

# Métodos de focalización en la política social en México

## Un estudio comparativo

Daniel Hernández Franco, Mónica Orozco Corona  
y Sirenia Vázquez Báez\*

Fecha de recepción: 15 de febrero de 2006; fecha de aceptación: 18 de septiembre de 2007.

*Resumen:* Las estrategias de focalización consisten en dirigir las acciones a una población o territorio definidos, con el fin de lograr la eficiencia en la gestión de los recursos. El objetivo de este trabajo es describir y analizar algunas de las herramientas que se utilizan en la política social, para focalizar los programas hacia la población que vive en condiciones de pobreza. Utilizando la información de la Encuesta Nacional de Ingresos y Gastos de los Hogares de 2002, se comparan tres métodos estadísticos y econométricos para identificar la población objetivo: un modelo de Análisis Discriminante, un modelo Logit y un modelo Logit Multinivel. Los resultados obtenidos indican que el modelo de Análisis Discriminante, método utilizado por Sedesol para identificar la población en condiciones de pobreza, tiene los errores de exclusión más bajos, de acuerdo con la línea de pobreza determinada y con su profundidad y severidad, por lo que se considera que éste es el método más eficiente para identificar la población objetivo.

*Palabras clave:* focalización, análisis discriminante, modelos jerárquicos, modelo Logit, error de inclusión, error de exclusión.

*Abstract:* This paper tries to describe and analyze some of the tools used in social policy to target poverty reduction programs in Mexico. Targeting strategy programs center public actions to a reasonable well defined population or territory, in order to achieve a more focused efficiency in the resource allocation stage. Using data from the National Household Survey of Income and Expenditure (ENIGH) 2002, three econometric models are used to identify the target population: Discriminant Analysis, a Logit Model and a Hierarchical Logit Model. The results show that Discriminant Analysis, which is the main tool used by Sedesol to identify individuals in poverty conditions, has the lowest undercoverage

---

\* Daniel Hernández Franco, hernandezfranco@gmail.com, Coordinador de Asesores de la Secretaría de Educación Pública. Mónica E. Orozco Corona, monicao33@prodigy.net.mx, Consultora. Sirenia Vázquez Báez, svazquez@vector.com.mx, Analista, VECTOR Casa de Bolsa, S.A. de C.V., Blvd. Manuel Ávila Camacho 24, piso 14, Lomas de Chapultepec, 11000 México, D.F. Tel. (55) 5262 3600, Ext. 3245.

error for the three measures of the Foster-Greer-Thorbecke poverty index. The result leads to conclude that this method proves to be the most efficient in identifying population in poverty conditions, given the available information.

*Keywords:* targeting, discriminant analysis, hierarchical models, Logit model, leakage error, undercoverage error.

*Clasificación JEL:* I32, I38

## **Introducción**

**L**as estrategias de focalización consisten en dirigir las acciones a una población o territorio definidos, con el fin de lograr la eficiencia en la gestión de los recursos. Esta orientación considera las peculiaridades de las poblaciones y las regiones, para desarrollar mecanismos adecuados que cumplan con los objetivos establecidos. Ante recursos escasos para atender a todos los individuos o a todas las necesidades, es importante asegurar que se beneficien quienes más lo necesitan y, al mismo tiempo, no destinar recursos a quienes no se encuentren en una situación apremiante.

En el contexto de política pública, la principal motivación para la focalización proviene de tres factores: 1) maximizar la reducción de la pobreza, 2) limitar eficientemente los recursos destinados a la disminución de la pobreza y 3) aprovechar el costo de oportunidad entre el número de beneficiarios y el monto de las transferencias. Con estos tres objetivos se busca obtener un mayor impacto per cápita que el que podría derivarse de una política general que se aplica por igual a toda la población. En la última década, tanto en México como en muchos otros países, se ha impulsado la implantación de políticas focalizadas dentro del ámbito del desarrollo social. Sin embargo, ésta no ha sido una tarea sencilla, pues para que la focalización se realice adecuadamente es necesario contar con información estratégica. La información de la población, sus comunidades y sus características, es una herramienta clave para la generación de evidencia y la evaluación de los resultados e impacto de los programas, como parte del proceso de mejora continua en términos de diseño e implementación.

Nuestra motivación en este documento es despertar el interés en una mayor investigación y mejora de los mecanismos de focalización. Conocer algunas de las metodologías que se utilizan, tanto en los programas que se implementan en México como en otros países, es fundamental para ubicar el contexto en el que se desarrolla actualmente la política social de nuestro país. Esto, a su vez, contribuirá al perfeccionamiento de los mecanismos implementados en términos de información, metodología y

aplicación. El objetivo general de este trabajo es documentar algunos de los métodos de focalización enfocados a la reducción de la pobreza, que se utilizan en la política social tanto en México como en otros países. Mediante el análisis comparativo de tres herramientas de focalización, se buscará validar aquel que se utiliza en la selección de beneficiarios de los programas sociales de Sedesol, incluyendo el programa Oportunidades, en términos de su eficacia para identificar a la población en condiciones de pobreza.

El documento se organiza de la siguiente manera: en la sección I se describe brevemente la bibliografía existente; en la sección II se explican los principios generales y características de los distintos tipos de apoyos focalizados; en la sección III se detalla la metodología estadística y econométrica, y se presentan los resultados comparativos de los métodos utilizados. Finalmente, en la sección IV se presentan las conclusiones.

## **I. Una breve discusión bibliográfica**

Una de las primeras contribuciones teóricas fue el análisis que sobre la distribución del ingreso óptimo mediante la focalización llevó a cabo Akerlof (1978). Algunos años después Besley y Kanbur (1988) propusieron un sencillo modelo de minimización de pobreza partiendo de una “solución ideal”, un escenario perfecto de focalización en el cual, una vez identificados los individuos que se encuentran bajo una línea de pobreza conocida, no existen costos de administración ni otros incentivos que impidan al Estado cerrar la brecha de los que se encuentran por debajo de esta línea. Este modelo fue ampliado por Ravallion y Chao (1989), al incluir un escenario de información imperfecta y cuantificar los beneficios de la focalización en programas implementados en Bangladesh, Indonesia, Filipinas y Sri Lanka.

Sobre el método más adecuado de focalización no existe un consenso en el nivel teórico, pero se han desarrollado estudios interesantes tales como el de Glewwe *et al.* (1990), en el que plantea un método de focalización a partir de un modelo de minimización de pobreza, utilizando programación lineal; Thorbecke (2003) contribuyó a la identificación de población en condición de pobreza utilizando un modelo de equilibrio general; Abul Naga (2003) desarrolló un modelo al cual incorporó la incertidumbre, entre otros.

La bibliografía empírica sobre focalización es amplísima; en ella podemos encontrar desde trabajos de documentación, análisis y evaluación para

un solo programa en un país determinado, hasta recopilaciones y comparaciones entre conjuntos de programas, ya sea clasificados por región, tipo de apoyo o método de focalización. Más adelante en el documento daremos una descripción más amplia sobre la experiencia internacional y en México con los programas focalizados. Es importante resaltar que una buena parte de éstos se ha llevado a cabo a nivel institucional, principalmente por el Banco Mundial, que durante los últimos 30 años ha recopilado, estudiado e incluso implementado una gran variedad de programas focalizados alrededor del mundo. Los estudios realizados por este último abarcan desde el análisis de implementación de políticas focalizadas para reducir la desnutrición y la pobreza en países en desarrollo (Reutlinger, 1976), hasta uno de los más recientes e importantes trabajos de recopilación y comparación de experiencias internacionales, llevado a cabo por Coady y Grosh (2004), pasando por otras relevantes contribuciones como las de Grosh (1994) sobre Latinoamérica, y Mateus (1983) con un análisis de costo-beneficio en programas sociales de Latinoamérica y Asia, entre otros. En México la mayoría de los estudios de caso han sido también a nivel institucional, llevados a cabo por el Banco Mundial (Orozco, 2005; Skoufias, 2006), universidades como la Facultad Latinoamericana de Ciencias Sociales (FLACSO) (Herrera, 2000), el Centro de Investigación y Docencia Económicas (CIDE) (Scott, 2000), El Colegio de la Frontera Norte (Santibáñez *et al.*, 2005), y organismos públicos como el Instituto Nacional de Salud Pública (INSP) (Gutiérrez *et al.*, 2003).

En los estudios mencionados anteriormente se llega a la conclusión de que, aunque con algunas limitaciones, en general los programas focalizados han funcionado. El argumento a favor de estos programas podría ser bien resumido, de acuerdo con Amartya Sen (1995), en que mientras más exacto sea un subsidio en llegar a los pobres, habrá un menor desperdicio de recursos y un menor costo para llegar al objetivo deseado, que en este caso es la reducción de la pobreza. Por su parte, los modelos teóricos, bajo ciertos supuestos, han permitido demostrar que en términos económicos, la focalización es un medio óptimo para reducir la pobreza, aun si se consideran ciertas restricciones que se presentan en la realidad, como incentivos a no trabajar cuando los hogares reciben la transferencia (Kremer, 1997), o problemas de información y evaluación imperfecta para identificar a los individuos que necesitan los apoyos (Sheshinki y Diamond, 1992). Existe incluso evidencia que podría afirmar que en aquellos casos en los que la focalización no ha funcionado ha sido debido a problemas de implementación y no a la estructura del programa. Coady

y Grosh (2004) proporcionan argumentos que aseguran la posibilidad de reducir costos mientras más precisa y eficiente sea la focalización.

Algunos de los análisis que se han llevado a cabo para medir la eficiencia de una política focalizada en comparación con una política de corte general, muestran por una parte que los apoyos generalizados implican tasas de fuga de recursos sustantivas y una menor eficiencia en la atención a los más pobres (Skoufias, Davis y Behrman, 2000), y por otra reflejan que se cumplen dos de los objetivos de las estrategias focalizadas: dirigir las acciones hacia los segmentos de población de menores recursos con el fin de mejorar sus condiciones de vida, y disminuir la desigualdad respecto de aquellos sectores de la población que cuentan con mayores recursos (Skoufias *et al.*, 2000).

