4	17.91
3 Isdersalud	100	19.5	20.7	11.2
4 Iptierra	100	0	1.8	6.3
5 Ihacin	5.5	96.5	0.9	0.3

Cuadro 3
(continuación)

<i>Indicador</i>	<i>Máximo</i>	<i>Mediana</i>	<i>Media</i>	<i>Des. est.</i>
6 Inosan	100	0.9	2	5.2
7 Inoagua	100	0	1.4	9.3
8 Inodren	100	0	0.6	5.2
9 Inoelec	100	0	1.6	4.6
10 Inolav	100	10.7	13.1	11.7
11 Inoref	100	2.4	4.7	7.1
<i>Mínimo = 0</i>				
<i>n = 35 852</i>				

Nota *La correlación entre la población de 15 años y más, analfabeta y la diferencia respecto al valor máximo del grado promedio de escolaridad fue de 0.52.

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Al realizar la prueba de Kolmogorov-Smirnov¹⁰ para verificar la operatividad del MMV se encontró que ninguno de los indicadores sigue una distribución normal. No obstante, podemos hacer uso del teorema del límite central si asumimos para cada indicador X_j que el conjunto de manzanas x_1, x_2, \dots, x_n son independientes e idénticamente distribuidas con media μ y varianza σ^2 distinta de cero; así podemos denotar a D_n como la sumatoria de las n variables aleatorias y entonces:

$$\lim_{n \rightarrow \infty} \Pr \left(\frac{(D_n - n\mu)}{\sigma\sqrt{n}} \leq z \right) = \Phi(z)$$

Previo a la utilización del ACP o del MMV para construir el índice presentamos los resultados de las pruebas de EB y de KMO. La prueba estadística EB indica que la matriz de correlaciones de los indicadores originales no es identidad, lo que sugiere con un nivel de significancia mayor a 0.01 el uso del ACP y del MMV. Por otro lado, el estadístico de KMO arrojó un valor de 0.747, lo que indica que es aceptable el uso de cualquier técnica de análisis factorial.

El cuadro 4 presenta los resultados del ACP y del MMV referidos al porcentaje de la varianza que es posible explicar según el método a

¹⁰ No se presentan los resultados pero están disponibles a solicitud.

utilizar. En general, y de acuerdo con el criterio de Kaiser, el uso de tres componentes derivados del ACP nos permite reflejar 68.4% de la variabilidad del total de los once indicadores estandarizados, es decir, con sólo tres variables fue posible explicar más de dos terceras partes del comportamiento original de las variables indicador. Por su parte, el mismo criterio citado, en el MMV sugiere el uso de tres factores que, aunque explican un menor porcentaje de la varianza original, 57%, la técnica tiene la ventaja de presentar mejores correlaciones reproducidas que el ACP.¹¹

Es importante aclarar que en el caso del MMV, contrastar la hipótesis sobre la estructura factorial de Σ y el número v de factores comunes contradice el uso de tres factores, de hecho¹² la prueba aconseja que la matriz reproducida es estadísticamente distinta de la observada. Sin embargo, hacemos hincapié en la evidencia empírica al respecto, la cual ha dejado ver que, cuando n es grande y v es pequeña relativa a p , generalmente la hipótesis nula será rechazada (Johnson y Wichern, 2005: 499).

Una vez identificado el número de componentes/factores a utilizar, cada uno de ellos es asociado a una dimensión del rezago en el cuadro 5. Podemos pues asociar el primer factor a la dimensión educativa indistintamente del modelo a utilizar; el segundo está correlacionado con el equipamiento del hogar y el tercero con las condiciones y espacios en la vivienda.

Al tener clara la cantidad de componentes a utilizar, y su asociación con las distintas dimensiones del rezago, lo que sigue es la construcción del índice. Para ello, la figura 1 pretende aportar una idea visual de este procedimiento, ubica en espacio euclídeo la posición que toma cada manzana respecto al valor de cada uno de los tres componentes principales.

