



Propuesta metodológica para la focalización individual de los programas sociales

JOSÉ A. VALDERRAMA TORRES Y JUAN PICHIHUA*

Se propone una metodología de identificación individual de potenciales beneficiarios de programas sociales. La metodología tiene dos etapas. En la primera etapa, se estima un índice de bienestar siguiendo una práctica extendida en la literatura especializada: análisis de componentes principales con escalamiento óptimo. En la segunda etapa, se determinan los puntos de corte del índice estimado que permiten distinguir a los hogares que califican como potenciales beneficiarios. El cálculo de los umbrales es lo novedoso de este trabajo, y se basa en la minimización de una función que depende de los errores de focalización a los que se enfrenta cualquier programa social: infiltración y subcobertura. A manera de ilustración, utilizando datos de la Encuesta Nacional de Hogares 2007 se presentan estimaciones para Perú comparando los niveles de pobreza oficial con los estimados a partir de la metodología propuesta.

Palabras Clave : Pobreza, focalización, índice de bienestar, programas sociales.

Clasificación JEL : I32.

El objetivo de este trabajo es proponer una metodología de dos etapas que permita identificar a los hogares que deberían ser atendidos de manera prioritaria por los programas sociales. En primer lugar, se estima un índice de bienestar que permite ordenar a los hogares según su nivel de pobreza (aproximado en base a información observable). En segundo lugar, se determina a partir de qué puntos de corte del índice de bienestar estimado un hogar puede ser clasificado como potencial beneficiario de programas sociales. Para la determinación del índice de bienestar, se utilizó la técnica de análisis de componentes principales con escalamiento óptimo. Se estimó un índice de bienestar distinguiendo entre ámbito urbano y rural, y, dentro de cada uno de estos grupos, se calcularon los umbrales teniendo en cuenta el dominio geográfico (Costa, Sierra, Selva y Lima Metropolitana). Para la determinación de los umbrales, se propone una estrategia endógena con sustento teórico y empírico basado en la minimización de una función que depende de los errores de focalización.

Las estimaciones tuvieron como fuente de datos la Encuesta Nacional de Hogares (ENAHO) correspondiente al año 2007. Los resultados fueron validados teniendo en cuenta criterios de consistencia

* Valderrama Torres: Dirección de Calidad del Gasto Público del Ministerio de Economía y Finanzas y Escuela de Economía de la Universidad de Lima, Jr. Junín 319, Lima 1, Teléfono +511 311 5930 (e-mail: jvalder@ulima.edu.pe). Pichihua: Facultad de Economía de la Universidad Nacional Agraria La Molina (e-mail: jpichihua@lamolina.edu.pe).

Los autores agradecen los comentarios de Edgard Abanto Millones, Julia Johannsen, Marcos Robles y de dos árbitros anónimos designados por el BCRP a versiones previas de este documento. Cualquier error u omisión es de exclusiva responsabilidad de los autores y no compromete a las instituciones a las que los autores representan.

CUADRO 1. Niveles de pobreza según dominio y ámbito geográfico, 2007 (porcentajes)

Dominio	Nivel de pobreza	Rural	Urbano	Total
Costa	Pobres extremos	10.5	2.1	4.0
	Pobres no extremos	27.6	23.1	24.1
	No pobres	61.9	74.9	72.0
Sierra	Pobres extremos	40.8	8.5	29.3
	Pobres no extremos	32.5	27.8	30.9
	No pobres	26.7	63.7	39.9
Selva	Pobres extremos	23.4	11.0	17.8
	Pobres no extremos	31.9	29.2	30.7
	No pobres	44.7	59.7	51.6
Lima Metropolitana	Pobres extremos	-	0.5	0.5
	Pobres no extremos	-	17.9	17.9
	No pobres	-	81.5	81.5
Total	Pobres extremos	32.9	3.5	13.7
	Pobres no extremos	31.7	22.3	25.6
	No pobres	35.4	74.3	60.7

FUENTE: INEI - ENAHO 2007. Elaboración propia.

interna y externa.¹ En el primer caso, se verificó que las variables que intervienen en la construcción del índice de bienestar tengan un aporte significativo al mismo y en el segundo se verificó que la estimación del índice empleando datos de la ENAHO 2008 pero con los parámetros estimados en el 2007 guarde una relación positiva y estadísticamente significativa con el gasto per cápita del 2008.

El documento se divide en cuatro partes. La sección 1 presenta estadísticas sobre los niveles de pobreza observados en el país y discute alternativas para su medición en ausencia de información sobre el gasto de los hogares. La sección 2 explica en detalle la metodología empleada tanto para la estimación de un índice de bienestar como para la determinación de los umbrales que determinan qué hogares deberían beneficiarse de los programas sociales. La sección 3 muestra los resultados obtenidos a partir de la metodología propuesta. La sección 4 resume las conclusiones del trabajo.

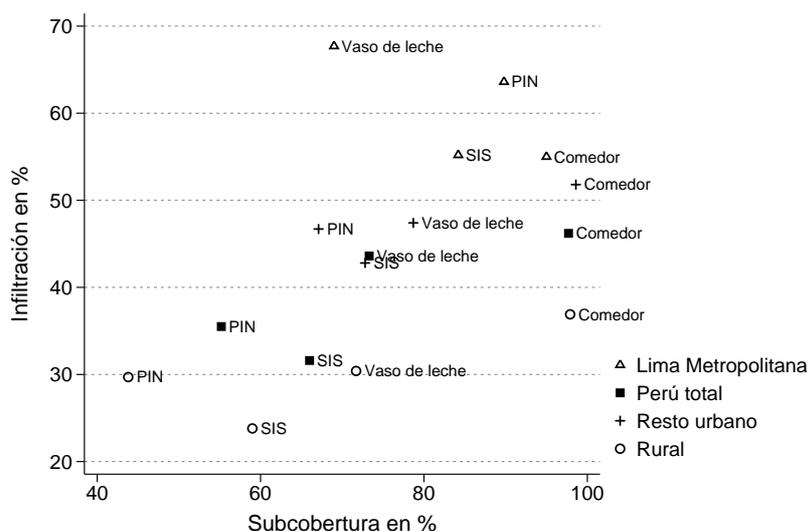
1 IDENTIFICACIÓN DE LA POBLACIÓN EN POBREZA EN PERÚ

Como se aprecia en el Cuadro 1, en Perú la población que padece de pobreza extrema representó el 13.7 por ciento de la población en el año 2007. Esta situación es mucho más grave en el ámbito rural respecto al urbano, pues en el primero este indicador es casi 10 veces mayor que el registrado en zonas urbanas.² La heterogeneidad en la incidencia de la pobreza extrema es aún mayor si se toma en cuenta los dominios geográficos, pues mientras en Lima Metropolitana esta condición afectó al 0.5 por ciento de la población, en la sierra rural este mismo indicador ascendió a 40.8 por ciento.

Mediante los programas sociales se busca atender a los más necesitados y con ello reducir los niveles de pobreza. Sin embargo, como puede apreciarse en el Gráfico 1, existe un importante porcentaje de

¹ La validez interna se refiere al nivel de confianza de los resultados dada una muestra, mientras que la validez externa se refiere a cuán generalizables son los resultados para otras muestras.