Los argumentos en contra de la focalización sostienen que su aplicación genera dos tipos de ineficiencia: que se puedan incluir personas cuando no necesitan el apoyo o que se excluyan algunas que sí lo necesitan. Sin embargo, hay casos en los cuales estos argumentos no son los únicos que cuestionan el desempeño de la focalización; por ejemplo, cuando intervienen otras variables como cambios de política que afectan la recaudación de impuestos y por lo tanto el presupuesto (Gelbach y Pritchett, 1997), o lo que Keen (1992) ha denominado “la paradoja de la focalización”, en la cual conforme aumentan las necesidades de la población atendida el presupuesto destinado a satisfacerlas se reduce. Esta paradoja también se puede aplicar cuando el tamaño o el tipo del hogar cambia, como lo demuestra Ebert (2005), ya que el impacto del beneficio focalizado será diferente en cada hogar, resultando en un impacto adverso en el agregado.

## II. Concepto y tipos de apoyos focalizados

De forma simplificada puede decirse que los apoyos focalizados son aquellos que se dirigen hacia grupos de población que presentan características específicas o que se encuentran habitando en zonas delimitadas territorialmente. Parten de la premisa de que no existe un acceso equitativo a los bienes o servicios para toda la población, y que sólo una direccionalidad intencionada ayuda a que quienes menos tienen puedan superar los obstáculos para el aprovechamiento de los apoyos.

Actualmente existen tantos métodos de focalización como programas para la reducción de la pobreza, cada uno con sus ventajas y limitaciones, sin que exista un consenso sobre cuál es el mejor. Con el fin de simplificar, consideramos seis tipologías en las cuales se pueden clasificar los

métodos de focalización. En el anexo A2 aparecen las seis tipologías de focalización, sus características, ventajas, desventajas, los programas sociales de la Sedesol focalizados según cada tipología, algunos ejemplos de programas en otros países y algunos de los estudios de evaluación que se han llevado a cabo.

Como se puede apreciar, prácticamente todos los programas focalizados de la Sedesol combinan distintos métodos. Por ejemplo, el programa de abasto social de leche Liconsa utiliza cuatro de estos tipos. El objetivo de este programa es contribuir a la alimentación y nutrición de los niños en edad preescolar, a las mujeres en periodo de lactancia y a los adultos mayores que viven en familias en condiciones de pobreza, mediante el suministro de leche fortificada a un precio mucho menor al de mercado, otorgando un beneficio al ingreso de estas familias (Sedesol, 2006). El primer método que utiliza es el de focalización geográfica, pues las lecherías Liconsa se encuentran en comunidades donde existe una mayor concentración de pobreza. En segundo lugar existe un proceso de autofocalización, pues la población es la que solicita pertenecer al programa para adquirir la leche. En tercer lugar se lleva a cabo una comprobación de medios aproximados para obtener un “puntaje”, y de acuerdo con las características socioeconómicas del hogar se evalúa si la familia es elegible para recibir el beneficio; además, este método se combina con una focalización demográfica, pues aquellas familias que tienen niños en edad preescolar, mujeres adolescentes o en lactancia, enfermos crónicos o con discapacidad, o adultos mayores, obtienen un puntaje mayor. De esta misma manera, programas como Oportunidades, Hábitat, 3x1 para Migrantes, el programa de atención a adultos mayores en zonas rurales, etc., combinan algunos de estos seis métodos de focalización.

### *II.1. Modelos de identificación de pobreza para México*

La metodología para la medición de la pobreza en México, elaborada por el Comité Técnico de Medición de la Pobreza (CTMP), que da lugar a las mediciones oficiales del gobierno de México, proporciona información para determinar la magnitud de la pobreza (Cortés *et al.*, 2000).<sup>1</sup> Sin embargo,

---

<sup>1</sup> En el año 2001, la Secretaría de Desarrollo Social convocó a un grupo de expertos académicos para que de manera independiente definieran una metodología para la medición de la pobreza en México. Como resultado, en el año 2002 se publicó dicha metodología, elaborada con la información de la Encuesta Nacional de Ingresos y Gastos de los Hogares, ENIGH 2000. La metodología se basa en la comparación del ingreso per cápita del hogar con el costo de una

estas medidas basadas en el ingreso no se utilizan en los programas focalizados del gobierno. La primera razón es que desde el año 2000 el enfoque de la política social en México ha tratado de ser integral, esto es, considerar la pobreza no solamente como la falta de ingreso, sino tal como lo define Amartya Sen (2003), como la carencia de una o varias necesidades básicas que impiden a un individuo alcanzar el desarrollo de todas sus capacidades.<sup>2</sup> La segunda razón se basa en un problema de información, pues en las encuestas de hogares, principal insumo con el que contamos para las mediciones de pobreza, no se cuenta con información detallada sobre el ingreso de cada miembro de la población, lo cual impide identificar de manera directa y exacta a los beneficiarios de dichos programas.

Tomando en cuenta estas dos cuestiones, la Secretaría de Desarrollo Social identifica la población en condición de pobreza mediante el método de comprobación de medios de vida aproximados. Dentro de este método existen distintas herramientas, tanto estadísticas como económicas, que permiten obtener un “puntaje” para la identificación de los pobres, y van desde las más sencillas, como el análisis discriminante (Wachter, 2000), análisis de componentes principales (Castaño, 2002) o modelos de regresión logit y probit (Glewwe y Kanaan, 1989), hasta algunos más sofisticados como análisis de conglomerados (Bitran *et al.*, 2005), análisis de correspondencias (Greenacre, 2005; Sutherland *et al.*, 2002) o conjuntos difusos (Makdissi, 2004). Si bien cada uno de ellos tiene sus ventajas y limitaciones y profundizar en el análisis de cada uno de éstos merece toda una línea de investigación, en este trabajo hemos elegido tres modelos: análisis discriminante, modelo logit y modelo logit multinivel, los cuales, aunque sencillos, nos permiten obtener resultados relevantes respecto a su capacidad de identificar a la población objetivo de los programas sociales en México.

En los programas que implementa Sedesol mediante comprobación de medios de vida aproximados se utiliza la herramienta de Análisis Discriminante, la cual, en el caso del programa de desarrollo humano Oportunidades, ha demostrado ser útil y eficiente en términos operativos

---

canasta de bienes. Para el caso particular que se analiza aquí se considera el nivel de pobreza de capacidades, obtenido a partir de una canasta de bienes que considera simultáneamente las necesidades de alimentación, salud y educación de la población.

<sup>2</sup> Véase Székely, *Hacia una nueva generación de política social*, “Cuadernos de Desarrollo Humano”, Secretaría de Desarrollo Social, México, 2002.

(Orozco *et al.*, 1999).<sup>3</sup> En etapas más recientes de la política social, este mecanismo de focalización se aplicó en el año 2001 para la identificación de la pobreza en manzanas, áreas geoestadísticas básicas (agebs), localidades, municipios, entidades federativas (Hernández *et al.*, 2003) y delimitación de polígonos de pobreza en las zonas urbanas en donde opera el Programa Hábitat.

Al ser esta herramienta la más utilizada en la identificación de individuos en los programas sociales que opera el gobierno federal, lo hemos incluido como nuestro modelo principal, el cual compararemos con otros dos.

Debido a la naturaleza discreta de los datos que se utilizan en este documento, un modelo Logit o Probit es más recomendable que el análisis de regresión lineal. Aunque no existe un criterio para seleccionar un probit o un logit, en este caso se ha utilizado el segundo.

El tercer modelo a comparar, Logit Multinivel, es una generalización del modelo Logit, en donde se considera que además de las variables que caracterizan los hogares, existen otras que reflejan características del territorio en donde habitan y que influyen en su condición de pobreza. Este modelo intenta responder dos preguntas: 1. ¿Estas características sólo se presentan a nivel hogar o también son explicativas en niveles superiores, digamos, regiones? Y 2. ¿Es posible que sólo algunas de estas características sean explicativas a nivel regional? Para ello busca incorporar información adicional y obtener mejores estimaciones del modelo Logit simple, considerando que las variables que explican la probabilidad de ser pobre no son únicamente aquellas que caracterizan el hogar, sino que existen patrones de correlación entre los hogares que habitan en determinados territorios, que los hacen ser más parecidos entre sí en comparación con hogares que habitan en otras regiones del país.

### III. Definición de variables y estadística descriptiva

Para todas las estimaciones se utiliza la misma fuente de datos: la Encuesta Nacional de Ingresos y Gastos de los Hogares 2002 (ENIGH).<sup>4</sup>

---

<sup>3</sup> El primer desarrollo fue concebido para la focalización a nivel de hogar del Programa de Educación, Salud y Alimentación (Progresá), en combinación con los niveles de marginación a nivel de localidades. En etapas posteriores, el mecanismo de clasificación de pobreza basado en el Análisis Discriminante ha sido aplicado a otros niveles de agregación, para identificar concentraciones de pobreza en las zonas urbanas.

<sup>4</sup> La muestra consta de 17,617 hogares y tiene representatividad a nivel nacional, en zonas rurales y urbanas.

En todos los modelos estimados se utilizan las mismas variables, con el fin de hacer comparaciones sobre la misma base de interés. Las variables son continuas y categóricas. En el caso de estas últimas, el valor de 1 indica una condición desfavorable de cada característica, y el 0 es la categoría de referencia representada por el complemento de cada variable.

*Variable dependiente:* Pobreza de capacidades. Es 1 si el hogar se clasifica en pobreza de capacidades de acuerdo con la medición de pobreza del CTMP; es 0 en otro caso. El nivel de pobreza de capacidades se define a partir de aquellos individuos que, si bien cuentan con un ingreso que les permite el acceso a una alimentación mínima adecuada, su ingreso no es suficiente para realizar una inversión aceptable en la salud y educación de los miembros del hogar. Siguiendo este concepto, el Programa Oportunidades está dirigido a la población que se encuentra por debajo de la línea de pobreza de capacidades, y es respecto a esta línea que se lleva a cabo su focalización. Para efectos de consistencia y para poder hacer comparaciones con otros estudios y con la información derivada del Programa Oportunidades, se ha tomado como variable dependiente este nivel de pobreza.

*Variables independientes:*

a. Características del hogar

- i. Estrato rural. Se toma como estrato rural aquellos hogares ubicados en localidades menores a 2,500 habitantes.
- ii. Hogar con piso de tierra.
- iii. Hogar sin excusado.
- iv. Hogar con excusado pero sin conexión de agua. (0 = no; 1 = sí)
- v. Enseres del hogar. Estufa de gas, lavadora, refrigerador, vehículo. (0 = no; 1 = sí)
- vi. Índice de hacinamiento. Variable continua que indica el número de miembros entre el número de cuartos del hogar.