Posterior al cómputo del índice y con la metodología de estratificación de Dalenius y Hodges (1959), el cuadro 6 muestra los puntos de corte tanto para el ACP como el MMV que representan los estratos del rezago social.

¹¹ La matriz de correlaciones reproducida se obtiene a partir de la solución factorial hallada. Si el modelo es bueno y el número de factores adecuado, la estructura factorial debe ser capaz de reproducir la matriz de correlaciones y en la diagonal de la matriz reproducida se encuentran las comunialidades.

¹² El estadístico con $v=3$ arrojó un $\chi^2=11141.9$, que excede el valor crítico $\chi^2_{33}(.05)=47.4$; para $v=4,5$ y 6 los estadísticos calculados fueron 2 070.5, 782.8 y 23.7, se excedieron los valores críticos 5% pero se apreció la sensibilidad de la prueba.

Cuadro 4
Valor de la varianza explicada

<i>Componente/ Factor</i>	<i>Eigenvalores iniciales</i>			<i>Suma de las saturaciones al cuadrado de la rotación</i>					
	<i>Total</i>	<i>% de la varianza</i>	<i>% acumu- lado</i>	<i>ACP</i>			<i>MMV</i>		
				<i>Total</i>	<i>% de la varianza</i>	<i>% acumu- lado</i>	<i>Total</i>	<i>% de la varianza</i>	<i>% acu- mulado</i>
1	3.7	33.2	33.2	3.0	27.3	27.3	2.6	23.2	23.2
2	2.5	22.3	55.5	2.7	24.3	51.6	2.4	21.5	44.7
3	1.4	13.0	68.4	1.9	16.8	68.4	1.3	12.0	56.7
4	0.8	7.0	75.5	-	-	-	-	-	-
5	0.7	6.3	81.8	-	-	-	-	-	-
6	0.6	5.0	86.8	-	-	-	-	-	-
7	0.5	4.6	91.4	-	-	-	-	-	-
8	0.4	3.8	95.2	-	-	-	-	-	-
9	0.3	2.5	97.7	-	-	-	-	-	-
10	0.1	1.2	98.9	-	-	-	-	-	-
11	0.1	1.1	100	-	-	-	-	-	-

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Cuadro 5
Matriz de componentes/factores rotados

Indicador	ACP				MMV			
	Componentes rotados				Factores rotados			
	1	2	3	Comuna- lidad	1	2	3	Comuna- lidad
1 Iesc	.9	.0	.1	.8	1.0	.0	.1	1.0
2 Iedbasinc	.9	.0	.1	.8	.8	.0	.2	.7
3 Isdersalud	.7	.0	.0	.4	.5	.0	.2	.2
4 Iptierra	.2	.1	.6	.4	.1	.1	.5	.3
5 Ihacin	.7	.1	.3	.6	.6	.1	.4	.5
6 Inosan	-.1	.9	.1	.8	-.1	.9	.1	.8
7 Inoagua	.1	.1	.8	.7	.1	.1	.5	.3
8 Inodren	.0	.0	.8	.6	.1	.1	.5	.2
9 Inoelec	-.1	.9	.0	.9	-.1	1.0	.0	1.0
10 Inolav	.5	.6	.2	.6	.4	.4	.4	.5
11 Inoref	.3	.8	.2	.8	.2	.7	.4	.7

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Cuadro 6
Manzanas, hogares y personas clasificadas según su grado de rezago social

