



BANCO CENTRAL DE RESERVA DEL PERÚ

## **Propuesta metodológica para la focalización individual de los programas sociales**

Jose Valderrama\* y Juan Pichihua\*\*

\* Universidad de Lima

\*\* Universidad Nacional Agraria La Molina

DT. N° 2010-006  
Serie de Documentos de Trabajo  
Working Paper series  
Julio 2010

Los puntos de vista expresados en este documento de trabajo corresponden a los autores y no reflejan necesariamente la posición del Banco Central de Reserva del Perú.

The views expressed in this paper are those of the authors and do not reflect necessarily the position of the Central Reserve Bank of Peru.

# Propuesta metodológica para la focalización individual de los programas sociales\*

José A. Valderrama Torres\*\*  
Juan Pichihua Serna\*\*\*

Mayo de 2010

## Resumen

Se propone una metodología de identificación individual de potenciales beneficiarios de programas sociales. Esta herramienta comprende dos etapas: en la primera se determina un índice de bienestar y en la segunda los puntos de corte que permiten distinguir a los hogares que califican como potenciales beneficiarios. La estimación del índice se basa en una práctica extendida en la literatura especializada: análisis de los componentes principales con escalamiento óptimo. 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.

## 1. Introducción

El objetivo del artículo es proponer una metodología que permita identificar a los hogares prioritarios que deberían ser atendidos por los programas sociales teniendo en cuenta las posibles diferencias existentes entre ámbitos geográficos.

La metodología se divide en dos componentes: en el primero se estima un índice de bienestar y en el segundo se determina a partir de qué puntos un hogar es clasificado como potencial beneficiario. Se estimó un índice de bienestar para cada 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).

La determinación del índice sigue un esquema estándar ampliamente utilizado en la literatura especializada: análisis de componentes principales con escalamiento óptimo. Para la determinación de los umbrales se propone una estrategia endógena con sustento teórico y empírico basado

---

\*Los autores agradecen los comentarios de Julia Johansen, Marcos Robles y de un árbitro anónimo designado por el BCRP a una versión previa 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

\*\*Escuela de Economía de la Universidad de Lima, Lima-Perú. Correo: jvalder@ulima.edu.pe

\*\*\*Facultad de Economía de la Universidad Nacional Agraria La Molina, Lima-Perú. Correo: jpichihua@lamolina.edu.pe

en la minimización de una función que depende de los errores de focalización.

Las estimaciones tuvieron como fuente de datos a la Encuesta Nacional de Hogares (ENAH) correspondiente al año 2007, los resultados han sido validados teniendo en cuenta criterios de consistencia interna y externa. En el primer caso, se ha verificado que las variables que intervienen en la construcción del índice tienen un aporte significativo al mismo y en el segundo se ha verificado que la estimación del índice empleando datos de la ENAH 2008 pero con los parámetros estimados en el 2007 guarda una relación positiva y estadísticamente significativa con el gasto per cápita del 2008.

Además de la introducción, el documento se encuentra dividido en cuatro partes. En la siguiente sección se presenta información de corte transversal sobre el estado de la pobreza en el país, en la tercera parte se detalla la metodología empleada, seguidamente se muestran los resultados y finalmente se presentan las principales conclusiones. Se incluye un anexo de los principales resultados que acompañaron las estimaciones.

## 2. Pobreza por ámbito geográfico

En el Perú, la población que padece de pobreza extrema representó el 14% 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 lo registrado en el Perú urbano<sup>1</sup>. La heterogeneidad en la incidencia de este indicador es aún mayor si se toman en cuenta los dominios geográficos, pues mientras en Lima Metropolitana esta condición afectó al 0.5% de la población, en la sierra rural este mismo indicador fue de 40.8% en la sierra rural (cuadro 1).

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 la figura 1, existe un importante porcentaje de personas que son beneficiados sin ser parte del público objetivo (infiltrados) 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<sup>2</sup>.

Como el presupuesto siempre es limitado, un mecanismo eficiente en la asignación del beneficio es redireccionar los fondos destinados incorrectamente a los infiltrados para orientarlos a los subcubiertos.<sup>3</sup> Este proceso se conoce como focalización, el cual enfrenta varios desafíos que tienen

---

<sup>1</sup>El tamaño del centro poblado define si este es considerado como urbano o rural. El límite que distingue uno del otro es de 2000 personas o 400 viviendas.

<sup>2</sup>La población objetivo en cada programa social fue estimada a partir de la ENAH y de acuerdo con los requisitos que deberían cumplir los potenciales beneficiarios. En el anexo A se detallan estas características.

<sup>3</sup>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.

Cuadro 1: Niveles de pobreza según dominio y ámbito geográfico, 2007

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

Fuente: INEI-ENAH0 2007

Elaboración: Propia

todos ellos como raíz el problema de identificar a los pobres. En el Perú, la determinación de la pobreza se basa en el gasto per cápita obtenido a partir de la Encuesta Nacional de Hogares; si este es menor a unos umbrales conocidos como líneas de pobreza y pobreza extrema, entonces el hogar es clasificado como pobre o pobre extremo, respectivamente. El reto que enfrenta cualquier sistema de focalización es crear un índice de bienestar del hogar, que si bien podría ser una estimación del gasto per cápita<sup>4</sup>, optar por un indicador de este tipo pasa por una decisión previa: ¿Qué tipo de pobreza se quiere medir: monetaria o no monetaria?