## b. Características de los miembros del hogar

- i. Sexo del jefe.
- ii. Dependencia demográfica. Número de miembros menores de 15 años y mayores de 65, entre el número de miembros con edades entre 15 y 65.
- iii. Edad del jefe.
- iv. Número de niños dentro del hogar.
- v. Seguridad social. Se considera un hogar con seguridad social aquel en el que al menos uno de los miembros cuenta con esta prestación.
- vi. Escolaridad del jefe. Se incluyen dos variables de escolaridad: sin instrucción y con primaria incompleta.

## c. Características de región

- i. Regiones. Son 14 regiones que se aplicaron a la muestra de la ENIGH, construidas con base en la experiencia operativa del Programa Oportunidades y la información registrada mediante Sistemas de Información Geográfica. Las variables regionales representan la proporción (o la media) de las variables a nivel hogar.

### *III.1. Estadística descriptiva*

El anexo A4 muestra las características socioeconómicas de los hogares que se encuentran en condiciones de pobreza de capacidades, en comparación con el perfil de los hogares a nivel nacional. A manera de ejemplo, mientras que 24% de los hogares a nivel nacional se ubican en el contexto rural, 50% de los hogares en pobreza de capacidades se encuentra en este estrato, y 30% de los hogares en pobreza tienen piso de tierra, una característica tres veces más frecuente en comparación con sólo 10% a nivel nacional. Existe una gran diferencia en el ingreso neto total per cápita promedio entre los hogares clasificados como pobres, respecto al promedio nacional, que es cinco veces mayor.

Como puede apreciarse, existe una diferencia clara entre las características de los hogares, dependiendo de su condición de pobreza. Es precisamente esta característica de la información la que permite aproximar la medición de la pobreza mediante métodos estadísticos.

### III.2. Descripción de los modelos

#### III.2.1. Modelo de Análisis Discriminante

Siguiendo a Fisher (1937), a partir de la estimación del análisis discriminante se obtiene una *función discriminante* y las correlaciones de las variables explicativas para las poblaciones pobres y no pobres. Dicha función “resume” las características del hogar, expresadas a partir de muchas variables, en una sola variable continua, denominada calificación discriminante. Esta variable es un índice que ordena los hogares de acuerdo con su nivel de pobreza. La combinación lineal a partir de la cual se estima la función discriminante tiene la siguiente forma funcional:

$$y = \lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k$$

donde  $y$  es la variable dependiente,  $x = (x_1, x_2, \dots, x_k)$  es el vector de características del hogar,  $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)$  son coeficientes tales que la razón de la diferencia al cuadrado entre  $\bar{y}_1$  y  $\bar{y}_2$  con la varianza de  $Y$  es máxima, es decir se maximiza la función:

$$\phi = \frac{[\lambda'(\bar{y}_1 - \bar{y}_2)]^2}{\lambda' \Sigma \lambda},$$

donde  $\mu_1$  y  $\mu_2$  es la media de  $X$  en los dos grupos,  $\Sigma_1$  y  $\Sigma_2$  las matrices de varianzas y covarianzas y  $\Sigma_1 = \Sigma_2 = \Sigma$ .

El producto de los coeficientes no estandarizados por las observaciones (o sea, la multiplicación de estos dos elementos) da como resultado el puntaje del discriminante  $y$ , a partir de un punto de corte, la clasificación de pobre y no pobre.

Entre las pruebas de significancia y de verificación de bondad de ajuste se encuentra el estadístico  $\Lambda$  de Wilks, que varía de 0 a 1, donde 0 significa que las medias entre grupos de esa variable son diferentes (*i. e.* que esa variable explica más la diferencia entre grupos)

y 1 que las medias entre ambos grupos son la misma. La prueba F de las lamdas muestra si la contribución de las variables es significativa y la correlación canónica es la correlación de la función con la calificación discriminante.

### III.2.2. Modelo de Regresión Logit

La forma funcional del modelo logit es:

$$y_i^* = \beta_1 + \sum_2^k \beta_i^k x_i^k + \varepsilon_i$$

donde  $x = (x_1, x_2, \dots, x_k)$  es el vector de características del hogar,  $y_i^*$  es la variable dependiente que no es observable; lo que se observa es una variable dicotómica definida por:  $Y_i = 1$  si  $Y_i^* > 0$   $Y_i = 0$  e. o. c y lo que se estima es la probabilidad de ocurrencia de la variable dependiente:

$$\begin{aligned} \text{Prob}(Y_i = 1) &= \text{Prob}(Y_i^* > 0) = \text{Prob}(\varepsilon_i > -\beta X_i) \\ \text{Prob}(Y_i = 1) &= 1 - F(-\beta X_i), \end{aligned}$$

donde F es la función de distribución acumulativa de  $\varepsilon$ , siendo  $\varepsilon_i \sim N(0, \sigma^2)$ , que al ser logística tiene la siguiente forma:

$$F(-\beta X_i) = \pi_i = \frac{\exp(-\beta X_i)}{1 + \exp(-\beta X_i)} = \frac{1}{1 + \exp(\beta X_i)}$$

La clasificación de pobre o no pobre se obtiene a partir de la probabilidad estimada  $\hat{p} = \frac{\exp(\beta_i)}{1 + \exp(\beta_i)}$  y un punto de corte de esta probabilidad.

Las medidas más recomendadas para verificar la bondad de ajuste de los modelos Logit y Probit pueden ser el valor de la función de verosimilitud en el máximo valor; el estadístico L que surge de una prueba de homoscedasticidad con distribución  $X^2(p)g.l.$  (donde p es el número de variables explicativas dentro del modelo). Si este estadístico es mayor que el valor crítico no rechazamos  $H_0$ , y este modelo es homoscedástico. El coeficiente de determinación pseudo  $R^2$  es también una medida de bondad de ajuste que toma en cuenta la restricción del rango de las variables cuantitativas.

### III.2.3. Modelo de regresión Logit Multinivel

El modelo Multinivel muestra una técnica para la estimación de los coeficientes del modelo Logit y su nivel de significancia, cuando existen variables a nivel regional que reflejan un posible anidamiento en los datos. A estas variables se les denomina comúnmente variables de segundo nivel. A las variables del hogar se les llama variables de primer nivel, y pueden variar de un hogar a otro incluso si ambos hogares se ubican en el mismo territorio. Es importante decir que si bien la ENIGH no es representativa de las 14 regiones que aquí se utilizan, pues no podríamos calcular estadísticas descriptivas con representatividad regional o estimar regresiones para cada región, los coeficientes de las variables que representan las regiones resultaron significativos en el modelo Logit, lo cual brinda cierta evidencia de las diferencias que existen entre distintas zonas del país. La función del modelo Logit multinivel estimado es de la siguiente forma:

Sea  $i = (1, \dots, n)$  el número de individuos de primer nivel (microunidades) anidados con  $j = (1, \dots, N)$  grupos de segundo nivel (macrounidades), la relación de primer nivel es:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij} \text{ donde } r_i \sim N(0, \sigma^2)$$

Las relaciones de segundo nivel son:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}W_j + u_{1j}$$

El modelo en forma combinada es:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}X_iW_j + u_{0j} + u_{1j}X_{ij} + r_{ij}$$

donde se asume que:

$$E \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \text{Var} \begin{bmatrix} u_{0j} \\ u_{1j} \end{bmatrix} = \begin{bmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{bmatrix} = T$$

$$Cov(u_{0j}, r_{ij}) = Cov(u_{1j}, r_{ij}) = 0$$

$B_{0j}$  y  $\beta_{1j}$  son los coeficientes de nivel 1. Éstos no son conocidos pero pueden ser estimados, y para este caso tienen un componente aleatorio de la forma:

$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$

$\gamma_{00}, \dots, \gamma_{11}$  son coeficientes de nivel 2 con efectos fijos.

$X_{ij}$  variables explicativas de nivel 1.

$W_{ij}$  variables explicativas de nivel 2.

$r_{ij}$  es el término de error de nivel 1.

$u_{0j}, u_{1j}$  son términos de error de nivel 2.

$\sigma^2$  es la varianza de nivel 1.

$\tau_{00}, \tau_{01}, \tau_{11}$  son componentes de varianzas y covarianzas de nivel 2.

De la definición anterior surgen dos modelos:<sup>5</sup>

- i. Un modelo de un nivel, donde las observaciones están a nivel de hogar y cuya ordenada al origen tiene un componente aleatorio.
- ii. Un modelo de dos niveles, a nivel de hogar y a nivel de región, con efectos aleatorios en la ordenada al origen y en dos variables de segundo nivel (piso de tierra y estufa de gas).

En este modelo la clasificación de pobre o no pobre también se obtiene de una función de probabilidad. Entre sus medidas de especificación y de bondad de ajuste se pueden mencionar el estadístico  $t$ , para probar la hipótesis de que un parámetro sea igual a cero ( $H_0: \gamma h = 0$ ); un coeficiente de determinación  $R^2$ , que combina las varianzas del modelo en ambos niveles; el coeficiente de correlación entre clases (ICC), que indica la proporción de la varianza que es explicada por cada nivel; la devianza, definida como  $d = -2 \ln(\text{valor de verosimilitud})$ , y aunque en muchos modelos no se pueden interpretar sus valores directamente, se comparan sus diferencias entre varios modelos con el mismo conjunto de datos.

### III.3. Resultados comparativos

En esta sección se presentan los resultados de cada uno de los modelos, haciendo un análisis comparativo basado en tres aspectos: significancia

---

<sup>5</sup> En los modelos no lineales de análisis multinivel no es posible utilizar factores de expansión con el software HLM. Para la estimación se expandió la base de datos a nivel poblacional, y se creó a partir de ella una muestra aleatoria autoponderada.

de las variables, capacidad de clasificación a partir de la medición de ingreso del CTMP y, el más relevante, errores de inclusión y exclusión. A partir de esta comparación, lo que se busca es identificar las ventajas y limitaciones de cada uno de los modelos de acuerdo con estos tres criterios, y justificar su utilización en la identificación de los beneficiarios de los programas sociales en México.