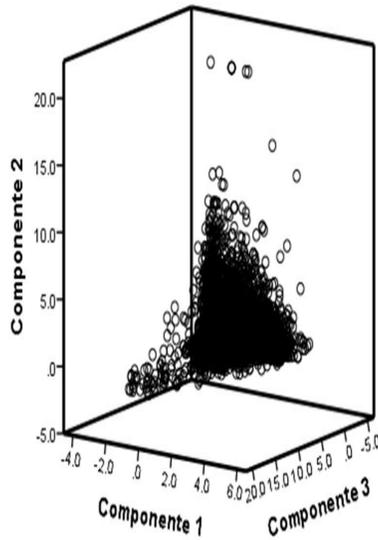
ACP				
Estrato	Punto de corte	Manzanas	Hogares	Habitantes
Muy bajo	- 2.094	7,500	185,437	671,722
Bajo	2.095 - 3.053	10 069	291 887	1 136 560
Medio	3.054 - 4.120	9 094	257 499	1 055 864
Alto	4.121 - 6.286	5 943	161 314	680 284
Muy alto	6.287 - 418.779	3 246	70 848	275 008
MMV				
Muy bajo	- 2.470	6 616	167 421	602 334
Bajo	2.471- 3.649	8 919	251 321	965 799
Medio	3.650-4.590	10 803	312 508	1 287 815

Cuadro 6
(continuación)

Alto	4.591	6.841	6 127	161 267	673 680
Muy alto	6.842	518.559	3 387	74 468	289 810
Total			35 852	966 985	3 819 438

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Figura 1
Dispersión de las manzanas según el valor de sus componentes, ACP



El cuadro 7 muestra el total del total de manzanas que coincidieron al utilizar ambas metodologías. Resultan particularmente interesantes los casos que se extrapolan, es decir, aquellas manzanas que bajo la primera metodología aparecen con un grado de rezago social muy bajo y con la segunda aparecen con un grado de rezago muy alto, en general se esperaría que la matriz de traslape concentrara

a la mayoría de las manzanas en la diagonal y que los valores fuera de ésta sean pequeños. Aunque tal discrepancia pudiera representar una adversidad al momento de la instrumentación de alguna política social focalizada, un breve análisis exploratorio de datos permite verificar que la correspondencia en el último decil fue de 95.5% de las manzanas, mientras que en el último cuartil fue de 76 por ciento. Es decir, uno de los criterios sugeridos podría ser el atender al decil más rezagado por ambas metodologías o a los últimos dos deciles. Estadísticamente, la discrepancia de resultados es atribuida a la forma en que operan ambas técnicas, en tanto que el ACP analiza toda la varianza, el MMV analiza únicamente la varianza compartida. Sin embargo, hasta este punto no se ha llevado a cabo ningún análisis de sensibilidad derivado de la instrumentación del algoritmo, por lo que sería recomendable, en trabajos posteriores, validar su sensibilidad y determinar los umbrales para los cuales debe operar.

Cuadro 7
Matriz de traslape ACP-MMV

		MMV					
		Muy bajo	Bajo	Medio	Alto	Muy alto	Total
ACP	Muy bajo	5 868	1 428	162	37	5	7 500
	Bajo	733	5 860	3 041	419	16	10 069
	Medio	14	1 550	5 391	2 086	53	9 094
	Alto	1	74	2 179	3 193	496	5 943
	Muy alto	0	7	30	392	2 817	3 246
	Total	6 616	8 919	10 803	6 127	3 387	35 852

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

El cuadro 8 agrupa al total de habitantes que se ubicaron en cada uno de los cinco estratos en ambas técnicas. Al comparar los resultados con trabajos anteriores, el cuadro 9 presenta en negritas nuestros resultados, se encontró que la aproximación más cercana es la proporcionada por Martínez, Treviño y Gómez (2009). Sin embargo, en el vértice opuesto están situadas las mediciones oficiales, en las que se estipula que a nivel localidad únicamente 14 individuos no tienen las condiciones necesarias para un desarrollo personal sano.