² En Perú, un centro poblado se define como urbano o rural según el tamaño de población. Según el INEI, El límite que distingue uno del otro es de 2000 personas o 400 viviendas.

GRÁFICO 1. Perú 2007. Infiltración y subcobertura en programas sociales

FUENTE: INEI - ENAHO 2007. Elaboración propia. SIS y PIN son siglas para Sistema Integral de Salud y Programa Infantil de Nutrición, respectivamente. “Comedor” se refiere a los Comedores Populares.

personas que son beneficiados sin ser parte del público objetivo (infiltración) o que, siendo parte del público objetivo, no son atendidos (subcobertura). Por ejemplo, 68 de cada 100 personas beneficiadas por el Vaso de Leche en Lima Metropolitana no deberían estar en el programa, mientras que 69 de cada 100 personas que sí tienen los atributos como para ser consideradas parte del público objetivo no recibieron este beneficio.³

Como el presupuesto siempre es limitado, un mecanismo eficiente en la asignación de los programas sociales es redireccionar los fondos destinados incorrectamente a los infiltrados para orientarlos a los subcubiertos.⁴ Este proceso se conoce como focalización, el cual enfrenta el desafío de identificar correctamente a los pobres. La mejor herramienta con la que se cuenta para aproximar el nivel de pobreza de un hogar es su nivel de consumo per cápita (pobreza monetaria). En Perú, la determinación de la pobreza monetaria se basa en el gasto per cápita obtenido a partir de la Encuesta Nacional de Hogares (ENAHO); si éste es menor a unos umbrales conocidos como líneas de pobreza y pobreza extrema, entonces el hogar es clasificado como pobre o pobre extremo, respectivamente. Sin embargo, esta pobreza no es observable para toda la población, sólo para los hogares encuestados en la ENAHO. En tal sentido, el reto que enfrenta cualquier sistema de focalización es identificar si un determinado hogar de la población es elegible o no en ausencia de información sobre pobreza monetaria.

El Sistema de Focalización de Hogares (SISFOH) es el órgano encargado de identificar a los hogares pobres en Perú. Uno de los insumos primordiales para este propósito es la aplicación presencial de una ficha socioeconómica en cada uno de los hogares. Esta ficha contiene tres tipos de pregunta: características

³ La población objetivo del Vaso de Leche y de otros programas sociales fue estimada a partir de la ENAHO de acuerdo con los requisitos que deberían cumplir los potenciales beneficiarios. Para ver el detalle de la población objetivo de los programas sociales estudiados ver por ejemplo Monge y otros (2009, p. 43).

⁴ Esto es una condición necesaria pero no suficiente, el problema pasa también por elevar los presupuestos destinados a los programas sociales. Este punto es notorio si se toma en cuenta que las zonas que se caracterizan por tener una alta incidencia en pobreza son también aquellas con los menores errores de focalización, como es el caso del ámbito rural.

de la vivienda, tenencia de activos y características demográficas de los miembros del hogar. A la fecha, el SISFOH ya cuenta con un registro de hogares a los cuales se ha aplicado la ficha. A fin de utilizar esta información para la identificación de los potenciales beneficiarios de programas sociales, la estrategia seguida por los organismos encargados es aproximar un índice de bienestar en base a la información observable en el padrón de hogares del SISFOH, construido de manera que esté altamente correlacionado con la pobreza monetaria. Los ponderadores de dicho índice de bienestar se obtienen de fuentes secundarias, utilizando información de la ENAHO, donde la pobreza monetaria sí es observable. A partir de este índice de bienestar, se definen puntos de corte (umbrales) que determinan la elegibilidad para los programas sociales. Los umbrales se estiman para distintos conglomerados geográficos con el fin de garantizar un costo admisible del error total de focalización, elevando el punto de corte de manera que se vaya identificando correctamente a un porcentaje de pobres hasta el momento en que las ganancias marginales de ampliar la cobertura de la población elegible conlleve a un incremento marginal del error relativamente alto. Ver detalles en MEF (2010).

2 METODOLOGÍA

Se propone una estrategia de identificación de la población elegible para programas sociales. Ésta consta de dos etapas. En la primera etapa, como la pobreza monetaria no es observable, se debe estimar un índice de bienestar, tal como la probabilidad de ser pobre (si se usa un modelo probabilístico), el gasto per cápita proyectado (por ejemplo, con un modelo de regresión lineal), o la creación de un índice estadístico utilizando técnicas multivariadas de procesamiento de grandes bases de datos. La segunda etapa consiste en la determinación del umbral o punto en el valor del índice a partir del cual un hogar puede ser considerado como potencial beneficiario de un programa social.

Para la determinación del índice se utiliza la técnica de componentes principales, la cual viene siendo empleada por el SISFOH (ver MEF, 2010).⁵ En este sentido, es la determinación de los umbrales el principal aporte de esta investigación. Para ello, se define una función de pérdida que depende de los errores de focalización, de la cual se determina óptimamente el punto que minimiza esta función.

Determinación del índice de bienestar

El análisis de componentes principales es una herramienta estadística que permite crear un indicador sintético capaz de explicar la mayor parte de la variabilidad de un conjunto de variables, y es por tanto un procedimiento pertinente para resumir la gran cantidad de información socioeconómica de los hogares con la que se cuenta.

Sin embargo, muchas de las características observadas de los hogares son variables cualitativas, es decir medidas en escala ordinal (por ejemplo, nivel educativo) o nominal (por ejemplo, sexo). Cuando las variables tienen esta característica no se puede emplear directamente esta metodología (Castaño y Moreno, 1994). Una forma de esta lidiar con este tipo de información es la asignación de un valor

⁵ Cabe mencionar que dentro del enfoque no monetario otra metodología tradicionalmente empleada para identificar pobres es el de las Necesidades Básicas Insatisfechas (NBI), propuesta por la CEPAL a comienzos de la década del 90. Consiste en el empleo de cinco indicadores que miden distintas dimensiones de la calidad de vida para formular un índice. Este enfoque se diferencia de lo propuesto en por lo menos tres aspectos: (a) se asigna la misma importancia a cada variable que interviene en su cálculo, a diferencia del enfoque de componentes principales que asigna un ponderador estimado para tomar en cuenta la volatilidad de la variable; (b) no permite hacer un ordenamiento extenso de hogares de tal forma que se detecte a los hogares más pobres, sino que sólo permite la clasificación de los hogares en seis grupos; (c) la selección de variables que intervienen en el cálculo del indicador es arbitraria. Esto lo diferencia del método propuesto, que toma en cuenta a aquellas variables que tengan un aporte significativo a la hora de identificar hogares pobres, dentro de un universo extenso de información.

arbitrario a cada una de las categorías en las que se divide una variable. No obstante, además de la subjetividad del proceso, este tipo de valoración no tiene en cuenta la relación de cada categoría con las categorías de otras variables de interés. Es decir, para cada variable cualitativa se estaría haciendo una valoración unidimensional, perdiendo información sobre la relación que tiene una variable con respecto a las demás. Para abordar técnicamente este inconveniente se usa el procedimiento denominado “escalamiento óptimo” (Young, 1981), que asigna valores numéricos a las categorías de las variables con el fin de maximizar las correlaciones entre todas las variables de interés. Ello es consistente con el procedimiento estándar de componentes principales, toda vez que busca resumir información de un conjunto de variables correlacionadas. Éste es, a su vez, precisamente el caso al que nos enfrentamos, ya que los hogares pobres tienden a tener valores bajos en casi todas las características medidas (viviendas precarias, bajo nivel educativo, acceso limitado a servicios básicos, entre otras) y, por ende, un valor también bajo en el indicador final de nivel de bienestar.