Si bien las medidas de bondad de ajuste también son relevantes para llevar a cabo una comparación, existen diferencias en los supuestos de los modelos que impiden comparar estas medidas. En cada modelo se eligió la especificación que de acuerdo con sus propios supuestos resultó tener un mejor ajuste. El análisis sólo se lleva a cabo a partir de los resultados relevantes para su comparación, pero en los anexos A5-A7 se encuentran los resultados de los modelos de manera individual. En el caso del modelo Multinivel se presentan dos modelos, sin embargo todo el análisis está hecho con base en el modelo ii, pues éste resultó el mejor en términos de ajuste.<sup>6</sup>

Hay que recordar que además de los efectos entre hogares y entre regiones, también existe un efecto aleatorio particular al hogar y a la región no explicado en el modelo, que aumenta (o disminuye) la probabilidad de hallarse en condición de pobreza.

### III.3.1. Significancia e impacto en el modelo

Al igual que en las medidas de ajuste, en cada modelo los coeficientes obtenidos se interpretan de distinta manera, sin embargo podemos hacer comparaciones de acuerdo con la relevancia (significancia) de estas variables para explicar el modelo y el impacto que éstas tienen sobre la clasificación de pobreza, es decir, podemos identificar qué variables tienen un efecto positivo (menos pobre, indicado con el signo +), un efecto negativo (más pobre, indicado con un signo -), o cuáles variables no son significativas (indicado con un 0) sobre la calificación o la probabilidad de ser pobre. El cuadro del anexo A3 muestra esta comparación.

---

<sup>6</sup> Para el segundo nivel las variables se eligieron a partir de un análisis exploratorio sobre las correlaciones de los residuales empíricos bayesianos (Empirical Bayes residuals), de acuerdo con Bryk & Raudenbusch (2002) estimados en el modelo i, más las variables de segundo nivel que podrían incluirse en el modelo. Con la intención de simplificar este documento, se han omitido los resultados del análisis exploratorio y sólo se presentan los resultados de la estimación final de los modelos propuestos. Sin embargo, los resultados de éste se pueden solicitar a los autores.

Como se aprecia, las únicas diferencias que existen son en términos de la significancia, pues las variables significativas en los tres modelos tienen el mismo impacto, ya sea positivo o negativo, sobre la variable explicativa. En el modelo de análisis discriminante todas las variables son significativas, excepto el estrato rural y las de educación, siendo las variables de región las que contribuyen a una mejor calificación discriminante. Para el caso del modelo Logit, las variables de sexo y edad del jefe son las únicas no significativas, y al igual que en el modelo discriminante, las variables de región contribuyen a reducir la probabilidad de ser un hogar en pobreza de capacidades. A partir de los resultados de este modelo podemos determinar que, dadas el resto de sus características, los hogares que se ubican en la región 4 tienen una probabilidad de 14% de ser pobres, que corresponde a hogares de los estados de Coahuila, Nuevo León y Tamaulipas.<sup>7</sup> En cambio los hogares de la región 8, conformada por hogares de algunas regiones de Hidalgo, Puebla, San Luis Potosí o Veracruz, tienen una probabilidad de clasificarse en condiciones de pobreza de 34%.

En el modelo Logit Multinivel las variables no significativas son la no instrucción del jefe de hogar y estufa de gas, ya que su significancia es absorbida por las variables de región. A diferencia de los dos modelos anteriores, en este caso el sexo del jefe resultó significativo y con un impacto positivo. De esta manera, si el hogar tiene piso de tierra o no tiene estufa de gas la probabilidad de encontrarse en pobreza aumenta, pero es menor cuando el hogar habita en una región con elevada proporción de jefas de hogar (mujeres).<sup>8</sup>

Dado que no existen diferencias sustanciales en términos de la significancia o el impacto de las variables entre un modelo y otro, no podríamos establecer una preferencia por uno u otro. Sin embargo, es importante mencionar que una ventaja de los modelos multinivel es que permiten ir más allá del análisis de las variables a nivel del hogar, en combinación con variables medidas en otros niveles (en este caso regiones), así como

---

<sup>7</sup> Véase anexo A1 para una descripción de los estados que pertenecen a cada región.

<sup>8</sup> Encontrar una especificación adecuada tanto en primer como en segundo nivel es una tarea que se realiza probando diferentes hipótesis y considerando diferentes especificaciones. En este trabajo se presentan los resultados finales de los numerosos modelos que se probaron. En el anexo se incluyen los dos modelos principales, pero dado que el modelo ii resultó con mejor ajuste, todos los análisis de este documento se llevaron a cabo con base en este modelo.

explorar el efecto que esta combinación tiene sobre la condición de pobreza. Si bien las características del hogar son importantes en el modelo, incluir variables agregadas a nivel de región demuestra que, efectivamente, existen variaciones en las condiciones de pobreza de los hogares, que son explicadas por variaciones entre regiones. No considerarlas provocaría un sesgo en los estimadores y su nivel de significancia, pues en algunos casos las variaciones de segundo o tercer nivel son más importantes incluso que a nivel del hogar.

El modelo Logit Multinivel aporta información adicional sobre comportamientos regionales; sin embargo, su estimación e interpretación son considerablemente más complejas que en el caso del Análisis Discriminante y el modelo Logit simple.

### III.3.2. Porcentaje de clasificación coincidente con el CTMP

Como un vistazo a la capacidad de los modelos para identificar la población en condiciones de pobreza según sus características socioeconómicas, en el siguiente cuadro se presenta el porcentaje de casos coincidentes de cada modelo con las medidas de pobreza, de acuerdo con el ingreso del CTMP.

**Cuadro 1.** Casos coincidentes con el método de ingreso del CTMP (%)

| <i>Análisis Discriminante</i> | <i>Logit</i> | <i>Logit Multinivel</i> |
|-------------------------------|--------------|-------------------------|
| 84.20%                        | 86.80%       | 86.30%                  |

*Fuente:* Cálculos propios con base en la ENIGH 2002.

La capacidad de clasificación correcta del modelo Logit respecto de la clasificación de pobreza del CTMP es 87%, mientras que la del Logit Multinivel es de 86%, en comparación con 84% del Análisis Discriminante. A partir de este criterio se puede decir que el modelo Logit coincide en un porcentaje ligeramente mayor que los otros dos modelos, acercándose más a la medida de ingreso del CTMP. Esto significa que si se ignorara el ingreso de los hogares entrevistados en la ENIGH y se quisiera identificar su condición de pobreza utilizando el modelo Logit, 4 de cada 5 hogares serían clasificados en la misma categoría de ingreso que el original (pobre de capacidades o no pobre), de acuerdo con sus características socioeconómicas.

### III.3.3. Errores de exclusión e inclusión (subcobertura y fuga)

En general, para cuantificar la precisión de focalización sobre cualquier método se utilizan dos tipos de medidas: el error de inclusión o tasa de fuga (denotado por L) y el error de exclusión o tasa de subcobertura (denotado por U). El error de inclusión (L) se define como el número de hogares pobres de acuerdo con cada modelo y que son no pobres de acuerdo con la metodología del CTMP, divididos entre el número total de hogares no pobres del CTMP. El error de exclusión (U) son los hogares clasificados como no pobres de acuerdo con cada modelo y pobres según la metodología del CTMP, divididos entre el número total de hogares pobres del CTMO. En primer lugar, se comparan las tasas de fuga (error de inclusión) y subcobertura (error de exclusión) de cada modelo, es decir, el porcentaje de casos que son clasificados en la misma categoría (pobre o no pobre) que la clasificación del CTMP.<sup>9</sup>

El siguiente cuadro muestra los resultados de los errores de inclusión y exclusión de cada modelo. La tasa de fuga (L) correspondiente al Análisis Discriminante es de 40 por ciento. Esto significa que 4 de cada 10 hogares clasificados como pobres por el Análisis Discriminante, de acuerdo con sus características socioeconómicas, son clasificados como no pobres por la metodología del CTMP. En los modelos Logit y Logit Multinivel las tasas de fuga son de 34%, porcentaje menor al del Discriminante.

La tasa de subcobertura (U), según el modelo de Análisis Discriminante, es de 27%, siendo ésta la menor de todos los modelos, mientras que tanto el modelo Logit como el modelo Multinivel tienen una tasa de subcobertura de 33%. Estas tasas de fuga reflejan que existen hogares por encima de la línea de pobreza de ingreso, cuyas características socioeconómicas son tan similares a las de los hogares que se encuentran por debajo de la línea de pobreza, que podrían considerarse como hogares pobres.

---

<sup>9</sup> La medición del CTPM es la que consideramos estándar para medir la pobreza. En descargo de las diferencias que se muestran en este documento, en donde se verifica mejor la aproximación de las técnicas estadísticas respecto del perfil socioeconómico en comparación con la metodología del CTPM, debe decirse que ésta no fue diseñada como herramienta de focalización; su principal objetivo es medir la pobreza. Hay que mencionar que en el caso del CTPM la Sedesol considera otros niveles de pobreza (pobreza patrimonial) más elevados, cuya adopción podría disminuir sustantivamente las tasas de subcobertura a costa de incrementar también sustantivamente las tasas de fuga para la población más pobre. Sin embargo, el objetivo en este trabajo se centró únicamente en la pobreza de capacidades establecida en la Sedesol.

**Cuadro 2.** Errores de inclusión (L) y exclusión (U) de los modelos estimados

| A.            |          | CTMP              |                |                |
|---------------|----------|-------------------|----------------|----------------|
| Discriminante |          | <i>No pobre %</i> | <i>Pobre %</i> | <i>Total %</i> |
|               | No pobre |                   | U = 27.0       | 74.60          |
|               | Pobre    | L = 40.1          |                | 25.40          |
|               | Total    | 79.20             | 20.80          | 100.00         |

  

| B.    |          | CTMP              |                |                |
|-------|----------|-------------------|----------------|----------------|
| Logit |          | <i>No pobre %</i> | <i>Pobre %</i> | <i>Total %</i> |
|       | No pobre |                   | U = 32.9       | 78.70          |
|       | Pobre    | L = 34.4          |                | 21.30          |
|       | Total    | 79.20             | 20.80          | 100.00         |

  

| C.  |          | CTMP              |                |                |
|-----|----------|-------------------|----------------|----------------|
| HLM |          | <i>No pobre %</i> | <i>Pobre %</i> | <i>Total %</i> |
|     | No pobre |                   | U = 33.3       | 78.80          |
|     | Pobre    | L = 34.3          |                | 21.20          |
|     | Total    | 79.20             | 20.80          | 100.00         |

Fuente: Cálculos propios con base en la ENIGH 2002.