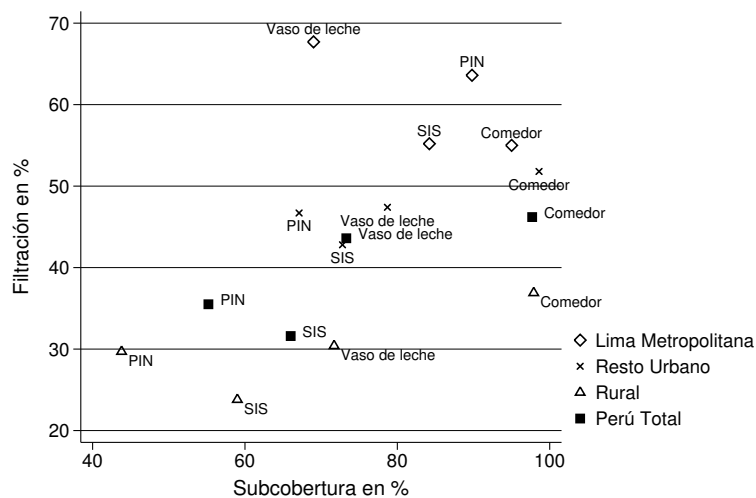
El Sistema de Focalización de Hogares (SISFOH) es el órgano encargado de identificar a los hogares en el Perú. Uno de los insumos primordiales para este propósito es la aplicación presencial de una ficha socioeconómica en cada uno de los hogares, procedimiento que permite identificar con pocas preguntas a los hogares pobres de los no pobres con base en las características de la vivienda, la posesión de activos y características de sus miembros.

A la fecha, el SISFOH ya cuenta con un registro de hogares urbanos a los cuales se ha aplicado la ficha. El siguiente paso entonces es contar con un sistema de identificación de los potenciales beneficiarios, lo cual da origen a la inquietud planteada en este documento. En lo que sigue se propone una metodología en dos etapas. En la primera se estima un índice de bienestar del hogar (IBH) y en la segunda se determina a partir de qué valores de este índice un hogar se puede considerar como candidato para ser beneficiario de un programa social.

---

<sup>4</sup>El cálculo del valor del gasto per cápita es impracticable, entre otras cosas, por el excesivo número de preguntas que requiere construir esta variable.

Figura 1: **Perú 2007. Errores de filtración y subcobertura en programas sociales**



SIS: Sistema Integral de Salud

PIN: Programa Infantil de Nutrición, antes: almuerzos y desayunos escolares.

Fuente: INEI-ENAH0 2007

Elaboración: Propia

### 3. Metodología

La estrategia de identificación de pobres sigue un esquema en dos etapas. En la primera se debe calcular un índice de bienestar, tal como la probabilidad de ser pobre, si se usa un modelo probabilístico; el ingreso proyectado, si se emplea un modelo basado en la minimización de los errores, entre otras opciones metodológicas. 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 un potencial beneficiario.

Para la determinación del índice se seguirá un enfoque multidimensional basado en la técnica de componentes principales. Si bien esta elección es arbitraria, estudios posteriores a este trabajo deberían evaluar otras metodologías enfocadas en la determinación de la pobreza monetaria. La determinación de los umbrales es 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.

#### 3.1. Determinación del índice de bienestar

Cuando el interés es construir un indicador de pobreza que resuma un conjunto de características socioeconómicas de los hogares, el análisis de componentes principales es una herramienta estadística adecuada, pues

por medio de este procedimiento es posible crear un indicador sintético capaz de explicar la mayor parte de la variabilidad del conjunto de variables originales. Sin embargo, muchas de las características observadas de los hogares son variables del tipo cualitativo, 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).

Un camino para esta valorización es el que harían expertos asignando un valor 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 variables. Para salvar técnicamente este inconveniente se usa el procedimiento denominado “escalamiento óptimo” (Young, 1981), el cual permite asignar valores numéricos a las categorías de las variables de forma tal que se maximizan las correlaciones entre todas las variables de interés; ello es consistente con el procedimiento estándar de componentes principales, pues este busca resumir información de un conjunto de variables correlacionadas y este es precisamente el caso al que nos enfrentamos, ya que los hogares pobres tienden a tener valores bajos en todas las características medidas (viviendas precarias, bajo nivel educativo, acceso limitado a servicios básicos, etc.), y por ende, un valor también bajo en el indicador final de nivel de pobreza, siendo esto una razón más para el empleo de esta metodología<sup>5</sup>.

Como el interés es contar con un índice de pobreza, se restringió el número de dimensiones a uno y se clasificó a todas las variables como ordinales, pues esta es la característica de las variables empleadas, donde las de menor orden indican situaciones desfavorables para el hogar.

Las variables tomadas en cuenta fueron aquellas que más se correlacionan con los niveles de pobreza, para lo cual se empleó el estadístico

---

<sup>5</sup>Cabe mencionar que dentro del enfoque no monetario otra metodología tradicionalmente empleada para identificar pobres es el de las Necesidades Básicas Insatisfechas (NBI). Esta metodología fue propuesta por la CEPAL a comienzos de la década del 90 y consiste en el empleo de cinco indicadores que miden distintos aspectos de la calidad de vida. Este enfoque se diferencia de lo propuesto en por lo menos tres características, que lo hacen poco atractivo para nuestros fines:

- 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 óptimamente para tomar en cuenta la volatilidad de la variable.
- No permite hacer un ordenamiento extenso de hogares de tal forma que se detecte a los hogares más pobres, sólo permite la clasificación de los hogares en seis grupos. El enfoque de componentes principales permite crear un índice de pobreza, y con ello se puede hacer un *ranking* extenso de hogares según su nivel de pobreza.
- La selección de variables que intervienen en el cálculo del indicador es arbitraria. Esto lo diferencia del método propuesto, que sólo tomó en cuenta a aquellas variables que tengan un aporte significativo a la hora de identificar hogares pobres.
- Los indicadores sugeridos hace cerca de veinte años han perdido vigencia, tanto así que el INEI está promoviendo el empleo de otros indicadores.

de Sommers. Las que clasifican como ordinales 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” (anexo B1). Con las variables elegidas se estimó el índice, que al ser una combinación lineal de los variables elegidas, también tiene la misma interpretación: menores valores indican que el hogar se encuentra en condiciones más desfavorables. Finalmente, se realizó un análisis ex-post, correlacionando los quintiles del índice estimado con todas las variables empleadas (anexo B2). Por la forma en que han sido recodificadas

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%, y se empleó el índice de consistencia interna de Cronbach<sup>6</sup> (anexos B3 y B4). Para el ejercicio de validación externa se generó el IBH y se halló la correlación de este índice con el logaritmo del gasto per cápita. En los anexos B5 y B6 se muestra la correlación haciendo la distinción entre el ámbito urbano y rural. Aunque moderada, la correlación es positiva y estadísticamente significativa<sup>7</sup>.