Como el interés es contar con un índice de pobreza, se clasificaron a todas las variables como ordinales y fueron recodificadas de tal forma que la primera categoría siempre haga referencia a la condición más desfavorable. Por ejemplo, en la variable “educación del jefe de hogar” la primera categoría es “sin ningún nivel educativo”. A partir de ese conjunto de variables se estima un índice que, al ser una combinación lineal de las variables elegidas, tiene la misma interpretación: menores valores indican que el hogar se encuentra en condiciones más desfavorables. Para la selección de variables se siguió la estrategia empleada por el SISFOH (MEF, 2010, p. 10) utilizando información de la ENAHO 2007. Así, las variables tomadas en cuenta fueron aquéllas con un mayor nivel de correlación con la pobreza monetaria según el coeficiente de asociación D de Somers, que permite medir el grado de asociación entre variables ordinales.

Umbral de bienestar

El siguiente paso es determinar a partir de qué nivel o umbral del índice construido un hogar puede ser clasificado como potencial beneficiario de un programa social. Las metodologías usuales se pueden distinguir según dos enfoques: relativo y absoluto. El primer tipo de metodología apela al uso de estadísticos de locación como quintiles o deciles. Por ejemplo, son potenciales beneficiarios aquellos hogares que se encuentran en los dos primeros quintiles de la distribución. En el caso del enfoque absoluto una alternativa es emplear análisis de conglomerados donde los puntos de corte son establecidos de tal forma que los grupos son similares al interior de los mismos pero distintos si se comparan entre grupos. En este trabajo se opta por una ruta alternativa inspirada en el principal impacto que espera alcanzar cualquier sistema de focalización: minimizar los errores de focalización. A continuación se discuten los fundamentos teóricos que justifican este procedimiento.

La matriz de clasificación según error de inclusión y exclusión es la base de la propuesta de tipificación. Por convención se define como “infiltrados” a aquellas personas que acceden a un programa social sin pertenecer a la población objetivo. Por su parte, “subcobertura” se refiere al porcentaje de la población objetivo que no recibe la atención de un programa social pese a ser elegible. Ambas categorías son *ex-post* al materializarse una vez que el programa social es implementado. En el panel izquierdo del Cuadro 2 (p. 70) se ilustran ambos conceptos.

Por su parte, las nociones alternativas de errores de inclusión y exclusión son definiciones que permiten clasificar a las personas *antes* de que estas sean o no beneficiarias de algún programa social, como se ilustra en el panel derecho del Cuadro 2. Para fines de nuestro análisis, se define el error de inclusión como el porcentaje de personas incorrectamente clasificadas como potenciales beneficiarios respecto a la población total, mientras que el error de exclusión alude al porcentaje de personas no beneficiarias que sí califican como parte de la población objetivo (respecto a la población total). Esta clasificación es posible debido a que se cuenta con un índice de bienestar del hogar, de tal forma que pueden ordenarse

CUADRO 2. Errores de focalización ex-post y ex-ante

		¿Es beneficiario?	
		Si	No
¿Es pobre?	Si		Subcobertura
	No	Infiltración	

		¿Potencial beneficiario?	
		Si	No
¿Es pobre?	Si		Error de Exclusión
	No	Error de Inclusión	

los hogares según su nivel de bienestar: desde el menor nivel posible (0) hasta el máximo nivel (100). En este punto el problema consiste en determinar el nivel del índice que distingue a un potencial beneficiario de un hogar que no califica como tal. El error de inclusión alcanza su mayor valor cuando el umbral toma el valor máximo posible. Esto se debe a que, cuando el umbral toma su valor máximo, toda la población es potencialmente elegible para programas sociales (el índice de bienestar de todos los individuos está por debajo del umbral de bienestar). En el otro extremo, el error de inclusión es cero cuando el umbral toma el valor mínimo posible, pues en este caso ningún individuo es elegible (todos los individuos tienen un índice de bienestar por encima del umbral). De manera análoga, el error de exclusión es cero cuando el umbral toma el valor máximo posible (toda la población es potencial beneficiaria) y alcanza su valor máximo cuando el umbral toma su valor mínimo (ningún individuo es potencial beneficiario).

Puede deducirse que las curvas que relacionan determinados niveles del umbral con el error de inclusión y exclusión tomarían formas similares a las de una distribución de probabilidad acumulada, tal como se muestra en el Gráfico 2, pues a medida que se acercan a los extremos el número de personas que se va incluyendo o excluyendo es cada vez menor hasta alcanzar un límite superior o inferior. Por tanto, no es necesario asumir ninguna forma funcional paramétrica: la relación no lineal que existe entre el error de inclusión y exclusión ante los posibles valores que podría tomar el nivel del umbral es implícita y depende de los datos.

Por construcción, este tipo de curvas de concentración no presenta mínimos interiores (son curvas monótonicas), por lo que buscar un umbral que minimice alguno de los errores es una tarea infructuosa.

GRÁFICO 2. Formas hipotéticas de las curvas de inclusión, exclusión y error de focalización total

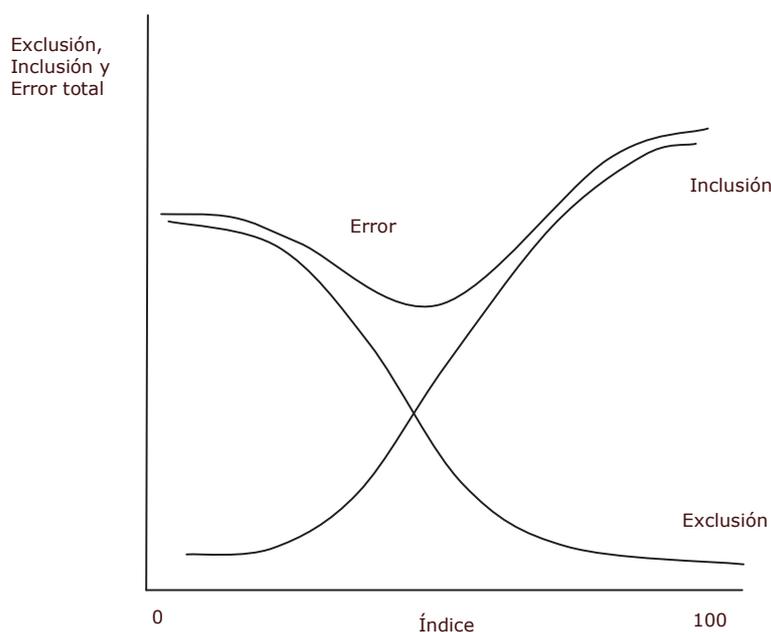
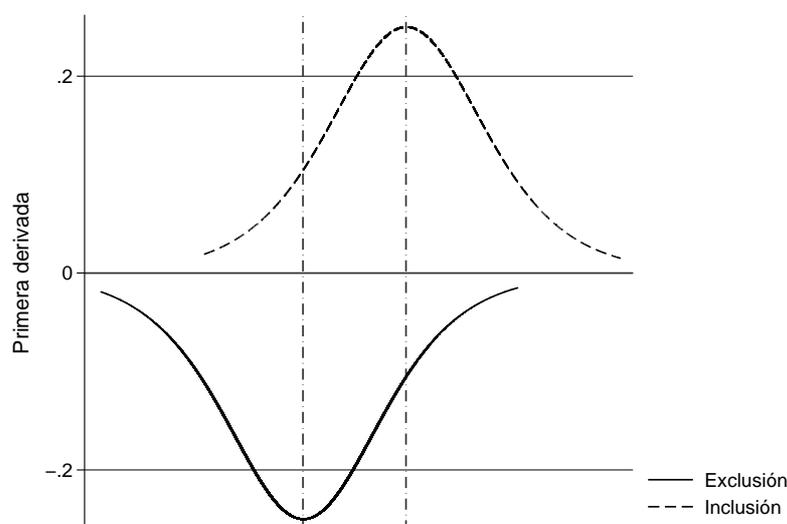