Las tasas de subcobertura y fuga indican resultados con base en el número de hogares pobres y no pobres de cada modelo y su porcentaje de clasificación respecto del criterio de referencia, pero no aportan mayor información sobre qué tan alejados de la línea de pobreza están estos hogares que se incluyen o excluyen. Por lo tanto, resulta conveniente contar con información sobre las tasas de fuga y subcobertura de la brecha, y la severidad de la pobreza. Para ello se consideran las medidas relativas provenientes de los índices Foster-Greer-Thorbecke (FGT). La tasa de subcobertura (error de exclusión) se calcula por medio de la siguiente fórmula:

$$U(\alpha) = \left( \frac{1}{N_{PC}} \right) \sum_{i=1}^q \left( \frac{z - i_i}{z} \right)^\alpha,$$

donde  $N_{PC}$  es el total de hogares clasificados como pobres de acuerdo con el criterio de ingreso;  $q$  es el número de hogares clasificados como no pobres según el modelo de comparación, y como pobres según el ingreso;  $z$  es la línea de pobreza;  $i_i$  es el ingreso corriente per cápita del  $i$ -ésimo hogar; y  $\alpha$  es el peso de la severidad de la pobreza.

La tasa de fuga (error de inclusión) se define como:

$$L(\alpha) = \left( \frac{1}{N_{ME}} \right) \sum_{i=1}^q \left( \frac{i_i - z}{z} \right)^\alpha,$$

donde  $N_{ME}$  son los hogares identificados como pobres por el correspondiente modelo estimado, y  $q$  es el número total de hogares no pobres según el criterio de ingreso, y pobres según el modelo a comparar.

El siguiente cuadro muestra los resultados de aplicar estas fórmulas a cada uno de los modelos. En cada cuadro hay tres columnas: el índice (0) corresponde al porcentaje de clasificación en pobreza de capacidades, y los índices (1) y (2) a las medidas de brecha (o profundidad) y severidad de la pobreza, respectivamente.

**Cuadro 3.** Tasas de subcobertura y fuga con el esquema de ponderación FGT

| <i>Cambios porcentuales con respecto al modelo Logit*</i> |                     |              |              |               |              |              |              |
|---|---------------------|--------------|--------------|---------------|--------------|--------------|--------------|
| <i>Modelo</i>   | <i>Subcobertura</i> |              |              | <i>Modelo</i> | <i>Fuga</i>  |              |              |
|   | <i>U(0)</i>         | <i>U(1)</i>  | <i>U(2)</i>  |               | <i>L(0)</i>  | <i>L(1)</i>  | <i>L(2)</i>  |
| Logit   | 0.329               | 0.094        | 0.08         | Logit         | 0.344        | 0.195        | 0.723        |
| HLM   | 0.333               | 0.097        | 0.081        | HLM           | 0.343        | 0.189        | 0.682        |
|   | <i>1.22</i>         | <i>3.19</i>  | <i>1.25</i>  |               | <i>-0.29</i> | <i>-3.08</i> | <i>-5.67</i> |
| Análisis  | 0.27                | 0.076        | 0.072        | Análisis      | 0.401        | 0.265        | 0.811        |
| Discriminante   | <i>-17.9</i>        | <i>-19.1</i> | <i>-10.0</i> | Discriminante | <i>16.6</i>  | <i>35.9</i>  | <i>12.2</i>  |

*Fuente:* Cálculos propios con base en la ENIGH 2002. \* Los números en cursiva representan la reducción (-) o incremento (+) en las tasas de subcobertura y fuga del modelo Logit Multinivel y el Análisis Discriminante respecto del modelo Logit.

Del lado izquierdo de la tabla, la columna U(0) presenta los mismos datos de subcobertura del cuadro anterior, en los cuales el análisis discriminante se ubica con un error de exclusión 18% por debajo del modelo Logit, y menor que el HLM. En las columnas U(1) y U(2) los menores errores de exclusión en términos de profundidad y severidad de la pobreza también pertenecen al Análisis Discriminante. Esto significa que aquellos hogares que excluye por error (vistos de menor a mayor ingreso) están más cerca de la línea de pobreza, es decir, son menos pobres, y con una brecha entre ellos menor que en los modelos Logit y Logit Multinivel. En términos relativos, la tasa de subcobertura de la profundidad de pobreza en el análisis discriminante es 19% menor a la del modelo Logit, y mucho menor que en el caso del Logit Multinivel, dado que este último presenta una tasa incluso 1.2% mayor en comparación con el Logit simple, mientras que en el caso de U(2) las brechas relativas entre los distintos modelos son menores.

En el lado derecho la columna L(0) muestra las tasas de fuga de los tres modelos. Se puede apreciar que mientras el modelo Multinivel está .3% por debajo del Logit, el modelo de Análisis Discriminante se encuentra con un error de inclusión de 16%. Sin embargo, respecto a la brecha L(1) y severidad L(2), si bien el análisis discriminante también tiene los mayores errores de inclusión (36% y 12% más que el modelo Logit respectivamente), vale decir a su favor que este modelo considera como pobres a aquellos que tienen una mayor profundidad y severidad de pobreza, es decir, aquellos que están más alejados de la línea de pobreza.

Esto podría dar muestras de que el Análisis Discriminante es un modelo más conservador, debido a que el rango de los criterios que aplica para evitar la exclusión de hogares que podrían considerarse como pobres de acuerdo con sus características socioeconómicas es más amplio que en el resto de los modelos presentados y que, por supuesto, del criterio de ingreso del CTMP.

Es importante tomar en cuenta que en la focalización de un programa social, además del costo monetario que implican los errores de inclusión y exclusión, existen otros criterios de justicia social que en algunas ocasiones no son cuantificables. No siempre es posible medir el costo económico que representa excluir a un hogar en condiciones de pobreza de los beneficios de un programa social, pues la falta de apoyos puede repercutir en varias esferas de la vida de las personas en el corto, mediano y largo plazos. Por ejemplo, en el desarrollo de sus capacidades, sus posibilidades de inserción en el mercado laboral, y la capacidad de contar con elemen-

tos para contribuir a romper el círculo intergeneracional de la pobreza, entre otros. En este sentido, podría optarse por un modelo conservador, con menores errores de subcobertura a costa de un mayor error de fuga. Bajo este criterio es que se ha preferido utilizar el Análisis Discriminante para la focalización de los programas sociales en México, pues identifica un mayor número de hogares con perfiles de elevadas carencias, aun cuando su ingreso pueda estar por encima de la línea de pobreza de capacidades.<sup>10</sup>

#### **IV. Conclusiones**

La intención de utilizar metodologías estadísticas tiene el objetivo de incorporar una medición más completa de la pobreza, aplicable a un concepto de focalización en donde se evita al máximo la interpretación subjetiva de quien realiza el análisis, respecto de los “pesos” o “importancia” que cada variable tiene dentro de la regla de decisión.

Según los resultados que se muestran en este documento, los tres modelos presentados cumplen con este objetivo, al ser especificaciones con un ajuste adecuado de acuerdo con sus supuestos, y sin grandes diferencias entre ellos en términos de significancia e impacto de las variables explicativas. Por lo tanto, la única diferencia entre estos tres modelos la hacen los criterios de errores de inclusión y exclusión. Como lo muestran los resultados obtenidos en este trabajo, el modelo de Análisis Discriminante, método utilizado por la Sedesol para identificar la población en condiciones de pobreza, tiene los errores de exclusión más bajos de acuerdo con la línea de pobreza determinada y de acuerdo con su profundidad y severidad.

Sin embargo, es crucial tomar en cuenta otros criterios al momento de elegir sobre uno u otro modelo. En términos teóricos existen múltiples referencias sobre la conveniencia de utilizar el modelo Logit en comparación con el Análisis Discriminante, cuando los supuestos de normalidad y/o identidad de matrices de covarianza no se cumplen. Sin embargo, autores como Lachenbruch (1975) o Klecka (1980) indican que el Análisis Discriminante es una técnica robusta que puede tolerar desviaciones de los supuestos. Documentos de años más recientes como Buja (1994), Hastie

---

<sup>10</sup> Éste es el caso de la focalización de los programas Oportunidades y Hábitat, en los cuales las decisiones sobre la utilización del Análisis Discriminante fue tomada dando preferencia a un criterio de justicia para los hogares, con base en su perfil socioeconómico.

(1994), James (2001) o Roth (1999) demuestran la conveniencia de utilizar el Análisis Discriminante bajo la premisa de que es una técnica de clasificación, a diferencia del Logit, que se utiliza preferentemente cuando se quieren establecer relaciones entre variables. En la práctica, si bien muchos de esos estudios se centran en sistemas de reconocimiento de voz y datos, así como en análisis y predicciones de riesgo financiero o decisiones de quiebra de empresas, su utilización en el ámbito de las ciencias sociales no es tan común, por lo que hay una posibilidad de desarrollo de la técnica, pues empíricamente resulta viable incorporando las recientes aportaciones de textos avanzados de estadística.

Otro aspecto que se debe considerar es el comportamiento de los modelos por separado para comunidades rurales y urbanas. En algunos países se ha demostrado que los resultados en términos de identificación de individuos en condiciones de pobreza para dos poblaciones por separado son tan similares como si se utilizara un solo modelo en su conjunto (Grosh, 1994). En el caso de México, si bien la metodología considera las localidades rurales y urbanas, los puntajes se utilizan de manera homogénea en todas las regiones (Oportunidades, 2000), y es por eso que en este estudio se utiliza un modelo general. Sin embargo, a medida que se encuentren mejores metodologías de focalización no se descarta la posibilidad de utilizar modelos distintos para cada región. Finalmente, en la evaluación de la eficiencia de cada método desarrollado es importante tomar en cuenta otros aspectos que son relevantes en el diseño de políticas públicas para la reducción de la pobreza. Como ejemplo tenemos los costos administrativos, que pueden variar incluso de acuerdo con el entorno político de cada país (Van de Walle, 1995), o los incentivos, ya sea positivos o negativos, de un programa focalizado, en particular de aquellos que consisten en transferencias monetarias. Sin duda, desarrollar posteriores investigaciones y superar las limitaciones de estudios anteriores implica varios retos. Cualquier esfuerzo en esta dirección será muy fructífero para la política social.