Cuadro 8

Número de habitantes según el grado de rezago social por municipio

<i>Municipios</i>									
<i>ACP</i>									
<i>Estrato</i>	<i>Apodaca</i>	<i>García</i>	<i>San Pedro</i>	<i>General Escobedo</i>	<i>Guadalupe</i>	<i>Juárez</i>	<i>Monterrey</i>	<i>San Nicolás</i>	<i>Santa Catarina</i>
Muy bajo	65 815	9 534	46 860	46 162	110 379	7 897	208 162	143 702	33 211
Bajo	226 797	23 815	25 612	101 469	213 571	75 175	208 695	159 451	101 975
Medio	134 395	53 137	18 938	98 185	212 502	89 619	285 933	92 449	70 706
Alto	55 851	31 841	13 403	74 828	93 979	45 602	289 457	32 732	42 591
Muy alto	24 672	15 578	9 509	26 190	38 802	23 186	112 783	9 343	14 945
<i>MMV</i>									
Muy bajo	48 438	9 700	44 225	38 332	101 111	5 142	199 681	130 061	25 644
Bajo	179 935	25 317	23 466	81 834	177 073	59 630	209 158	132 669	76 717
Medio	169 201	57 323	19 303	145 403	225 546	109 302	353 906	110 485	97 346
Alto	77 039	28 137	16 131	56 321	119 061	44 459	230 924	52 647	48 961
Muy alto	32 917	13 428	11 197	24 944	46 442	2 946	111 361	11 815	14 760
Total	507 530	133 905	114 322	346 834	669 233	241 479	1 105 030	437 677	263 428

Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Cuadro 9

Población en situación de rezago y marginación según estrato para el AMM

<i>Estrato</i>	IRS 2010 <i>(Manzana ACP)</i>	%	IRS 2010 <i>(Manzana MMV)</i>	%	<i>IRS 2005</i> <i>(Manzana)</i>	%	<i>IRS 2010</i> <i>Coneval</i> <i>(Localidad)</i>	%	<i>IM 2010</i> <i>Conapo</i> <i>(Ageb-ZM-Monterrey)</i>	%
Muy bajo	671 722	17.6	602 334	15.8	1 048 781	29.3	3 921 942	99.8	1 804 849	44.7
Bajo	1 136 560	29.8	965 799	23.3	868 599	24.2	4 537	0.1	1 038 807	25.7
Medio	1 055 864	27.6	1 287 815	33.7	713 103	19.9	2 472	0.1	1 009 952	25.0
Alto	680 284	17.8	673 680	17.6	624 465	17.4	738	0.0	159 172	3.9
Muy alto	275 008	7.2	289 810	7.6	330 453	9.2	14	0.0	23 935	0.6

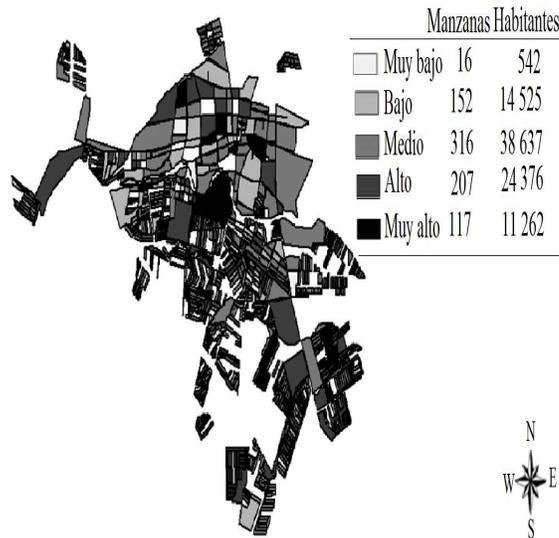
Fuente: Elaboración propia con base en información obtenida de Martínez, Treviño y Gómez (2009), Coneval (2011) y Conapo (2012).

Por su parte el Conapo, en una postura menos ortodoxa, señala que cerca de uno por ciento de la población vive alguna situación de marginación.¹³

Una vez identificada la población objetivo es posible utilizar instrumentos de focalización como herramientas para el combate a la pobreza. El objetivo es determinar geográficamente la ubicación de los habitantes más rezagados, lo que eventualmente facilitaría la operación de los programas sociales dirigidos a la atención de los grupos de la población más necesitada.