GRÁFICO 3. Región donde se garantiza un mínimo en la función de error de focalización total

Sin embargo, es posible encontrar condiciones que garanticen que la suma de ambos errores tenga un mínimo global. Teóricamente, esta suma es atractiva en el sentido que se está ponderando o dando la misma importancia a las personas que se encuentran clasificadas en los dos tipos de errores. Nótese que esta característica de la variable impide que se trabaje con las tasas de infiltración y subcobertura convencionales, pues estas se encuentran expresadas como porcentajes de poblaciones distintas (en el caso de infiltración el total son los beneficiarios y en el caso de la subcobertura es la población objetivo). Además, si se trabajara con estos indicadores no se puede garantizar continuidad en la función que resulte de la sumatoria, condición fundamental para analizar la existencia de óptimos. Por estas razones, se emplean errores de inclusión y exclusión ya definidos, que no son otra cosa que la población afectada en cada caso dividida por el total de la población.

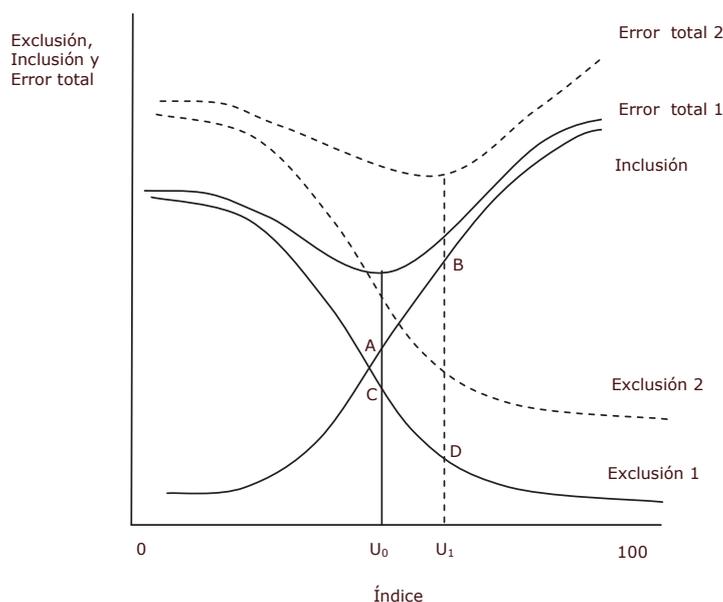
Sea I el error de inclusión, E el error de exclusión y T la suma de ambos tipos de errores. U representa un índice de bienestar definido entre 0 y 100 y que es el argumento del cual dependen I y E . Así, se tiene que $T(U) = I(U) + E(U)$. La existencia de un mínimo (U^*) exige que se cumplan las condiciones de primer (CPO) y segundo orden (CSO):

$$\text{CPO : } \frac{\partial T}{\partial U} = \frac{\partial I}{\partial U} + \frac{\partial E}{\partial U} = 0 \quad \text{y} \quad \text{CSO : } \frac{\partial^2 T}{\partial U^2} = \frac{\partial^2 I}{\partial U^2} + \frac{\partial^2 E}{\partial U^2} > 0. \quad (1)$$

Una condición suficiente para que exista un mínimo es que las segundas derivadas de $I(U)$ y $E(U)$ sean ambas positivas. De ser éste el caso, la función de error, que es resultado de la suma de ambos tipos de errores, toma la forma convexa mostrada en el Gráfico 2, donde el punto mínimo de la función T estaría ubicado en algún punto ubicado en el área encerrada por el intervalo determinado por las líneas verticales del Gráfico 3, pues en esa área ambas pendientes (segundas derivadas de I y E) son positivas.

La función de pérdida definida inicialmente puede generalizarse para dar una ponderación distinta a cada tipo de error: $T(U) = I(U) + \theta E(U)$, donde $\theta > 0$. El parámetro θ es una variable de política que determina la importancia que se le da al error de exclusión relativo al error de inclusión. En particular, se tiene que si $\theta < 1$, se le da más importancia al error de inclusión; si $\theta > 1$, se le da más importancia al error de exclusión; y si $\theta = 1$, ambos errores tienen la misma importancia.

El Gráfico 4 (p. 72) ilustra cómo una mayor ponderación al error de exclusión se refleja en un desplazamiento hacia la derecha del umbral que determina el error mínimo posible que se puede cometer.

GRÁFICO 4. Desplazamiento de las curvas debido a una mayor ponderación del error de exclusión

Ante un incremento de θ , el error de inclusión aumentaría pasando del punto A al punto B y el de exclusión se reduciría al pasar del punto C al punto D. La sensibilidad de los cambios dependerá de la elasticidad de las curvas respecto al valor del umbral. Cuanto más elásticas, mayor será el impacto sobre los errores.

En principio, asumir un ponderador mayor a 1 es razonable si el Estado es más sensible al error por exclusión que al error por inclusión. En la exploración empírica se fija el valor de $\theta = 2$. Una manera de interpretar este supuesto es que el punto que se elija como umbral es aquél donde se permite que en el margen se infiltren dos personas, con tal de que la subcobertura disminuya en una persona. Matemáticamente, ello se deduce de la condición de primer orden:

$$T(U) = I(U) + 2 \cdot E(U) \quad \Rightarrow \quad \text{CPO} : -\frac{\partial I / \partial U}{\partial E / \partial U} = 2. \quad (2)$$

Cabe mencionar que la estrategia de obtención del umbral o punto de corte es similar a lo desarrollado por Wodon (1997), quien propone definir una función de bienestar que depende de indicadores de sensibilidad, el porcentaje de pobres identificados correctamente, y especificidad, el porcentaje de no pobres identificados correctamente.