### 3.2. Determinación del umbral

Con la propuesta de índice el siguiente paso es determinar a partir de qué nivel o umbral un hogar puede ser clasificado como potencial beneficiario. Dependiendo del enfoque. Los enfoques se pueden agrupar en un esquema relativo y uno absoluto. El primero apela al uso de indicadores de posición como quintiles o deciles (v.g. Son potenciales beneficiarios aquellos hogares que se encuentran en los dos primeros quintiles de la distribución) mientras que el enfoque absoluto una alternativa es emplear análisis de conglomerados<sup>8</sup>.

En este trabajo se optó por un camino alternativo inspirado 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 el procedimiento planteado.

La matriz de clasificación según error de inclusión y exclusión es la base de la propuesta de tipificación. Por convención se define “infiltración” a las personas que acceden a un programa social y no pertenecen a la población objetivo y “subcobertura” al porcentaje de la población objetivo que no recibe la atención de un programa social pese a ser elegible. En la figura 2-a se ilustran estos conceptos.

A diferencia de los conceptos antes mencionados, los 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. El

---

<sup>6</sup>Indicador que mide el grado común que tienen todas las variables. Varía entre 0 y 1, valores mayores a 80% son considerados como muy buenos.

<sup>7</sup>En estricto rigor esta validación requiere de una predicción fuera de muestra. En encuestas de diseño complejo como la ENAHO, este procedimiento no es trivial, pues extraer una submuestra provoca cambios que exigen un ajuste en los factores de expansión y en algunos casos también del tamaño de los conglomerados

<sup>8</sup>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.



error de inclusión se define como las personas que incorrectamente son clasificadas como potenciales beneficiarios, mientras que los errores de exclusión comprenden a aquellas que incorrectamente son excluidas cuando en realidad califican como parte de la población objetivo; ambos expresados como porcentaje de la población total. Esta clasificación es posible debido a que se tiene un índice de bienestar del hogar, de tal forma que se pueden ordenar a 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 (ver panel derecho de la figura 2). Precisamente la determinación de la ubicación de este nivel es parte del objetivo del estudio.

Figura 2: Errores de focalización ex-post y ex-ante

		¿Es beneficiario?	
		Si	No
¿Es pobre?	Si		<i>Subcobertura</i>
	No	<i>Filtración</i>	

		¿Potencial beneficiario?	
		Si	No
¿Es pobre?	Si		<i>Exclusión</i>
	No	<i>Inclusión</i>	

De la figura 2-b es posible verificar que cuando el umbral toma el menor valor posible, el error de inclusión desaparece mientras que el error de exclusión toma su mayor valor. En el otro extremo, cuando el umbral tome el máximo valor posible ocurre todo lo contrario: la inclusión aumentará a su máximo nivel y la exclusión será cero, con lo cual se muestra la disyuntiva entre ambos tipos de errores al momento de elegir el nivel del umbral: aumentarlo provocaría una disminución en el error de exclusión y un aumento en el error de inclusión, y lo inverso cuando disminuye este valor.

Además de la característica mencionada, es fácil deducir que las curvas tomarían formas similares a las de una distribución acumulada, como se muestra en la figura 3, pues a medida que se acercan a los extremos el número de personas que se van incluyendo o excluyendo es cada vez menor hasta alcanzar un límite superior o inferior.

Por construcción, estas curvas de concentración no tienen mínimos, por lo que buscar un umbral que minimice alguno de los errores sería una tarea infructuosa. Sin embargo, como se mostrará a continuación, 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; esta característica de la variable impide que se trabaje con las tasas de infiltración y subcobertura convencionales, pues estos se encuentran expresados 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. Alternativamente, se emplearon los errores de

inclusión y exclusión ya definidos, que no es otra cosa que la población afectada dividida por el total de la población.

Sea  $I$  el error de inclusión y  $E$ , el error de exclusión; donde  $T$ , es la suma de ambos tipos de errores y  $U$ , el argumento del cual dependen. Entonces:

$$T(U) = I(U) + E(U) \quad (1)$$

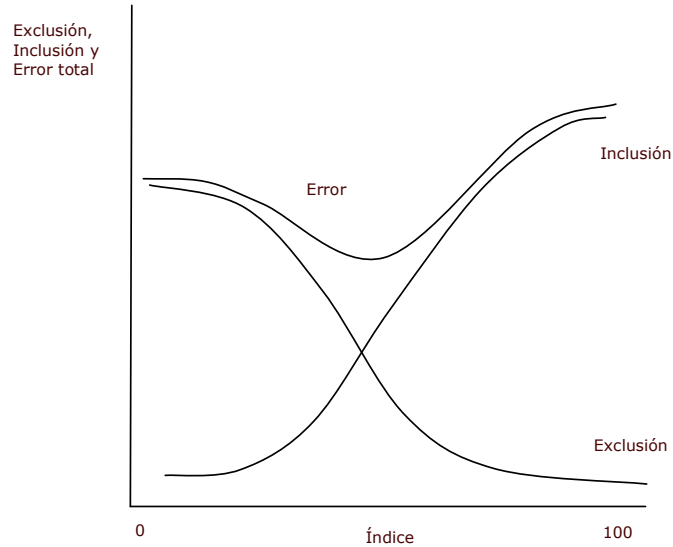
La existencia de un mínimo ( $U^*$ ) exige que se cumplan las condiciones de primer y segundo orden:

$$CPO: \frac{\partial T}{\partial U} = \frac{\partial I}{\partial U} + \frac{\partial E}{\partial U} = 0 \quad (2)$$

$$C2O: \frac{\partial^2 T}{\partial U^2} = \frac{\partial^2 I}{\partial U^2} + \frac{\partial^2 E}{\partial U^2} > 0 \quad (3)$$