### Referencias bibliográficas

- Abul Naga, R. (2003), "The Allocation of Benefits under Uncertainty: A Decision-Theoretic Framework", *Economic Modelling*, 20, no. 4, pp. 873-893.
- Akerlof, G. A. (1978), "The Economics of 'Tagging' as Applied to the Optimal

- Income Tax, Welfare Programs, and Manpower Planning”, *The American Economic Review* 68, no. 1, pp. 8-19.
- Besley, T. J., S. M. Ravi Kanbur (1988), “Food Subsidies and Poverty Alleviation”, *Economic Journal* 98, no. 392, pp. 701-719.
- Bitran R. *et al.* (2005), “Keeping Healthy in an Urban Environment: Public Health Challenges for the Urban Poor”, en *The Urban Poor in Latin America*, Washington, World Bank.
- Bryk A y S. Raudenbush (2002), “Hierarchical Linear Models, Applications and Data Analysis Methods”, *Advanced Quantitative Techniques in the Social Sciences Series*, no. 1, Sage, 2da edición.
- Buja, A., T. Hastie y R. Tibshirani (1994), “Penalized Discriminant Analysis”, *Annals of Statistics*, vol. 23, pp. 73-102.
- Castaño, Elkin (2002), “Proxy Means Test Index for Targeting Social Programs: Two Methodologies and Empirical Evidence”, *Lecturas de Economía*, 56, pp. 133-144.
- Coady, David P. (2006), “The Welfare Returns to Finer Targeting: The Case of the Progresa Program in Mexico”, *International Tax and Public Finance* 13, no. 2-3, pp. 217-239.
- Coady, D. (2000), “The Application of Social Cost-Benefit Analysis to the Evaluation of Progresa”, *International Food Policy Research Institute*.
- Coady, D., M. Grosh (2004), “Targeting of transfers in developing countries: Review of lessons and experience”, *Regional and Sectoral Studies*, Washington, World Bank.
- Coady, D. y E. Skoufias (2001), “On the Targeting and Redistributive Efficiencies of Alternative Transfer Instruments”, *International Food Policy Research Institute*.
- Cortés, F. *et al.* (2002), “Evolución y características de la pobreza en México en la última década del siglo XX”, *Serie: Documentos de investigación II*, México, Sedesol.
- Duclos, Jean-Yves, Makdissi, Paul y Quentin Wodon (2005), “Poverty-Dominant Program Reforms: The Role of Targeting and Allocation Rules”, *Journal of Development Economics* 77, no. 1, pp. 53-73.
- Ebert, Udo (2005), “Optimal Anti-Poverty Programmes: Horizontal Equity and the Paradox of Targeting”, *Economica* 72, no. 287, pp. 453-468.
- Elbers, Chris *et al.* (2007), “Poverty Alleviation Through Geographic Targeting: How Much Does Disaggregation Help?” *Journal of Development Economics* 83, no. 1, pp. 198-213.
- Fisher, R.A. (1937), *The Design of Experiments*, 2ª ed., Londres, Oliver and Boyd.

- Gelbach, Jonath B. y Lant Pritchett (1997), *More for the Poor Is Less for the Poor: The Politics of Targeting*, Washington, World Bank.
- Glewwe, Paul y Jacques van der Gaag (1990), "Identifying the Poor in Developing Countries: Do Different Definitions Matter?" *World Development* 18, no. 6, pp. 803-814.
- Glewwe, P. y O. Kanaan (1989), "Targeting Assistance to the Poor: A Multivariate Approach Using Household Survey Data. *Policy, Planning and Research Working Paper 225*, Washington, World Bank.
- González de la Rocha, Mercedes (2003), "De los 'recursos de la pobreza' a la 'pobreza de recursos' y a las desventajas acumuladas", *Latin American Research Review*, 39, no. 1, pp. 192-195.
- Greenacre, M. y R. Pardo (2005), "Multiple Correspondence Analysis of a Subset of Response", *Economics Working Papers*, Department of Economics and Business, Universitat Pompeu Fabra.
- Grosh, M. (1994), *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*, Washington, World Bank.
- Gutiérrez J. P., S. Bertozzi y P. Gertler (2003), "Evaluación de la identificación de familias beneficiarias en el medio urbano", en *Evaluación de resultados de impacto del Programa de Desarrollo Humano Oportunidades*, México, Instituto Nacional de Salud Pública.
- Hastie, T. y R. Tibshirani (1994), "Discriminant Analysis by Gaussian Mixtures", *Journal of the Royal Statistical Society*, vol. 176, no. 16, pp. 58-155.
- Hernández, D. et al. (2003), "Concentración de hogares en condición de pobreza en el medio urbano", *Cuadernos de Desarrollo Humano*, México, Secretaría de Desarrollo Social.
- Herrera Ramos J. (2000), *Instituciones, focalización y combate a la pobreza*, México, Facultad Latinoamericana de Ciencias Sociales.
- James, G. M. y T. Hastie (2001), "Functional Linear Discriminant Analysis for Irregularly Sampled Curves", *Journal of the Royal Statistical Society*, vol. 63, no. 3, pp. 533-550.
- Kanbur, Ravi y Timothy Besley (1990), "The Principles of Targeting", Washington, World Bank.
- Keen, Michael (1992), "Needs and Targeting", *Economic Journal*, 102, no. 410, pp. 67-79.
- Klecka, W. R. (1980), "Discriminant Analysis", *Quantitative Applications in the Social Sciences* 19, Beverly Hills, SAGE University Paper.
- Kremer, Michael (1997), *Tax Incentives for Youth Employment*, Washington, World Bank.

- Lachenbruch, P.A. (1975), *Discriminant Analysis*, Nueva York, Hafner Press.
- Makdissi, Paul y Quentin Wodon (2004), "Fuzzy Targeting Indices and Orderings", *Bulletin of Economic Research* 56, no. 1, pp. 41-51.
- Mateus A. (1983), "Targeting Food Subsidies for the Needy: the Use of Cost-benefit Analysis and Institutional Design", *Staff Working Paper* 617, World Bank.
- Oportunidades (2000), *Diseño-ejecución del Programa Oportunidades*, México.
- Orozco, M., C. Hubert (2005), "La focalización en el Programa Oportunidades de México", *Unidad de la Protección Social, Red de Desarrollo Humano, Serie de Informes sobre Redes de Protección Social*, World Bank.
- Orozco, M., J. Gómez de León y D. Hernández (1999), "La identificación de los beneficiarios de Progresá", en *Más oportunidades para las familias pobres. Evaluación de Resultados del Programa de Educación, Salud y Alimentación*, México, Secretaría de Desarrollo Social.
- Ravallion, Martin, y Kalvin Chao (1989), "Targeted Policies for Poverty Alleviation under Imperfect information: Algorithms and Applications", *Journal of Policy Modeling* 11, no. 2, pp. 213-224.
- Ravallion, Martin (1999), "Is More Targeting Consistent with Less Spending?" *International Tax and Public Finance* 6, no. 3, 411 pp. 453-468.
- Reutlinger, Schlomo (1976), *Desnutrición y pobreza: magnitudes y opciones de política*, Tecnos.
- Roth, V., V. Steinhage (1999), "Nonlinear Discriminant Analysis Using Kernel Functions", *Department of Computer Science III, Technical Report IAI-TR-99-7*, Bonn, Bonn University.
- Santibáñez, J. et al. (2005), *Evaluación externa del programa Hábitat 2003-2004*, Tijuana, El Colegio de la Frontera Norte.
- Scott, J. (2000), *Descentralización, focalización y pobreza en México*, México, Miguel Angel Porrúa.
- Sedesol (2006), *Manual ciudadano*, México.
- Sen, Amartya (1995), "The Political Economy of Targeting", en *Public Spending and the Poor: Theory and Evidence*, 11-24. Baltimore y Londres.
- (2003), "Development as Capability", en *Readings in Human Development: Concepts, Measures and Policies for a Development Paradigm*, 3-16, Nueva Delhi.
- Sheshinski, E. y P. Diamond (1992), "Economic Aspects of Optimal Disa-

- bility Benefits”, *Working papers from Massachusetts Institute of Technology*, Cambridge, Department of Economics (MIT).
- Skoufias E., B. Davis, J. Behrman (2000), “Evaluación de la selección de hogares beneficiarios en el (Progres) Programa de Educación, Salud, y Alimentación”, en *Más Oportunidades para las Familias pobres*, México, SEDESOL.
- (2006), “Conditional Cash Transfers, Adult Work Incentives and Poverty”, *Impact evaluation series*, no. 5, World Bank.
- Sutherland, H., R. Taylor, J. Gomulka (2002), “Combining Household Income and Expenditure Data in Policy Simulations”, *Review of Income and Wealth* 48, no. 4, pp. 517-536.
- “Targeting Outcomes Redux.” *World Bank Research Observer* 19, no. 1, pp. 61-85.
- Thorbecke, E. (2003), “Poverty Analysis and Measurement within a General Equilibrium Framework”, en *Reducing poverty in Asia: Emerging Issues in Growth, Targeting, and Measurement*, 45-78. Cheltenham, U.K. and Northampton.
- Van de Walle, Dominique and Kimberly Nead (1995), *Public Spending and the Poor: Theory and Evidence*, Baltimore and London, The Johns Hopkins University Press for the World Bank.
- Wachter, S. y S. Galiani (2000), “Optimal Income Support Targeting”.
- (2006), “Optimal Income Support Targeting”, *International Tax and Public Finance*, 13, 6, pp. 661-684.