La metodología propuesta difiere de la de Wodon en un aspecto fundamental: los argumentos de las funciones objetivos son distintos. Al emplear los indicadores de sensibilidad y especificidad, Wodon relativiza la calidad de la predicción respecto al total de pobres y no pobres, respectivamente. Así, si se tuviera un universo de pobres y no pobres de 1000 y 100, respectivamente, y así mismo se lograra identificar correctamente a 100 y 10 pobres y no pobres, respectivamente, los indicadores de sensibilidad y especificidad serían ambos de 10 por ciento. Con los mismos datos, la metodología propuesta (con $\theta = 1$) señalaría que se estaría cometiendo un error de exclusión de 900 personas (1000-100) y un error de inclusión de 90 personas (100-10); es decir, el problema de focalización es más grave en el caso de la exclusión.⁶

⁶ Con fines computacionales la metodología propone dividir ambos errores por la población total, debe notarse que esta operación no altera el valor óptimo del umbral pues este parámetro es una constante en la función de pérdida, o dicho de otro modo es una transformación monótonica.

3 RESULTADOS

Determinación del índice de bienestar

El índice de bienestar se estimó utilizando información de la ENAHO 2007, siguiendo los criterios establecidos previamente. A partir de un conjunto inicial de alrededor de 40 variables, se seleccionaron las variables mostradas en el Cuadro 3 para cada ámbito geográfico. Cada una de estas variables registró un coeficiente de correlación (*D* de Somers) estadísticamente distinto de cero con el indicador de pobreza monetaria (columna 1). Asimismo, a manera de análisis *ex-post*, se muestra la correlación entre los quintiles del índice estimado y las variables empleadas (columna 2).

El Cuadro 4 (p. 74) muestra las puntuaciones empleadas para la construcción del índice de bienestar. Éstas se encuentran reescaladas, de tal modo que se garantiza que el índice varíe entre 0 y 100.

CUADRO 3. Variables que conforman el índice de bienestar según ámbito geográfico

Dominio	Variable	Asociación con pobreza ^(d)	Asociación con índice ^(d)
Urbano	Hacinamiento	0.480	0.403
	Nivel educativo del jefe del hogar	0.358	0.346
	Combustible empleado para cocinar	0.379	0.265
	Tenencia de teléfono fijo	0.375	0.334
	Tenencia de cocina a gas	0.256	0.146
	Tenencia de refrigeradora/congeladora	0.396	0.322
	Tenencia de televisor a color	0.290	0.190
	Tenencia de plancha eléctrica	0.339	0.220
	Tenencia de equipo de música	0.300	0.234
	Bienes identificadores de riqueza ^(a)	0.459	0.459
	Material predominante en las paredes exteriores	0.365	0.283
	Material predominante en los pisos	0.415	0.408
	Material predominante en los techos	0.364	0.326
	Tipo de abastecimiento de agua en el hogar	0.221	0.161
	Tipo de servicio higiénico en el hogar	0.337	0.242
	Número de artefactos eléctricos ^(b)	0.581	0.491
	Afiliados a algún seguro de salud (excluye SIS)	0.409	0.341
Rural	Número de artefactos eléctricos ^(b)	0.335	0.317
	Hacinamiento	0.294	0.281
	Miembros del hogar menores de 6 años	0.307	0.280
	Tenencia de cocina a gas	0.223	0.211
	Tenencia de televisor a color	0.205	0.194
	Tasa de ocupados en el hogar ^(c)	0.206	0.190
	Material predominante en los pisos	0.190	0.180
	Tenencia de plancha eléctrica	0.183	0.171
	Mujeres adultas analfabetas	0.185	0.167
	Vivienda con piso de tierra (en la Sierra)	0.204	0.157
	Afiliados a algún seguro de salud (excluye SIS)	0.121	0.112

NOTAS: (a) Computadora, refrigeradora, cable, internet y lavadora. (b) Televisor a color, equipo de sonido, DVD, computadora, plancha eléctrica, licuadora, cocina a gas, refrigeradora/congeladora, lavadora y microondas. (c) El numerador considera a los mayores de 18 años ocupados y el denominador al total de adultos. (d) La columna “Asociación con pobreza” reporta las correlaciones (*D* de Somers) específicas entre cada variable y el nivel de pobreza; la columna “Asociación con índice” muestra la correlación (*D* de Somers) entre los quintiles del índice estimado y las variables.

CUADRO 4. Puntuaciones empleadas para la estimación del índice de bienestar del hogar

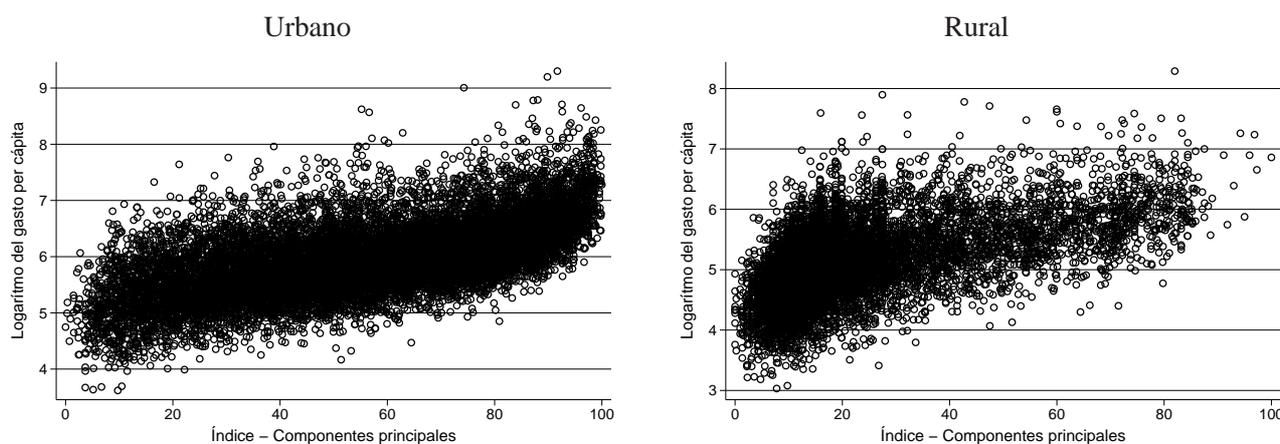
Variable y categorías	Urbano	Rural	Variable y categorías	Urbano	Rural
Hacinamiento			El material predominante en los techos es:		
6 a más	-0.955	-1.062	Otro	-0.872	-
Entre 4 y 6	0.268	0.246	Paja	-0.872	-
Entre 2 y 4	1.751	1.018	Estera	0.613	-
Entre 1 y 2	4.217	3.584	Plancha	1.603	-
Menos de 1	5.561	4.728	Caña	1.603	-
Educación del jefe del hogar			Tejas	1.603	-
Sin educación	0.434	-	Madera	3.231	-
Inicial	1.051	-	Concreto	5.668	-
Primaria	1.909	-	Abastecimiento de agua en el hogar procede de :		
Secundaria	3.324	-	Otro	-0.308	-
Técnica	4.865	-	Rio	-0.308	-
Universitario	6.256	-	Pozo	0.391	-
Post Grado	7.478	-	Camión	0.833	-
Tipo de combustible para cocinar			Pilón	0.833	-
No cocinan	-0.719	-	Fuera	1.499	-
Otro	-0.719	-	Dentro	4.409	-
Leña	-0.719	-	Servicio higiénico que tiene el hogar esta conectado a :		
Carbón	0.904	-	No tiene	-1.231	-
Kerosene	1.100	-	Rio	0.178	-
Gas	4.736	-	Pozo ciego	0.570	-
Electricidad	6.262	-	Pozo séptico	0.570	-
Teléfono fijo			Fuera	1.480	-
No	1.637	-	Dentro	4.829	-
Si	5.887	-	Número de artefactos electrodomésticos		
Cocina a gas			Ninguno	-1.360	-0.711
No	-0.392	0.314	Sólo 1	-1.360	2.447
Si	4.501	9.487	Sólo 2	-0.276	4.932
Refrigeradora/Congeladora			Sólo 3	1.042	7.387
No	0.858	-	Sólo 4	2.164	10.386
Si	5.944	-	Sólo 5	3.626	12.801
TV a color			Sólo 6	4.809	14.935
No	-0.407	0.215	Sólo 7	6.052	16.140
Si	4.673	9.655	Sólo 8	7.258	17.685
Plancha eléctrica			Sólo 9	8.145	17.685
No	-0.237	0.669	Sólo 10	8.145	17.685
Si	4.870	11.898	Afiliados a algún seguro de salud		
Equipo de música			Ninguno	1.566	1.363
No	1.776	-	Uno	3.751	5.911
Si	5.541	-	Dos	5.107	9.827
Bienes identificadores de riqueza			Tres	5.897	12.423
Ninguno	0.527	-	Más de 3	6.057	12.423
1	4.632	-	Tasa de ocupados en el hogar		
2	6.714	-	No ocupados	-	0.973
3	7.895	-	De 0 a 25%	-	0.973
4	7.895	-	De 25% a 50%	-	1.886
El material predominante en las paredes exteriores es :			De 50% a 75%	-	3.339
Otro	0.043	-	De 75% a 100%	-	3.339
Madera o estera	0.043	-	Menores de 6 años		
Piedra con barro	0.043	-	Más de 2	-	-0.672
Quincha	0.194	-	Sólo 2	-	0.552
Adobe o tapia	0.443	-	Sólo 1	-	2.483
Piedra o sillar con cal o cemento	1.098	-	Ninguno	-	3.210
Ladrillo o bloque de cemento	5.075	-	Mujeres adultas analfabetas		
El material predominante en los pisos es :			Más de 1	-	-0.446
Otro	-0.361	-0.329	Sólo 1	-	0.241
Tierra	-0.361	0.672	No hay	-	3.534
Cemento	3.770	7.999	Piso de tierra (en la Sierea)		
Madera	3.770	7.999	Si	-	-0.316
Losetas	6.768	17.375	No	-	3.623
Láminas	6.948	18.459			
Parquet	7.204	20.419			