En las figuras 2, 3 y 4 se ejemplifica el uso de los instrumentos de focalización geográfica para el municipio de García con el uso del ACP, su elección como ejemplo se basó en el hecho de que el municipio concentra la mayor proporción de habitantes en situación de muy alto rezago respecto a su población total.

Figura 2
Mapa de rezago social, García



¹³ Vale la pena reiterar que estos estudios reflejan un análisis comparativo entre todas las localidades/Agabs del país, lo que implica un ordenamiento de observaciones que contienen a toda la población de México. No obstante, hacemos notar que los contextos en los que viven los habitantes de determinadas zonas geográficas aunque pertenezcan a un mismo país, generalmente son muy distintos.

Figura 3
Casos extremos de rezago social, García

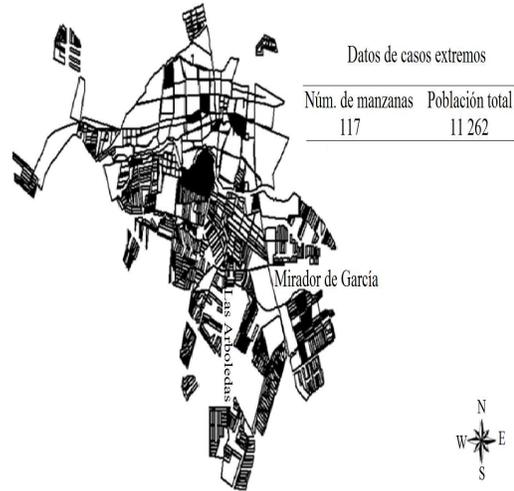
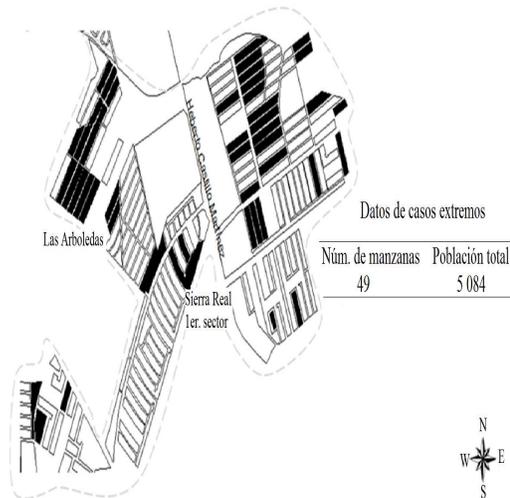


Figura 4
Polígono de rezago social, García



Fuente: Cálculos propios con base en el *Censo de población y vivienda 2010* (INEGI, 2011).

Nótese que el total de habitantes no coincide con los reportados en el cuadro 8, debido a que se excluyeron dos localidades que geográficamente no son colindantes con la región mostrada. La figura 4 es de especial importancia, ya que representa un polígono de rezago social, el cual, por ejemplo, agrupa a un total de 49 manzanas y 5 084 habitantes. La utilidad de esta representación visual del fenómeno radica en su facilidad para entender las condiciones de precariedad en que allí se vive, además de la reproducción geoespacial del rezago social al tratarse de una zona con alta concentración de habitantes en esta condición.

5. Conclusiones

En esta investigación se presenta una metodología que permite estimar el grado de rezago social para las manzanas de las áreas metropolitanas de México, se complementa así la información existente respecto a rezago social. A diferencia de los 12 indicadores utilizados por el Coneval para la construcción del indicador, esta propuesta utiliza once, sustituye el porcentaje de la población de 15 años y más analfabeta por la diferencia respecto al valor máximo del grado promedio de escolaridad, además de no considerar el porcentaje de la población de 6 a 14 años que no asiste a la escuela. Estos cambios fueron necesarios por dos razones; para la primera variable, el *Censo de Población y Vivienda 2010* no reporta información suficiente y para la segunda, al trabajar con manzanas hay un gran número de observaciones en donde no existe población entre 6 y 14 años.