## **Anexo**

### **A1. Regiones definidas para la estimación de los modelos estadísticos**

---

**Región 1.** Baja California, Baja California Sur, Coahuila, Chihuahua, Durango, Sonora.

**Región 4.** Coahuila, Nuevo León, Tamaulipas.

**Región 5.** Sinaloa, Sonora.

**Región 6.** Aguascalientes, Guanajuato, Jalisco, San Luis Potosí, Zacatecas.

**Región 7.** Durango, Jalisco, Nayarit, Sinaloa.

**Región 8.** Guanajuato, Hidalgo, Puebla, Querétaro, San Luis Potosí, Tlaxcala, Veracruz.

**Región 10.** Campeche, Quintana Roo, Tabasco, Yucatán.

**Región 11.** Guanajuato, Jalisco, Michoacán, Querétaro.

**Región 12.** Colima, Guerrero, Jalisco, Michoacán, Oaxaca.

**Región 13.** Guanajuato, Guerrero, México, Michoacán, Morelos, Querétaro.

**Región 14.** Distrito Federal, México.

**Región 15.** México, Morelos, Oaxaca, Puebla, Veracruz.

**Región 16.** Chiapas, Guerrero, Morelos, Oaxaca.

**Región 18.** Chiapas.

### A2. Cuadro comparativo tipos de focalización

| Método de focalización                     | Características  | Unidad de evaluación | Ventajas   | Desventajas   | Programas implementados en México por Sedesol     | Experiencia internacional  | Estudios de caso   |
|--|--|----------------------|--|---|---|--|--|
| Comprobación de medios de vida             | Se evalúa directamente hogar por hogar o por individuo, y se verifica el ingreso de cada uno.  | Hogar/ Individuo     | Si se logra obtener información veraz, puede ser muy precisa   | Puede generar incentivos perversos. Puede haber errores de reporte del ingreso.   | NA  | Garantía de ingresos mínimos Rumania (2002); subsidio a la vivienda (1980) Zambia. | Tesliuc <i>et al.</i> (2003); K. Subbarao <i>et al.</i> (1997).  |
| Comprobación de medios de vida aproximados | Consiste en calcular una calificación o "puntaje" a partir de ciertas características socioeconómicas del hogar, y a partir de un punto de corte determinado se puede hacer una clasificación entre pobre y no pobre. Este se lleva a cabo ya sea por medio de un análisis de regresión, análisis de componentes principales, análisis discriminante, etc. | Hogar/ Individuo     | Las características del hogar son fáciles de verificar. No hay manipulación en la información. Tiene menos probabilidad de crear incentivos perversos. | Se necesita un sistema eficiente de recopilación de datos, incluyendo tecnología y personal. La selección del punto de corte corre el riesgo de ser arbitraria. No capta cambios drásticos en el bienestar, como los que ocurren cuando hay una crisis económica. | Oportunidades; Hábitat; Programa de abasto social | Bolsa Scola (2001), Brasil; CAS-Subsidio Familiar (1980-1990), Chile.              | Grosh (1994) para Latinoamérica. Glewwe y Kanaana (1989), con información de Costa de Marfil. Grosh (1992) para Jamaica. |

## A2. Cuadro comparativo tipos de focalización (continuación)

| Método de focalización          | Características   | Unidad de evaluación | Ventajas  | Desventajas  | Programas implementados en México por Sedesol   | Experiencia Internacional  | Estudios de caso  |
|---------------------------------|---|----------------------|---|--|---|--|---|
| <i>Focalización comunitaria</i> | El líder o un grupo de miembros de una comunidad decide que miembro de la comunidad podría recibir los beneficios.  | Hogar/ Individuo     | Toma en cuenta las necesidades y el bienestar locales. Reduce los costos de identificación de los beneficiarios, transfiriéndolos a la comunidad. | Podría generar incentivos perversos en los líderes o miembros de la comunidad. Podría disminuir la cohesión social. El utilizar definiciones locales de bienestar dificulta la evaluación y la hace ambigua. | Oportunidades   | Programa de Fondo Social (1990), Nicaragua; Programa de Apoyo Económico, (1995), Albania.            | J. Van Domelen (2001); H. Alderman (2000).  |
| <i>Focalización geográfica</i>  | El lugar de residencia determina la elegibilidad para obtener los apoyos. Utiliza métodos de referenciación geográfica, que pueden abarcar desde mapas de pobreza hasta sistemas de posicionamiento global (GPS). | Área geográfica      | Es sencilla de administrar. Es fácil de combinar con otros métodos. No crea "estigmas" dentro de los miembros de la comunidad beneficiada.        | Depende de la precisión en la identificación de la comunidad. Tendrá un impacto pequeño si no existen áreas de concentración de la pobreza.  | Hábitat; Programa de Abasto Rural Diconsa; Estrategia de Microregiones; Programa de abasto social de leche Liconsa. | Fondo Social de Bolivia (1991); Bolsa Scola (2001), Brasil; Fondo de Bienestar Social (1996), Yemén. | Dalenius y Hodges (1957); M. Prahdan, L. Rawlings (1998); F. Bourguignon, F. Ferreira (2002); D. van de Walle (2002). |

**A2. Cuadro comparativo tipos de focalización (continuación)**

| Método de focalización          | Características  | Unidad de evaluación | Ventajas  | Desventajas  | Programas implementados en México por SEDESOL  | Experiencia Internacional   | Estudios de caso   |
|---------------------------------|--|----------------------|---|--|--|---|--|
| <i>Focalización demográfica</i> | La elegibilidad está determinada por el género, la edad y otras características demográficas.                        | Individual           | Es sencilla de administrar.<br>No crea "estigmas" en los individuos beneficiados.<br>Suele ser popular en términos políticos. | No tiene un gran impacto cuando esas características no están correlacionadas con la pobreza.                                      | Programa de Adultos Mayores;<br>Mujeres con Oportunidades;<br>Jornaleros Agrícolas;<br>Programa 3x1; Programa de abasto social de leche Liconsa. | Pensiones no contributivas (1974), Costa Rica; Programa de Pensiones en Hungría (1990); Ayuda alimentaria, trabajo por comida (1980-1990), Etiopía. | Grosh (1994); C. Grootaert (1997); D. Clay <i>et al.</i> (1999). |
| <i>Autofocalización</i>         | Programa abierto a toda la población, pero está diseñado de tal manera que su aceptación sea mayor entre los pobres. | Individual           | Bajos costos de administración.<br>Tiene menores posibilidades de generar incentivos perversos.                               | Puede crear "estigmas" entre la población que lo recibe.<br>Es difícil encontrar beneficios adecuados a este tipo de focalización. | Programa de Empleo Temporal; Programa de abasto social de leche Liconsa.   | Bono Madre Infantil (1990), Honduras; Subsidio al consumo de alimentos (1990), Marruecos.   | M. Grosh (1994); S. Razmara <i>et al.</i> (1999).                |

Fuente: Elaboración propia.

**A3. Cuadro de variables explicativas por modelo**

| <i>Significancia e impacto de las variables en cada modelo</i> |                                   |              |                             |
|--|-----------------------------------|--------------|-----------------------------|
|  | <i>Análisis<br/>Discriminante</i> | <i>Logit</i> | <i>Logit<br/>Multinivel</i> |
| <i>Variables del hogar</i>                                     |                                   |              |                             |
| Hacinamiento   | -                                 | -            | -                           |
| Sin refrigerador   | -                                 | -            | -                           |
| Sin estufa de gas  | -                                 | -            | 0                           |
| Piso de tierra   | -                                 | -            | -                           |
| Sin lavadora   | -                                 | -            | -                           |
| Estrato rural  | 0                                 | -            | NA                          |
| Sin excusado   | -                                 | -            | -                           |
| Sin conexión de agua   | -                                 | -            | -                           |
| Sin vehículo   | -                                 | -            | -                           |
| <i>Variables de los miembros del hogar</i>                     |                                   |              |                             |
| Número de niños  | -                                 | -            | -                           |
| Sin seguridad social   | -                                 | -            | -                           |
| Dependencia demográfica  | -                                 | -            | -                           |
| Escolaridad del jefe: primaria incompleta                      | -                                 | -            | -                           |
| Escolaridad del jefe: sin instrucción                          | 0                                 | -            | 0                           |
| Sexo del jefe  | 0                                 | 0            | +                           |
| Edad del jefe  | -                                 | 0            | -                           |
| <i>Variables de región</i>                                     |                                   |              |                             |
| Región 8   | +                                 | +            | NA                          |
| Región 14  | +                                 | +            | NA                          |
| Región 1   | +                                 | +            | NA                          |
| Región 16  | +                                 | +            | NA                          |
| Región 4   | +                                 | +            | NA                          |
| Región 15  | +                                 | +            | NA                          |
| Región 11  | +                                 | 0            | NA                          |
| Región 10  | +                                 | 0            | NA                          |
| Región 12  | +                                 | 0            | NA                          |
| Región 13  | +                                 | 0            | NA                          |
| Región 7   | +                                 | 0            | NA                          |
| Región 5   | +                                 | 0            | NA                          |
| Región 6   | +                                 | 0            | NA                          |

**A3. Cuadro de variables explicativas por modelo (continuación)***Variables de segundo nivel*

|         |    |    |   |
|---------|----|----|---|
| PRURURB | NA | NA | — |
| PESTGAS | NA | NA | — |
| PMUJER  | NA | NA | + |
| PISO_TI | NA | NA | 0 |
| PBAÑO2  | NA | NA | 0 |

*Fuente:* Cálculos propios con base en la ENIGH 2002. — Significativo y negativo (más pobre). + Significativo y positivo (menos pobre). 0 No significativo. NA: No Aplica.

**A4. Características socioeconómicas de los hogares**

| <i>Variable</i>  | <i>Hogares a nivel nacional</i> | <i>Hogares en pobreza de capacidades<sup>4</sup></i> |
|--|---------------------------------|--|
| Hogares rurales <sup>1</sup>                           | 23.6                            | 50.1   |
| Hogares con piso de tierra <sup>1</sup>                | 9.7                             | 30.3   |
| Hogares sin excusado <sup>1</sup>                      | 6.7                             | 18.1   |
| Hogares sin conexión de agua <sup>1</sup>              | 7.5                             | 18.1   |
| Hogares sin estufa de gas <sup>1</sup>                 | 13.5                            | 39.1   |
| Hogares sin refrigerador <sup>1</sup>                  | 23.8                            | 58.3   |
| Hogares sin lavadora <sup>1</sup>                      | 43.2                            | 76.1   |
| Hogares sin vehículo <sup>1</sup>                      | 52.8                            | 70.4   |
| Hogares sin seguridad social <sup>1</sup>              | 59.0                            | 86.5   |
| Hogares con mujeres jefas de familia <sup>1</sup>      | 20.0                            | 17.0   |
| Hogares con jefes sin instrucción <sup>1</sup>         | 13.7                            | 26.9   |
| Hogares con jefes con primaria incompleta <sup>1</sup> | 22.6                            | 34.3   |
| Hogares en pobreza de capacidades <sup>2</sup>         | 21.1                            | 100.0  |
| Índice de hacinamiento <sup>2</sup>                    | 1.8                             | 2.9  |
| Edad del jefe <sup>2</sup>                             | 47.1                            | 46.7   |
| Número de niños dentro del hogar <sup>2</sup>          | 1.0                             | 1.8  |
| Índice de dependencia demográfica <sup>2</sup>         | 0.7                             | 1.1  |
| Ingreso neto total per cápita <sup>3</sup>             | 2,328                           | 444  |

*Fuente:* Cálculos propios con base en la ENIGH 2002. <sup>1</sup> Porcentaje; <sup>2</sup> Promedios; <sup>3</sup> Ingreso promedio a pesos de agosto de 2002; <sup>4</sup> Con base en la metodología del CTMP.