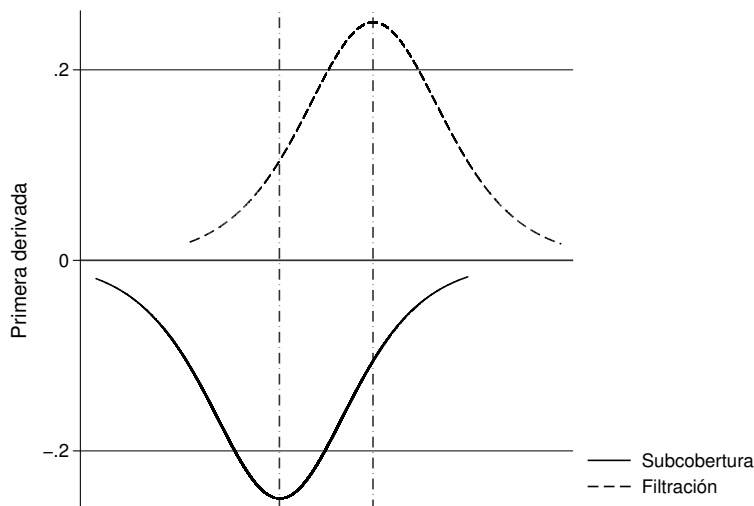
Una condición suficiente para que exista mínimo es que cada una de las segundas derivadas sea definida positiva simultáneamente<sup>9</sup>, si ese es el caso, la función de error, que es resultado de la suma de ambos tipos de errores, toma la forma convexa mostrada en la figura 3, donde el punto mínimo de la función  $T$  estaría ubicado en algún punto ubicado en el área encerrada por el intervalo determinado por las líneas verticales de la figura 4, pues en esa área ambas pendientes (segundas derivadas de las funciones originales) son positivas.

Figura 3: Formas hipotéticas de las curvas de inclusión, exclusión y error de focalización total



<sup>9</sup>Naturalmente existen otras posibilidades. Por ejemplo, una de las segundas derivadas puede ser negativa y la otra positiva y lo suficientemente alta como para superar a su par.

Figura 4: **Región en la que se garantiza la presencia de un mínimo en la función de error de focalización total**



Una generalización a lo planteado en la expresión (1) sería modificar la función de pérdida anterior para darle una ponderación distinta a cada tipo de error:

$$T = I + \alpha E \quad (4)$$

La elección del parámetro  $\alpha$  es una variable de política, toma un valor positivo con distintas implicancias dependiendo de si este es menor o mayor a uno:

- Si  $\alpha < 1$ , se le da más importancia al error de inclusión.
- Si  $\alpha > 1$ , se le da más importancia al error de exclusión.
- Si  $\alpha = 1$ , ambos errores tienen la misma importancia.

En este documento, a manera de aplicación, se usó como ponderador al número 2. La implicancia de lo anterior es que en el punto que se elija como umbral es aquel donde se permite que en el límite o margen se filtren dos personas, con tal de que la subcobertura disminuya en una persona. Matemáticamente, ello se deduce de la condición de primer orden:

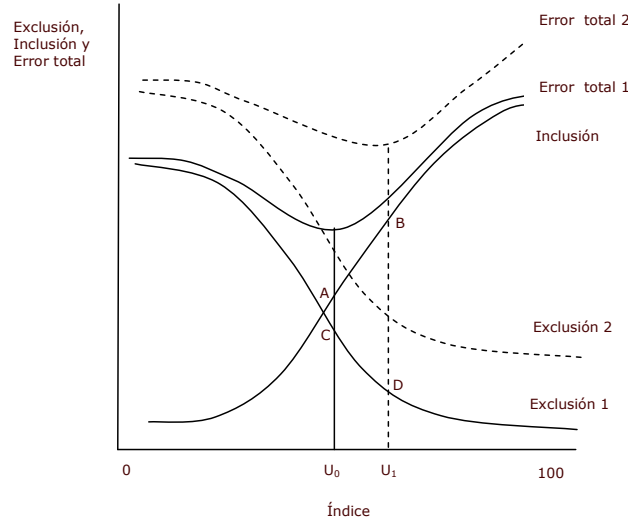
$$T = I + 2 * E \Rightarrow CPO : \frac{\partial I / \partial U}{\partial E / \partial U} = -\frac{2}{1} \quad (5)$$

Al ser una elección arbitraria, extensiones de esta investigación podrían preocuparse por los efectos de la elección de este parámetro sobre los errores de focalización, tratando tal vez de encontrar un valor óptimo para este parámetro.

Gráficamente, una mayor ponderación al error de exclusión significaría un desplazamiento hacia la derecha del umbral que determina el error mínimo posible que se puede cometer; con lo que el error de inclusión

aumentaría pasando del punto A al punto B y la exclusión se reduciría al pasar del punto C al punto D (ver figura 5). La sensibilidad de los cambios dependerá de la elasticidad de las curvas respecto al valor del umbral. Mientras más elásticas sean estas, mayor será el impacto sobre los errores.

Figura 5: **Desplazamiento de las curvas debido a una mayor ponderación del error de exclusión**



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 los indicadores de sensibilidad (SE) y especificidad (SP)<sup>10</sup>:  $W(SE, 1 - SP)$  con  $\partial W/\partial SE$  positivo,  $\partial W/\partial(1 - SP)$  negativo,  $\partial^2 W/\partial SE^2$  negativo, y  $\partial^2 W/\partial(1 - SP)^2$  positivo. Si la sociedad maximiza la función de bienestar sujeto a la restricción impuesta por la curva ROC<sup>11</sup> representada por  $g_x(1 - SP) = SE$ :

$$\text{Max } W(SE, 1 - SP) \text{ sujeto a } g_x(1 - SP) = SE \quad (6)$$

Siendo la condición de primer orden que soluciona el problema:

$$\frac{\partial g_x}{\partial(1 - SP)} = -\frac{\partial W/\partial(1 - SP)}{\partial W/\partial SE} \quad (7)$$

La discusión planteada por Wodon asume la existencia de un índice de pobreza. En su versión logística, dicho índice sería la probabilidad de ser pobre, y el umbral o punto de corte antes aludido, el punto a partir del cual un hogar es clasificado como pobre. En esta línea es que el autor

<sup>10</sup>Ambos indicadores forman parte de la terminología de las curvas ROC. El primero es el porcentaje de pobres identificados correctamente y el segundo es el porcentaje de no pobres identificados correctamente.