La validación del índice propuesto tuvo dos componentes: uno interno y otro externo. En el primer caso se excluyó a aquellas variables con una contribución a la varianza de menos de 1 por ciento, y se empleó el índice de consistencia interna “ α ” de Cronbach. Este indicador mide el grado de asociación que tienen todas las variables en conjunto, varía entre 0 y 1, y valores mayores a 0.8 son considerados como muy buenos. Los resultados se muestran en el Cuadro 5. Tanto para el caso urbano como rural los resultados son favorables, pues el índice de Cronbach es 0.91 y 0.80, respectivamente.

Para el ejercicio de validación externa se generó el índice de bienestar y se halló la correlación entre dicho índice y el logaritmo del gasto per cápita utilizando datos de la ENAHO 2007. En el Gráfico 5 (p. 76) se muestra la relación entre ambas variables. Aunque moderada, la correlaciones son positivas y estadísticamente significativas (0.50 y 0.46 para el ámbito urbano y rural, respectivamente). En estricto rigor esta validación requiere de una predicción fuera de muestra. Por ello, se estimó el índice de bienestar usando datos de la ENAHO 2008 pero con las puntuaciones obtenidas de la ENAHO 2007, para luego hallar la correlación entre dicho índice y el logaritmo del gasto per cápita de 2008. Los resultados (disponibles a solicitud) fueron similares a los aquí reportados, confirmando así la relación positiva y estadísticamente significativa entre el índice y el logaritmo del gasto per cápita.

CUADRO 5. Validación interna: varianza explicada, aporte a la varianza y α de Cronbach

	Aporte a la varianza explicada	Aporte porcentual
Urbano - varianza total: 10.151, varianza explicada: 6.849 (67% del total), α de Cronbach: 0.91		
Número de artefactos eléctricos	0.799	11.7
Bienes identificadores de riqueza	0.686	10.0
Tenencia de refrigeradora/congeladora	0.558	8.1
Material predominante en los pisos	0.464	6.8
Tenencia de plancha eléctrica	0.429	6.3
Tenencia de teléfono fijo	0.388	5.7
Tenencia de televisor a color	0.380	5.5
Material predominante en los techos	0.380	5.5
Tipo de conexión del servicio higiénico	0.368	5.4
Material predominante en las paredes exteriores	0.359	5.2
Afiliados a algún seguro de salud (excluye SIS)	0.337	4.9
Tenencia de cocina a gas	0.320	4.7
Tiene equipo de música	0.306	4.5
Hacinamiento	0.236	3.4
Educación del jefe del hogar	0.228	3.3
Tipo de abastecimiento de agua en el hogar	0.220	3.2
Rural - varianza total: 7.351, varianza explicada: 3.649 (50% del total), α de Cronbach: 0.80		
Número de artefactos eléctricos	0.859	23.5
Tenencia de televisor a color	0.578	15.8
Tenencia de plancha eléctrica	0.574	15.7
Tenencia de cocina a gas	0.536	14.7
Material predominante en los pisos	0.378	10.4
Afiliados a algún seguro de salud (excluye SIS)	0.249	6.8
Vivienda con piso de tierra (en la Sierra)	0.145	4.0
Hacinamiento	0.123	3.4
Mujeres adultas analfabetas	0.110	3.0
Miembros del hogar menores de 6 años	0.069	1.9
Tasa de ocupados en el hogar	0.027	0.7

GRÁFICO 5. Validación externa: Asociación entre el gasto per cápita y el índice de bienestar

Determinación de los umbrales

Sobre la base del índice estimado y siguiendo el desarrollo teórico planteado anteriormente se estimaron los umbrales de tal forma que se minimice el error total, definido como la suma de los errores de inclusión y el doble del error de exclusión. Para tener en cuenta la heterogeneidad por ámbito y dominio geográfico, el ejercicio se realizó de manera independiente para la Costa, Sierra, Selva y Lima Metropolitana. Los resultados se reportan en el Cuadro 6, mientras que el Gráfico 6 (p. 77) muestra las curvas que dan origen a los umbrales.