En aras de cuantificar el grado de rezago social para un tamaño de muestra mayor y ante los criterios de confidencialidad del INEGI se propone el uso de un algoritmo de imputaciones en donde, por diferencia e imputar promedios de valores por Ageb, es posible recuperar valores para los once indicadores e incrementar el tamaño de muestra. La evidencia empírica para el caso del AMM sugiere que el algoritmo resulta eficiente, ya que permitió estimar el nivel de rezago de 92.2% del total de observaciones, mientras que sin algoritmo el porcentaje hubiera sido de 70.4.

Para la elaboración del índice se propuso el uso dos técnicas distintas, lo que permitió robustecer los resultados. En primer lugar, igual que el Coneval, se hizo uso del análisis de componentes principales, asimismo, se propuso el método de máxima verosimilitud como técnica de reducción. Para el caso de los componentes principales los pesos tienden a estar sobre estimados, particularmente si el número de

variables a reducir es pequeño, por lo que hay que tomar los resultados con cautela ya que lo que se está analizando es toda la varianza, y es más frecuente que un indicador tenga pesos en varios componentes, es decir, su varianza se puede explicar por varios componentes. Por su parte, el método de máxima verosimilitud puede dar estimaciones más conservadoras al tener en cuenta la varianza no compartida de las variables, y termina por reproducir mejor la matriz de correlaciones. Otra de las ventajas del uso de este segundo método es que nos brinda la posibilidad de realizar una prueba para comprobar si el número de factores a considerar es adecuado, y encuentra que para el caso del área metropolitana de Monterrey es necesario utilizar un factor más en comparación con el análisis de componentes principales.

En el ejercicio para el área metropolitana de Monterrey, ambos métodos arrojaron resultados similares –cuadro 8– lo que sugiere la validez de las aseveraciones aquí expuestas, aunque habría que seguirlas con precaución ya que se encontraron casos atípicos que posicionaban a una misma manzana en estratos completamente opuestos. Es por ello que es recomendable realizar ambos análisis para quedarnos con la solución más clara.

Una importante diferencia respecto al índice propuesto por el CONEVAL es el número de componentes/factores a utilizar y la forma en cómo se construye; mientras que a nivel estatal, municipal y de localidad es suficiente el uso de un solo componente, los resultados empíricos sugieren que al estimarlo a nivel manzana es necesario utilizar tres de ellos. Asimismo, otra de las aportaciones de la investigación es el uso de la “manzana ideal” como pivote para estimar el grado de rezago, lo que sugiere una visión positiva del fenómeno.

Por otro lado, el hecho de que la medición multidimensional de la pobreza no pueda ser vinculada a características geográficas definidas y que el índice de rezago social/marginación sea calculado cuando más desagregado para localidades/Agabs, constituye una seria limitación en el combate a la pobreza en México. En este sentido, nuestro trabajo hace una importante aportación en materia de focalización, pues al vincular el grado de rezago social para cada una de las manzanas con la cartografía urbana es posible identificar, espacialmente, aquella población que necesita de manera prioritaria el servicio de los programas de atención gubernamental.

Futuras extensiones de este trabajo son el cómputo del índice para el resto de las AMMX y el respectivo contraste respecto a los resultados oficiales, así como la elaboración de los mapas que permitan identificar el grado de rezago del total de las 59 AMMX, ya que así sería posible identificar lo urgente de lo necesario, además de priorizar

el gasto social. Conjuntamente, se sugiere validar el algoritmo para el resto de las AMMX y verificar el efecto que tiene en aquellas manzanas que se traslapan entre los diferentes grados de rezago, según la metodología instrumentada, pues a pesar de que se establecieron criterios para su utilización habría que estimar el índice para diversas áreas metropolitanas y poder determinar umbrales de imputación acorde a la sensibilidad. Finalmente, sin ánimo de ser concluyentes, se reconocen las limitaciones de la metodología propuesta y se plantea la adecuación del método acorde a las necesidades y disponibilidad de la información.