<sup>11</sup>La curva ROC establece la relación entre  $SE$  y  $1 - SP$ . Se genera al modificar el punto de corte o umbral que permite predecir si un hogar es pobre o no. Si el umbral tiende a cero todas las personas son clasificadas como pobres con lo cual  $SE = 1$  y al mismo tiempo  $SP = 0$ .

plantea elegir a partir de un *set* de modelos, aquél que consiga el máximo nivel de bienestar. Es decir, si se tiene el conjunto  $G = g_{x1}, \dots, g_{xn}$ , es posible hallar el punto de corte óptimo para cada especificación o modelo, calcular la función de utilidad indirecta y escoger el modelo que permita obtener el mayor nivel de bienestar. Entonces, el mejor modelo sería:

$$g_{x_i}^* = \text{argMax}[v^*(g_{x1}), \dots, v^*(g_{xN})] \quad (8)$$

La metodología propuesta se diferencia de la de Wodon en por lo menos tres aspectos:

- La función de optimización es una función de error o pérdida, por lo tanto el problema es de minimización. Si bien se optó por un tipo particular de función (la suma ponderada de los dos tipos de errores) esta puede ser generalizada en estudios posteriores. Más aún, en su versión simple ( $\alpha = 1$ ), la forma funcional es atractiva pues la suma tiene una interpretación natural: personas incorrectamente identificadas.
- Los argumentos de la función son distintos. Al emplear los indicadores de sensibilidad y sensibilidad, 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 logran identificar correctamente a 100 y 10 pobres y no pobres, respectivamente; ambos indicadores de sensibilidad y sensibilidad serían 10%. Con los mismos datos, la metodología propuesta 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<sup>12</sup>.

En la línea de lo sugerido por Wodon, futuras extensiones de esta investigación deberían incorporar un criterio que permita discriminar cuál es el mejor modelo generador del índice de bienestar, tal que se obtengan los menores errores de focalización. Así, de un *set* de modelos candidatos, se puede elegir el punto de corte que minimice la función de pérdida en cada uno de los modelos planteados y quedarse con aquel modelo que presente la menor pérdida. Bajo esta extensión, el mejor modelo sería aquel que soluciona el siguiente problema:

$$g_{x_i}^* = \text{argMin}[v^*(g_{x1}), \dots, v^*(g_{xN})] \quad (9)$$

## 4. Resultados

### 4.1. Determinación del índice de bienestar

Las variables seleccionadas para formar parte del índice de bienestar fueron motivadas por el grado de correlación entre la variable en cuestión y el nivel de pobreza (pobreza extrema, pobreza no extrema, no pobre). En este sentido, a partir de un conjunto inicial de alrededor de 40 variables

---

<sup>12</sup>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 monotónica.

Cuadro 2: Variables seleccionadas según ámbito geográfico

Variable	Urbano	Rural
Hacinamiento	x	x
Nivel educativo del jefe de hogar	x	
Combustible empleado para cocinar	x	
Tenencia de teléfono fijo	x	
Tenencia de cocina a gas	x	x
Tenencia de refrigeradora/congeladora	x	
Tenencia de TV a color	x	x
Tenencia de plancha eléctrica	x	x
Tenencia de equipo de música	x	
Bienes identificadores de riqueza 1/	x	
Número de artefactos en el hogar 2/	x	x
Material que predomina en las paredes	x	
Material que predomina en los techos	x	
Material que predomina en los pisos	x	x
Abastecimiento de agua en el hogar	x	
Tipo de servicio higiénico	x	
Afiliados a algún seguro de salud menos SIS	x	x
Tasa de ocupados 3/		x
Menores de 6 años		x
Mujeres adultas analfabetas		x
Vivienda con piso de tierra en la sierra		x

Notas:

1/ Se consideraron la tenencia de computadora, refrigeradora, cable, internet y lavadora

2/ Los artefactos considerados son: TV a color, equipo de sonido, DVD, computadora, plancha eléctrica, licuadora, cocina a gas, refrigeradora/congeladora, lavadora y microondas

3/ El numerador considera a los mayores de 18 años ocupados y el denominador al total de adultos

se seleccionaron las variables mostradas en el cuadro 2, donde además se hace la distinción entre ámbitos geográficos<sup>13</sup>.

En el cuadro 3 se muestran las puntuaciones empleadas para la construcción del índice de bienestar. Estas se encuentran reescaladas, de tal modo que se garantiza que el índice varíe entre 0 y 100.

#### 4.2. Determinación de los umbrales

Siguiendo el desarrollo teórico planteado en la sección 3.2, se estimaron los umbrales de tal forma que se minimiza 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<sup>14</sup>.

<sup>13</sup>En el anexo B se detalla la metodología seguida para la especificación del modelo.

<sup>14</sup>En el anexo C se encuentran las curvas que dan origen a los umbrales.