**A5. Coeficientes de la función discriminante**

| <i>Variable</i>                            | <i>Coeficientes de la función</i> | <i>Matriz de estructura*</i> |
|--|-----------------------------------|------------------------------|
| Hacinamiento                               | -0.23                             | -0.592                       |
| Sin refrigerador                           | -0.48                             | -0.588                       |
| Sin estufa de gas                          | -0.66                             | -0.544                       |
| Piso de tierra                             | -0.46                             | -0.487                       |
| Sin lavadora                               | -0.26                             | -0.465                       |
| Estrato rural                              | -0.13                             | -0.435                       |
| Número de niños                            | -0.29                             | -0.430                       |
| Sin seguridad social                       | -0.51                             | -0.385                       |
| Dependencia demográfica                    | -0.19                             | -0.384                       |
| Escolaridad del jefe: primaria incompleta  | -0.19                             | -0.353                       |
| No tiene excusado                          | -0.32                             | -0.303                       |
| Sí tiene excusado pero no conexión de agua | -0.21                             | -0.267                       |
| Escolaridad del jefe: sin instrucción      | -0.06                             | -0.265                       |
| Región 8                                   | 0.48                              | -0.253                       |
| Vehículo motorizado                        | -0.13                             | -0.239                       |
| Región 14                                  | 1.09                              | 0.197                        |
| Región 1                                   | 1.11                              | 0.131                        |
| Región 16                                  | 0.87                              | -0.124                       |
| Región 4                                   | 1.17                              | 0.124                        |
| Región 15                                  | 0.75                              | -0.102                       |
| Región 11                                  | 1.00                              | 0.075                        |
| Región 10                                  | 0.62                              | -0.072                       |
| Sexo del jefe                              | 0.02                              | 0.047                        |
| Región 12                                  | 1.03                              | -0.045                       |
| Región 13                                  | 1.11                              | 0.018                        |
| Edad del jefe                              | -0.01                             | 0.007                        |
| Región 7                                   | 0.92                              | -0.006                       |
| Región 5                                   | 0.79                              | -0.002                       |
| Región 6                                   | 0.80                              | 0.001                        |
| Constante                                  | 1.010                             |                              |
| Lambda de Wilks                            |                                   | 0.612                        |
| Prueba F (29) g.l.                         |                                   | 11689415                     |
| Correlación canónica                       |                                   | 0.62                         |

84.2% de los casos clasificados coincidentes con el método del CTMP

*Fuente:* Cálculos propios con base en la ENIGH 2002. \* Correlaciones intragrupo combinadas entre las variables discriminantes y las funciones discriminantes canónicas tipificadas. Variables ordenadas por el tamaño de la correlación con la función.

**A6. Resultados de la regresión Logit**

| <i>Variable</i>                                  | <i>Coefficiente</i> | <i>Estadístico t</i> | <i>Razón de Momios</i> | <i>Probabilidad</i> |
|--|---------------------|----------------------|------------------------|---------------------|
| Rural  | 0.11                | (1.75)               | 1.12                   | 0.53                |
| Con piso de tierra                               | 0.35                | (4.32)**             | 1.42                   | 0.59                |
| Sin excusado                                     | 0.34                | (3.71)**             | 1.40                   | 0.58                |
| Sin conexión de agua                             | 0.22                | (2.48)*              | 1.25                   | 0.55                |
| Sin estufa de gas                                | 0.63                | (8.09)**             | 1.88                   | 0.65                |
| Sin lavadora                                     | 0.56                | (8.99)**             | 1.75                   | 0.64                |
| Sin refrigerador                                 | 0.48                | (7.38)**             | 1.61                   | 0.62                |
| Sin vehículo                                     | 0.43                | (7.36)**             | 1.53                   | 0.61                |
| Hacinamiento                                     | 0.34                | (15.84)**            | 1.41                   | 0.58                |
| Sexo del jefe                                    | -0.12               | (-1.82)              | 0.89                   | 0.47                |
| Dependencia demográfica                          | 0.28                | (6.27)**             | 1.33                   | 0.57                |
| Edad del jefe                                    | 0.01                | (3.77)**             | 1.01                   | 0.50                |
| Número de niños                                  | 0.51                | (16.37)**            | 1.66                   | 0.62                |
| Sin seguridad social                             | 1.20                | (18.27)**            | 3.31                   | 0.77                |
| Escolaridad del jefe:<br>sin instrucción         | 0.07                | (0.96)               | 1.07                   | 0.52                |
| Escolaridad del jefe:<br>con primaria incompleta | 0.43                | (6.55)**             | 1.54                   | 0.61                |
| Región 1   | -1.58               | (-9.53)**            | 0.21                   | 0.17                |
| Región 4   | -1.78               | (-9.69)**            | 0.17                   | 0.14                |
| Región 5   | -0.85               | (-4.65)**            | 0.43                   | 0.30                |
| Región 6   | -0.82               | (-4.64)**            | 0.44                   | 0.30                |
| Región 7   | -1.16               | (-4.61)**            | 0.31                   | 0.24                |
| Región 8   | -0.67               | (-4.40)**            | 0.51                   | 0.34                |
| Región 10  | -0.59               | (-3.56)**            | 0.55                   | 0.36                |
| Región 11  | -1.35               | (-8.38)**            | 0.26                   | 0.21                |
| Región 12  | -1.38               | (-7.24)**            | 0.25                   | 0.20                |
| Región 13  | -1.57               | (-10.06)**           | 0.21                   | 0.17                |
| Región 14  | -1.70               | (-10.83)**           | 0.18                   | 0.15                |
| Región 15  | -0.95               | (-6.28)**            | 0.39                   | 0.28                |
| Región 16  | -1.16               | (-7.09)**            | 0.31                   | 0.24                |
| Constante  | -4.06               | (-22.00)**           | 0.02                   | 0.02                |
| LR chi(29)                                       | 6574.75             |                      |                        |                     |
| Log Likelihood                                   | -5207.59            |                      |                        |                     |
| R <sup>2</sup> ajustada                          | 0.39                |                      |                        |                     |

Casos clasificados coincidentes con el método del CTMP 86.6%

*Fuente:* Cálculos propios con base en la ENIGH 2002. Variable dependiente: Pobreza de capacidades.

\*Estadístico t significativo al 5% \*\*Estadístico t significativo al 1%.

**A7. Modelos multinivel**

| <i>Variable</i>         | <i>Modelo i</i> |          | <i>Modelo ii</i> |          |
|-------------------------|-----------------|----------|------------------|----------|
|                         | <i>Coef</i>     | <i>T</i> | <i>Coef</i>      | <i>T</i> |
| Constante               | -5.07           | -35.67   | -5.73            | -27.57   |
| <i>PRURURB</i>          |                 |          | 2.25             | 3.73     |
| Piso de tierra          | 0.41            | 16.32    | 2.04             | 2.62     |
| <i>PPISO_TI</i>         |                 |          | -0.90            | -0.54    |
| <i>PBAÑO2</i>           |                 |          | 0.92             | 0.44     |
| <i>PMUJER</i>           |                 |          | -8.03            | -2.08    |
| Sin excusado            | 0.35            | 13.13    | 0.43             | 15.34    |
| Sin conexión de agua    | 0.26            | 9.86     | 0.25             | 9.23     |
| Sin estufa de gas       | 0.69            | 29.38    | 0.11             | 0.96     |
| <i>PESTGAS</i>          |                 |          | 2.40             | 4.92     |
| Sin refrigerador        | 0.44            | 22.25    | 0.44             | 22.06    |
| Sin lavadora            | 0.59            | 31.07    | 0.61             | 31.62    |
| Sin vehículo            | 0.39            | 21.73    | 0.39             | 21.60    |
| Hacinamiento            | 0.33            | 50.43    | 0.33             | 49.81    |
| Sexo del jefe           | -0.11           | -5.26    | -0.11            | -5.17    |
| Dependencia demográfica | 0.27            | 19.36    | 0.27             | 18.81    |
| Edad del jefe           | 0.01            | 10.89    | 0.01             | 10.32    |
| Número de niños         | 0.53            | 55.95    | 0.53             | 55.95    |
| Sin seguridad social    | 1.19            | 59.98    | 1.19             | 59.92    |
| Escolaridad del jefe:   |                 |          |                  |          |
| sin instrucción         | 0.02            | 0.90     | 0.01             | 0.22     |
| Escolaridad del jefe:   |                 |          |                  |          |
| con primaria incompleta | 0.46            | 22.88    | 0.46             | 22.85    |

*Estimación del componente de la varianza*

|                |       |       |
|----------------|-------|-------|
| Constante      | 0.259 | 0.129 |
| Piso de tierra | -     | 0.119 |
| Estufa de gas  | -     | 0.062 |

*Componentes de varianzas y covarianzas de nivel 2*

|                              |   |        |
|------------------------------|---|--------|
| Constante-piso de tierra     | - | 0.015  |
| Constante-estufa de gas      | - | -0.041 |
| Estufa de gas-piso de tierra | - | -0.010 |

*Probabilidades*

|                       |       |       |
|-----------------------|-------|-------|
| Log (P/1-P)           | -1.70 | -1.82 |
| Probabilidad estimada | 0.242 | 0.238 |

### A7. Modelos Multinivel (continuación)

| <i>Variable</i>                      | <i>Modelo i</i>  |          | <i>Modelo ii</i> |          |
|--------------------------------------|------------------|----------|------------------|----------|
|                                      | <i>Coef</i>      | <i>T</i> | <i>Coef</i>      | <i>T</i> |
| <i>Análisis de varianza</i>          |                  |          |                  |          |
| Varianza intraclase                  | 0.41             |          | 0.41             |          |
| Varianza entre grupos                | 0.26             |          | 0.13             |          |
| Varianza del residual                | 3.29             |          | 3.29             |          |
| Varianza total                       | 3.96             |          | 3.83             |          |
| R <sup>2</sup>                       | 0.10             |          | 0.11             |          |
| Ln (Valor de verosimilitud)          | -246223.10       |          | -243369.70       |          |
| Devianza                             | <b>492446.20</b> |          | <b>486739.40</b> |          |
| Coficiente de correlación intraclase | 0.073            |          |                  |          |

*Fuente:* Cálculos propios con base en la ENIGH 2002. Variable dependiente: Pobreza de capacidades. Los resultados representan los coeficientes de la regresión logística multinivel.