En el caso de la Costa urbana por ejemplo, cuando el objetivo es reducir los errores de focalización

CUADRO 6. Determinación de los umbrales que minimizan el error de focalización

Ámbito Geográfico	Nivel de pobreza	Errores (%)		Total	Umbral	Intervalo de confianza
		Inclusión	Exclusión			
Costa urbana	PE	0.3	1.7	3.8	8.2	[8.0, 8.4]
	PT	10.8	9.7	30.1	43.1	[42.5, 43.8]
Sierra urbana	PE	4.3	4.0	12.3	18.4	[18.1, 18.7]
	PT	19.0	4.6	28.2	53.1	[52.6, 53.7]
Selva urbana	PE	9.0	4.6	18.2	18.4	[18.1, 18.7]
	PT	18.8	7.6	34.1	46.1	[45.5, 46.6]
Lima Metropolitana	PE	0.3	0.4	1.1	13.8	[13.4, 14.2]
	PT	11.0	8.4	27.9	50.7	[50.4, 51.0]
Costa rural	PE	3.4	6.7	16.7	12.4	[12.2, 12.6]
	PT	23.6	5.3	34.2	34.8	[33.9, 35.6]
Sierra rural	PE	36.0	4.2	44.4	23.3	[22.7, 23.9]
	PT	20.5	36.0	92.5	52.2	[51.9, 52.6]
Selva rural	PE	7.4	15.1	37.6	11.1	[11.0, 11.3]
	PT	30.4	3.6	37.5	35.8	[34.9, 36.6]

NOTAS: se obtuvo dos umbrales por cada dominio geográfico. El de menor valor se obtuvo al asumir que la población objetivo son únicamente los pobres extremos (PE) y el de mayor valor cuando la población objetivo es el total de pobres (PT). Los intervalos al 95% de confianza fueron obtenidos utilizando la técnica de *bootstrapping* con 200 réplicas asumiendo normalidad en la distribución del umbral. El remuestreo fue sobre la ENAHO 2007.

GRÁFICO 6. Representación gráfica de la determinación de los umbrales

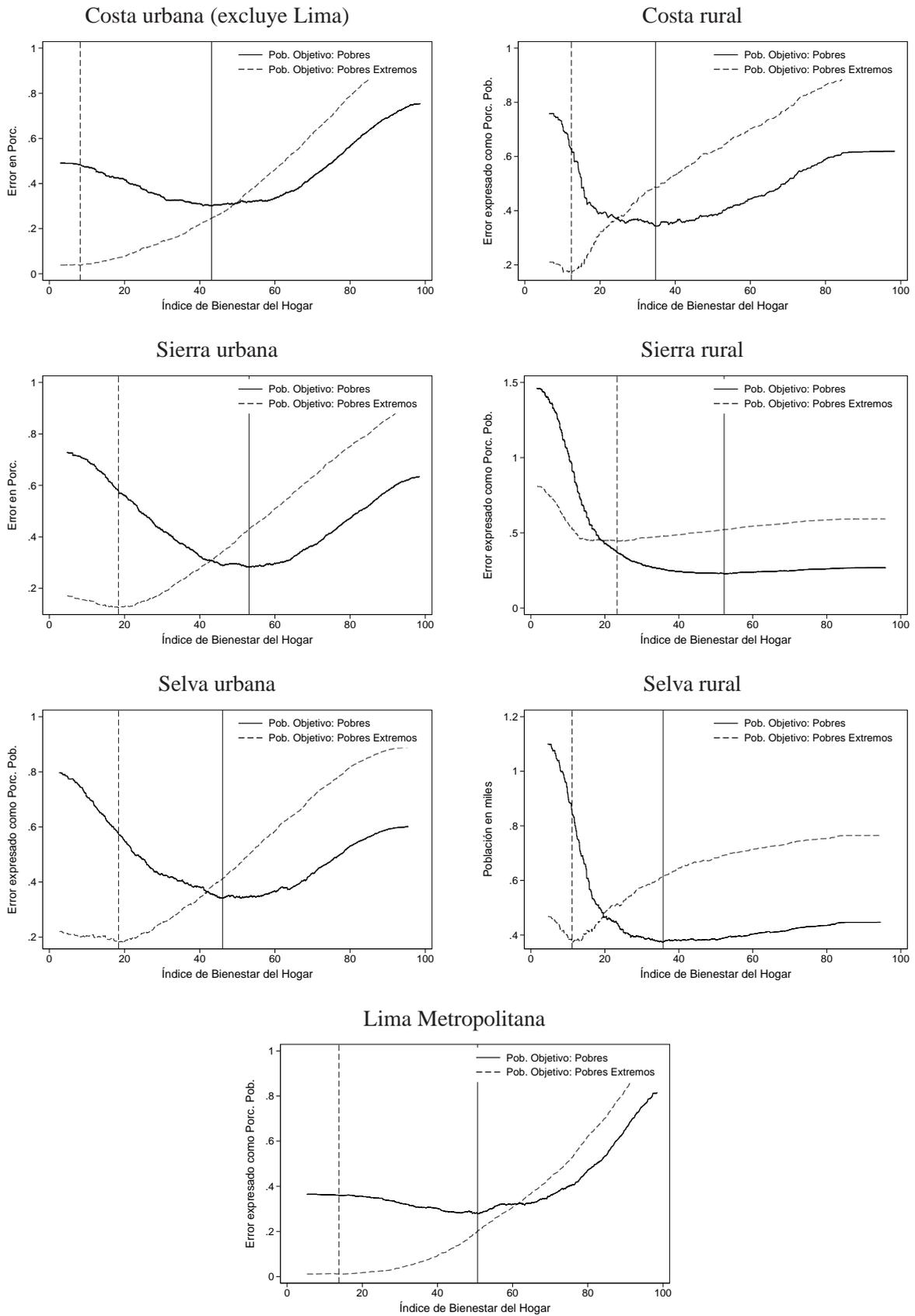
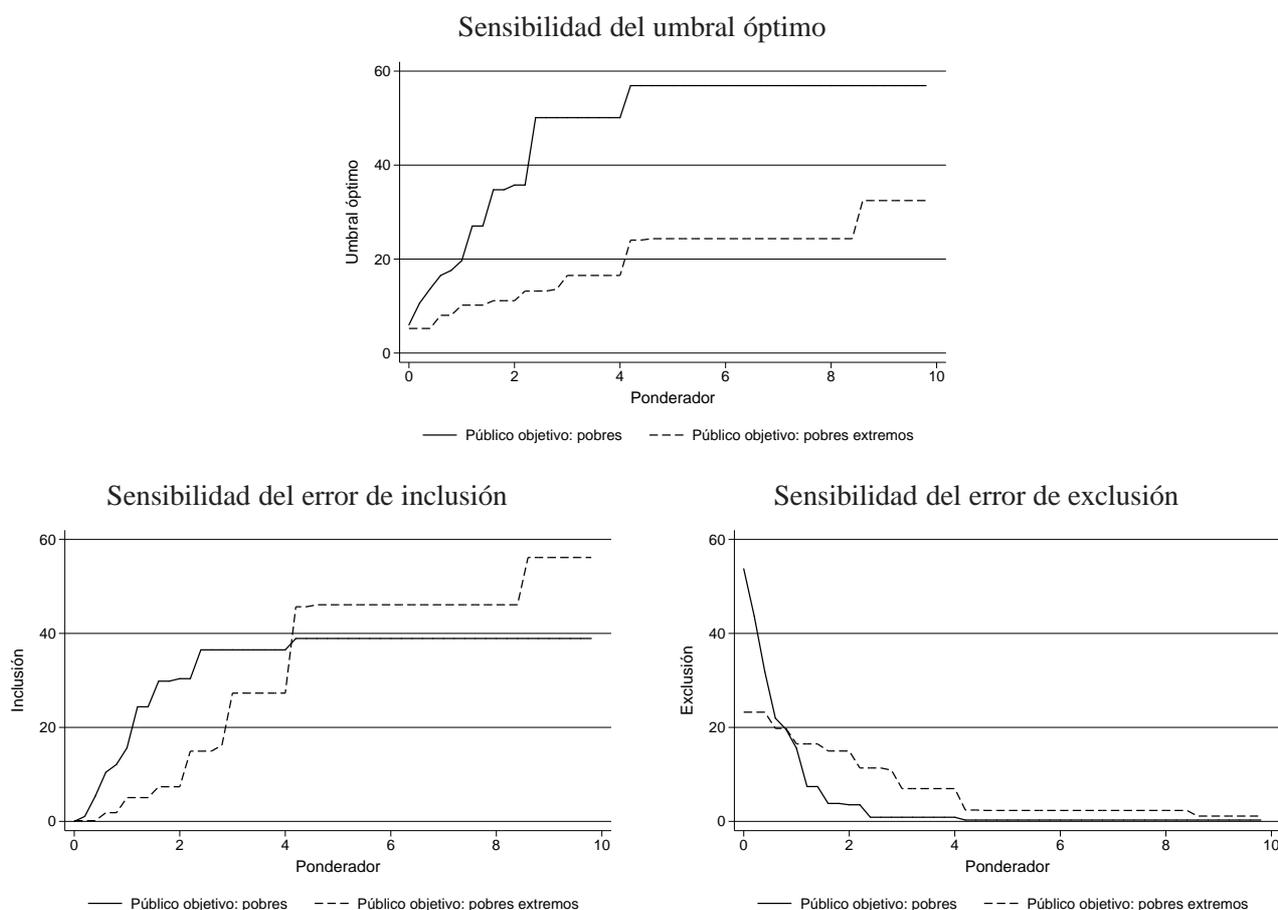