**Cuadro 3: Puntuaciones empleadas para la estimación del índice de bienestar del hogar**

<b>Variable y categorías</b>	<b>Urbano</b>	<b>Rural</b>	<b>Variable y categorías</b>	<b>Urbano</b>	<b>Rural</b>
<b>Hacinamiento</b>			<b>Material predominante en los techos es:</b>		
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 o caña	1.603	-
Menos de 1	5.561	4.728	Tejas	1.603	-
<b>Educación del jefe</b>			Madera	3.231	-
Sin nivel	0.434	-	Concreto	5.668	-
Inicial	1.051	-	<b>Abastecimiento de agua en su hogar procede de :</b>		
Primaria	1.909	-	Otro	-0.308	-
Secundaria	3.324	-	Rio	-0.308	-
Técnica	4.865	-	Pozo	0.391	-
Universitario	6.256	-	Camión	0.833	-
Post Grado	7.478	-	Pilón	0.833	-
<b>Tipo de combustible para cocinar</b>			Fuera	1.499	-
No cocinan	-0.719	-	Dentro	4.409	-
Otro	-0.719	-	<b>Servicio higiénico que tiene su hogar esta conectado a :</b>		
Leña	-0.719	-	No tiene	-1.231	-
Carbón	0.904	-	Rio	0.178	-
Kerosene	1.100	-	Pozo ciego	0.570	-
Gas	4.736	-	Pozo séptico	0.570	-
Electricidad	6.262	-	Fuera	1.480	-
<b>Teléfono fijo</b>			Dentro	4.829	-
No	1.637	-	<b>Número de artefactos</b>		
Si	5.887	-	Ninguno	-1.360	-0.711
<b>Cocina a gas</b>			Sólo 1	-1.360	2.447
No	-0.392	0.314	Sólo 2	-0.276	4.932
Si	4.501	9.487	Sólo 3	1.042	7.387
<b>Refrigeradora/Congeladora</b>			Sólo 4	2.164	10.386
No	0.858	-	Sólo 5	3.626	12.801
Si	5.944	-	Sólo 6	4.809	14.935
<b>TV a color</b>			Sólo 7	6.052	16.140
No	-0.407	0.215	Sólo 8	7.258	17.685
Si	4.673	9.655	Sólo 9	8.145	17.685
<b>Plancha eléctrica</b>			Sólo 10	8.145	17.685
No	-0.237	0.669	<b>Afiliados a algún SS</b>		
Si	4.870	11.898	Ninguno	1.566	1.363
<b>Equipo de música</b>			Uno	3.751	5.911
No	1.776	-	Dos	5.107	9.827
Si	5.541	-	Tres	5.897	12.423
<b>Bienes identificadores de riqueza</b>			Más de 3	6.057	12.423
Ninguno	0.527	-	<b>Tasa de ocupados</b>		
1	4.632	-	No ocupados	-	0.973
2	6.714	-	De 0 a 25%	-	0.973
3	7.895	-	De 25% a 50%	-	1.886
4	7.895	-	De 50% a 75%	-	3.339
<b>El material predominante en las paredes exteriores es :</b>			De 75% a 100%	-	3.339
Madera o estera	0.043	-	<b>Menores de 6 años</b>		
Piedra con barro	0.043	-	Más de 2	-	-0.672
Quincha	0.194	-	Sólo 2	-	0.552
Tapia	0.194	-	Sólo 1	-	2.483
Adobe	0.692	-	Ninguno	-	3.210
Piedra o sillar con cal o cemento	1.098	-	<b>Mujeres adultas analfabetas</b>		
Ladrillo o bloque de cemento	5.075	-	Más de 1	-	-0.446
<b>el material predominante en los pisos es :</b>			Sólo 1	-	0.241
Otro	-0.361	-0.329	No hay	-	3.534
Tierra	-0.361	0.672	<b>Piso de tierra en la siera</b>		
Cemento	3.770	7.999	No	-	-0.316
Madera	3.770	7.999	Si	-	3.623
Losetas	6.768	17.375			
Láminas	6.948	18.459			
Parquet	7.204	20.419			

**Cuadro 4: Determinación de los umbrales urbanos que minimizan el error de focalización 1/**

Ámb. Geográf.	Nivel de pobreza 2/	Tipo de error			Umbral	Interv. de Conf. al 95% 3/	
		Inclusión	Exclusión	Total		Lím. Inf.	Lím. Sup.
Costa	PE	0.3%	1.7%	3.8%	8.2	8.0	8.4
Costa	PT	10.8%	9.7%	30.1%	43.1	42.5	43.8
Sierra	PE	4.3%	4.0%	12.3%	18.4	18.1	18.7
Sierra	PT	19.0%	4.6%	28.2%	53.1	52.6	53.7
Selva	PE	9.0%	4.6%	18.2%	18.4	18.1	18.7
Selva	PT	18.8%	7.6%	34.1%	46.1	45.5	46.6
LM	PE	0.3%	0.4%	1.1%	13.8	13.4	14.2
LM	PT	11.0%	8.4%	27.9%	50.7	50.4	51.0

1/ Se obtuvieron dos umbrales por cada dominio. El de menor valor se obtuvo al asumir que la población objetivo son únicamente los pobres extremos y el de mayor valor cuando la población objetivo es el total de pobres.

2/ PE: Personas que deben ser atendidas para minimizar el error de identificación cuando el público objetivo son los pobres extremos. PNE: Personas que deben ser atendidas para minimizar el error de identificación cuando el público objetivo son los pobres extremos.

3/ Bootstrap con 200 réplicas asumiendo normalidad en la distribución del umbral. El remuestreo fue sobre la ENAHO 2007, en cada una de las 200 réplicas se generó un umbral.

**Cuadro 5: Determinación de los umbrales rurales que minimizan el error de focalización 1/**

Ámb. Geográf.	Nivel de pobreza 2/	Tipo de error			Umbral	Interv. de Conf. al 95% 3/	
		Inclusión	Exclusión	Total		Lím. Inf.	Lím. Sup.
Costa	PE	3.4%	6.7%	16.7%	12.4	12.2	12.6
Costa	PT	23.6%	5.3%	34.2%	34.8	33.9	35.6
Sierra	PE	36.0%	4.2%	44.4%	23.3	22.7	23.9
Sierra	PT	20.5%	36.0%	92.5%	52.2	51.9	52.6
Selva	PE	7.4%	15.1%	37.6%	11.1	11.0	11.3
Selva	PT	30.4%	3.6%	37.5%	35.8	34.9	36.6

1/ Se obtuvieron dos umbrales por cada dominio. El de menor valor se obtuvo al asumir que la población objetivo son únicamente los pobres extremos y el de mayor valor cuando la población objetivo es el total de pobres.