Referencias

- Bigman, D. y H. Fofack. 2000. *Geographical Targeting for Poverty Alleviation. Methodology and Applications*, World Bank Regional and Sectoral Studies, Washington, D.C.
- Bistrain, C. 2010. Revisión de los índices de marginación elaborados por el Conapo, *Estudios Demográficos y Urbanos*, 25(1): 175-217.
- Cahill, M. 2007. Is the Human Development Index Redundant? *Eastern Economic Journal*, 31(1): 1219.
- Camberos, M. y J. Bracamontes. 2007. Marginación y políticas de desarrollo social: un análisis regional para Sonora, *Problemas del Desarrollo*, 38(149): 113-135.
- Cárdenas, O. 2010. Cardenalización del índice de marginación: una metodología para evaluar la eficiencia del gasto ejercido en el Ramo 33, *EconoQuantum*, 7(1): 41-66.
- Conapo. 2006. *Índice de marginación*, 2005, México.
- . 2011. *Índice de marginación por entidad federativa y municipio, 2010*, México.
- . 2012. *Índice de marginación urbana, 2010*, México.
- Coneval. 2007. *Los mapas de la pobreza en México*, México.
- . 2011. *Índice de rezago social, 2010*, México.
- Cuadras, C. 2012. *Nuevos métodos de análisis multivariante*, CMC Editions, Barcelona.

- De la Torre, R. (2004). *El índice de desarrollo humano y la asignación del gasto público por entidad federativa en México*, Estudios sobre Desarrollo Humano, PNUD, México.
- Delenius, T. y J. Hodges. 1959. Minimum Variance Stratification, *Journal of American Statistical Association*, 54(285): 88-101.
- Evalúa DF. 2011. *Índice de desarrollo social de las unidades territoriales del Distrito Federal*, Consejo de Evaluación del Desarrollo Social del Distrito Federal, México.
- Gutiérrez, H. y V. Gama. 2010. Limitantes de los índices de marginación de Conapo y propuesta para evaluar la marginación municipal en México, *Papeles de Población*, 16(66): 227-257.
- INEGI. 2005. *Conteo de población y vivienda 2005*, <http://www.inegi.org.mx/est/contenidos/proyectos/ccpv/cpv2005/default.aspx>
- . 2011. *Censo de población y vivienda 2010*, Aguascalientes, México
- Johnson, R. y D. Wichern. 2005. *Applied Multivariate Statistical Analysis*, Prentice Hall, Fifth Edition.
- Kaiser, H. (1970). A Second Generation Little Jiffy, *Psychometrika*, 35(4): 401-415.
- LGDS. 2004. *Ley general de desarrollo social*, Cámara de Diputados del H. Congreso de la Unión, México, DOF 20-1-2004.
- López, L., C. Rodríguez y R. Vélez. 2006. Estimación del IDH estatal en México, análisis de sensibilidad a distintas decisiones metodológicas y comparaciones internacionales, en M. Székely y L. López (comps.) *La medición del desarrollo humano en México*, Fondo de Cultura Económica, México.
- Martínez, I., J. Treviño y M. Gómez. 2009. *Mapas de pobreza y rezago social área metropolitana de Monterrey*, Consejo de Desarrollo Social, Nuevo León, México.
- PNUD. 2010. *Informe sobre desarrollo humano 2010*, <http://hdr.undp.org/es/informes/mundial/idh2010/>
- Ringen, S. 1988. Direct and Indirect Measures of Poverty, *Journal of Social Policy*, 17(3): 351-365.
- Spicker, P. 1999. *Poverty and Social Security*, Routledge, Londres.
- Székely, M. y L. López. 2006. *La medición del desarrollo humano en México*, Fondo de Cultura Económica, México.