GRÁFICO 7. Selva rural: sensibilidad de los resultados ante distintos valores del ponderador θ 

al tratar de identificar a los pobres extremos, se comete un error de inclusión de 0.3% y un error de exclusión de 3.7%, que de acuerdo a la función de error planteada (el error de exclusión se pondera el doble) totaliza un error de 3.8%. El punto de corte en el índice planteado en este escenario es de 8.2. La variabilidad muestral en la estimación de este umbral o punto de corte fue estimado a partir de la técnica del *bootstrapping*, el intervalo de confianza a un nivel de 95% (última columna del Cuadro 6) se estima en el rango que va de 8.0% al 8.4%.

Como se explicó, los cálculos mostrados asumen un valor de $\theta = 2$. Aunque ésta es una variable de política, resulta también de interés conocer los efectos que provoca diversos valores de este parámetro. Información adicional de este tipo permitiría eventualmente ayudar en la decisión de la elección de este valor. Por ejemplo, en el caso de la Selva rural, los resultados son sensibles cuando el ponderador se encuentra entre 0 y 4, pero tiende a generar resultados robustos a partir de 4. Ver Gráfico 7.

Población elegible versus pobreza oficial

A continuación se procede a comparar los porcentajes de la población elegible que se derivan de la metodología propuesta con los niveles oficiales (Cuadro 7, p. 79). En la mayoría de casos los porcentajes de la población elegible sobrestiman al nivel oficial. Sólo en el caso de la Selva rural se aprecia que la pobreza extrema oficial supera ligeramente a su contraparte estimada. Se esperaba este comportamiento, pues el objetivo de este documento no ha sido estimar el porcentaje de pobres en el país, sino determinar los umbrales o puntos de corte óptimos que permiten conseguir los menores errores de focalización

CUADRO 7. Pobreza oficial y porcentaje de población elegible por ámbito geográfico (en porcentajes)

	Costa urbana	Sierra urbana	Selva urbana	Lima	Costa rural	Sierra rural	Selva rural
Pobreza							
Total	25.1	36.3	40.3	18.5	38.1	73.3	55.3
Extrema	2.1	8.5	11.1	0.5	10.5	40.8	23.4
Población elegible							
Cuando objetivo es pobreza total	26.3	51.6	51.5	21.1	56.6	92.6	82.6
Cuando objetivo es pobreza extrema	0.5	9.5	16.0	0.5	8.3	73.5	16.3

posibles.

Mas aún, estos resultados dependen fuertemente del ponderador θ elegido, pues el hecho de que con un umbral óptimo se obtenga un mayor porcentaje de personas elegibles que con el índice de pobreza oficial parece un reflejo de la mayor sensibilidad al error de exclusión (relativo al error de inclusión) en la función objetivo, caracterizada por $\theta > 1$.

4 CONCLUSIONES

Se propone un algoritmo cuyo objetivo es identificar a los potenciales beneficiarios de los programas sociales. El valor agregado del proyecto ha sido plantear una metodología para determinar los puntos de corte que permitan calificar o identificar a los potenciales beneficiarios de los programas sociales.

El algoritmo consta de dos partes. En la primera se estima un índice de bienestar siguiendo la técnica de análisis de componentes principales. La segunda parte es lo novedoso del documento, pues se propone como umbral o punto de corte a aquel punto que minimiza una función que depende de los errores de focalización.

Para la creación del índice y por la naturaleza de los datos se empleó la técnica de componentes principales para datos cualitativos. Como resultado del empleo de esta metodología se tiene un índice urbano y rural que fue reescalado para que varíe entre 0 y 100. Cabe mencionar que otras estrategias para la determinación del índice, tales como mínimos cuadrados ordinarios, modelos logísticos o regresiones cuantílicas, son extensiones naturales de este trabajo. El cambio en el tipo de modelo tiene implicancias sobre la definición de pobreza que se quiere abordar. En particular, el análisis de componentes principales se ajusta más a la definición de pobreza multidimensional, mientras que los otros métodos mencionados se ajustan más a un enfoque de pobreza monetaria.

Para el caso del punto de corte o umbral, se identificó el punto que minimiza el error de focalización, definido como la suma entre el error de inclusión más dos veces el error de exclusión, ambos expresados como porcentaje de la población. Este ejercicio se realizó por dominio geográfico para tener en cuenta la heterogeneidad de los niveles de pobreza por zonas. Algunas extensiones de este documento deberían considerar formas funcionales alternativas de la función de pérdida a optimizar, analizar la sensibilidad de los resultados ante cambios en el ponderador y la elección del mejor modelo generador del índice de bienestar empleando otras técnicas como el análisis de regresión o modelos de elección binaria.

REFERENCIAS

- Castaño, E. y H. Moreno (1994), “Selección y cuantificación de las variables que conforman el sistema de selección de beneficiarios de programas sociales SISBEN”, Departamento Nacional de Planeación de Colombia, *Planeación y Desarrollo*, 25(edición especial de julio), 259-283.
- Ministerio de Economía y Finanzas - MEF (2010), “Metodología de cálculo del índice de focalización de hogares”, Dirección General de Asuntos Económicos y Sociales.
- Monge, A., E. Vásquez y D. Winkelried (2009), “¿Es el gasto público en programas sociales regresivo en el Perú?”, *Diagnóstico y Propuesta*, 43, Lima: Consorcio de Investigación Económica y Social.
- Wodon, Q. (1997), “Targeting the poor using ROC curves”, *World Development*, 25(12), 2083-2092.
- Young, F. W. (1981), “Quantitative analysis of qualitative data”, *Psychometrika*, 46(4), 357-388.