2/ PE: Personas que deben ser atendidas para minimizar el error de identificación cuando el público objetivo son los pobres extremos. PNE: Personas que deben ser atendidas para minimizar el error de identificación cuando el público objetivo son los pobres extremos.

3/ Bootstrap con 200 réplicas asumiendo normalidad en la distribución del umbral. El remuestreo fue sobre la ENAHO 2007, en cada una de las 200 réplicas se generó un umbral.

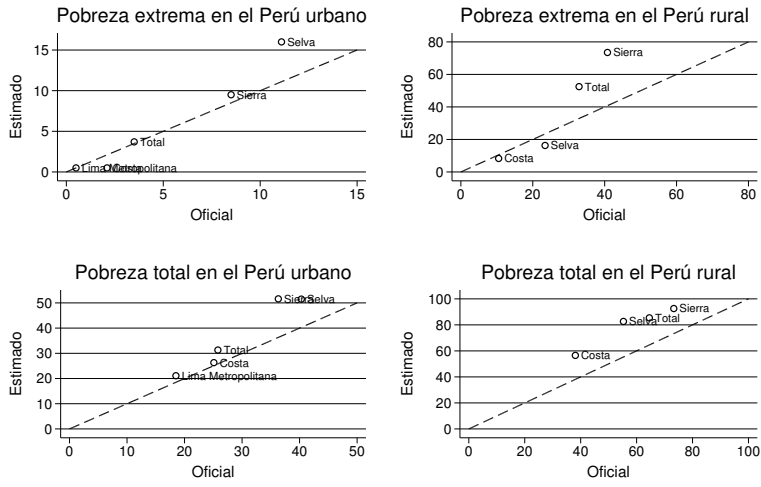
La variabilidad muestral fue tomada en cuenta a partir de la construcción de los intervalos de confianza (últimas dos columnas de los cuadros 4 y 5).

### 4.3. Pobreza estimada versus pobreza oficial

En la figura 6 se comparan los indicadores de pobreza extrema y pobreza total que se derivan de la metodología sugerida y los niveles oficiales. Los puntos ubicados en la línea punteada representan coincidencia perfecta entre ambos criterios, mientras que valores por encima de la línea señalarían que los niveles predicho son mayores a los oficiales. En general, los niveles de pobreza estimados en la mayoría de los casos 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.



Figura 6: Niveles de pobreza estimados y nivel oficial (en porcentajes)



## 5. Discusión y 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 de tipo de modelo tiene implicancias sobre la definición de pobreza que se quiere abordar: El análisis de componentes principales se ajustan 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 otras funciones de pérdida a optimizar como por ejemplo una cuadrática:  $T = I^2 + \alpha E^2$ , 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

- [1] Castaño, Elkin; Moreno, Hernando, *Selección y cuantificación de las variables que conforman el sistema de selección de beneficiarios de programas sociales (SISBEN)*, Misión de apoyo a la descentralización y la focalización de los servicios sociales, (1994)
- [2] Castaño, Deutsch y Véñez, *An economic interpretation of Colombia's SISBEN: A composite Welfare Index*, Manuscrito, (1998)
- [3] Madueño, Miguel, Javier Linares, y Alessandra Zurita, *Instrumento estandarizado de identificación de beneficiarios para programas sociales en el Perú*. Bethesda, MD: Socios para la reforma del sector salud, Abt Associates Inc., (2004)
- [4] Madueño, Miguel, *Revisión del algoritmo de focalización individual del SISFOH*. PRAES. Manuscrito, (2007)
- [5] MEF-DGAES, *Informe de evaluación de los instrumentos para la focalización individual utilizados por el SISFOH con recomendaciones para perfeccionarlo*. Manuscrito, (2007)
- [6] Young, *Quantitative Analysis of Qualitative Data*. *Psychometrika*, 46, 357-388, (1981)
- [7] Wodon, Quentin, *Targeting the Poor Using ROC Curves*. *World Development*, 25, 2083-2092, (1997)

## Anexos

### Anexo A: Población objetivo de programas sociales seleccionados

Los requisitos de acceso a un programa social varían dependiendo del objetivo del mismo, aunque todos tienen en común que la población a ser atendida necesariamente tiene que ser pobre. Así por ejemplo, en el caso del vaso de leche, los beneficiarios tienen que pertenecer a hogares pobres y además contar con alguna de las siguientes características: en primera prioridad a los niños menores de 6 años, madres gestantes y lactantes y clasificados como segunda prioridad a los niños de 7 a 13 años junto a los adultos mayores de 65 años. En la figura 1 se señalan los públicos objetivos de otros tres programas sociales simbólicos.

Figura 7: Programas sociales seleccionados y población objetivo

Programa	Norma	Población objetivo
Vaso de Leche	Ley N° 27470 (Ley que establece Normas Complementarias para la Ejecución del Programa del Vaso de Leche)	Los niños menores de 06 años, madres gestantes y lactantes en situación de pobreza en primera prioridad Los niños entre los 7 y 13 años y los adultos mayores de 65 años en situación de pobreza en segunda prioridad. Se considera a los hogares pobres
Comedores Populares	De acuerdo al PRONAA la Atención a Comedores busca mejorar las condiciones de acceso a la alimentación de personas de bajos recursos, constituyéndose los comedores como agentes de cambio, con carácter comunitario, desempeño dinámico y desarrollo organizacional.	Población pobre
Desayunos y almuerzos escolares	De acuerdo al PRONAA tienen como objetivo contribuir a mejorar el nivel nutricional de la población en edad escolar, prioritariamente en zonas con alta tasa de desnutrición infantil y situación de extrema pobreza, lo que conlleva a mejorar su capacidad de aprendizaje y asistencia a la escuela.	Niños de zonas pobres entre 5 y 14 años que asisten a escuela pública.
Seguro Integral de Salud	Normas SIS.	Población pobre sin seguro de salud

### Anexo B: Pasos empleados en la especificación de los modelos

#### 1. Selección Variables iniciales

Sobre la base de la correlación de las variables con los niveles de pobreza el test de Sommers permite elegir aquellas que identifican mejor la pobreza. En el algoritmo propuesto las variables fueron:

**Figura B1:** Asociación entre niveles de pobreza y variables seleccionadas

URBANO		RURAL	
VARIABLES	Sommers	VARIABLES	Sommers
Hacinamiento	0.403	Número de artefactos(a)	0.317
Número de artefactos	0.491	Hacinamiento	0.281
Bienes identificadores de riqueza	0.459	Menores de 6 años(a)	0.280
el material predominante en los pisos es :	0.408	Cocina a gas	0.211
Educación del jefe	0.346	TV a color	0.194
Afiliados a algún SS(a)	0.341	Tasa de ocupados(a)	0.190
Teléfono fijo	0.334	el material predominante en los pisos es :(a)	0.180
el material predominante en los techos es:	0.326	Plancha eléctrica	0.171
Refrigeradora/Congeladora	0.322	Mujeres adultas analfabetas(a)	0.167
El material predominante en las paredes exteriores es :	0.283	Piso de tierra en la sierra	0.157
Tipo de combustible para cocinar	0.265	Afiliados a algún SS(a)	0.112
el servicio higiénico que tiene su hogar esta conectado a :	0.242		
Equipo de música	0.234		
Plancha eléctrica	0.220		
TV a color	0.190		
el abastecimiento de agua en su hogar procede de :	0.161		
Cocina a gas	0.146		

## 2. Análisis Ex-Post

**Figura B2:** Asociación entre los quintiles del índice de bienestar y variables seleccionadas (Sommers)

URBANO		RURAL	
VARIABLES	Sommers	Variable	Sommers
Hacinamiento	0.403	Número de artefactos	0.317
Educación del jefe	0.346	Hacinamiento	0.281
Tipo de combustible para cocinar	0.265	Menores de 6 años	0.280
Teléfono fijo	0.334	Cocina a gas	0.211
Cocina a gas	0.146	TV a color	0.194
Refrigeradora/Congeladora	0.322	Tasa de ocupados	0.190
TV a color	0.190	el material predominante en los pisos	0.180
Plancha eléctrica	0.220	Plancha eléctrica	0.171
Equipo de música	0.234	Mujeres adultas analfabetas	0.167
Bienes identificadores de riqueza	0.459	Piso de tierra en la sierra	0.157
Material predominante en las paredes exteriores	0.283	Afiliados a algún seguro social	0.112
Material predominante en los pisos es :	0.408		
Material predominante en los techos es:	0.326		
Abastecimiento de agua	0.161		
Servicio higiénico	0.242		
Número de artefactos	0.491		
Afiliados a algún SS(a)	0.341		

3. Validación interna

**Figura B3:** Perú urbano, Varianza explicada, aporte a la varianza e indicador Cronbach

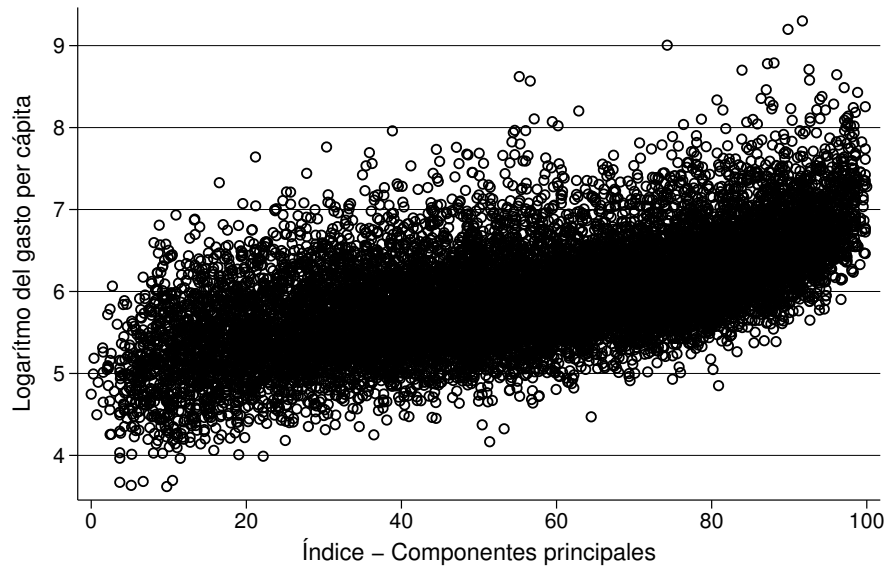
	Total	Aporte a la varianza explicada
Número de artefactos	0.799	11.7%
Bienes identificadores de riqueza	0.686	10.0%
Refrigeradora/Congeladora	0.558	8.1%
el material predominante en los pisos es :	0.464	6.8%
Plancha eléctrica	0.429	6.3%
Tipo de combustible para cocinar	0.394	5.7%
Teléfono fijo	0.388	5.7%
TV a color	0.380	5.5%
Material predominante en los techos es:	0.380	5.5%
el servicio higiénico que tiene su hogar esta conectado a :	0.368	5.4%
el material predominante en las paredes exteriores es :	0.359	5.2%
Afiliados a algún SS	0.337	4.9%
Cocina a gas	0.320	4.7%
Equipo de música	0.306	4.5%
Personas/cuarto	0.236	3.4%
Educación del jefe	0.228	3.3%
Abastecimiento de agua en su hogar procede de :	0.220	3.2%
<b>Varianza explicada</b>	<b>6.849</b>	<b>100.0%</b>
Varianza total	10.151	
Aporte a la varianza total	67%	
Alfa de Cronbach	91%	

**Figura B4:** Perú rural, Varianza explicada, aporte a la varianza e indicador Cronbach

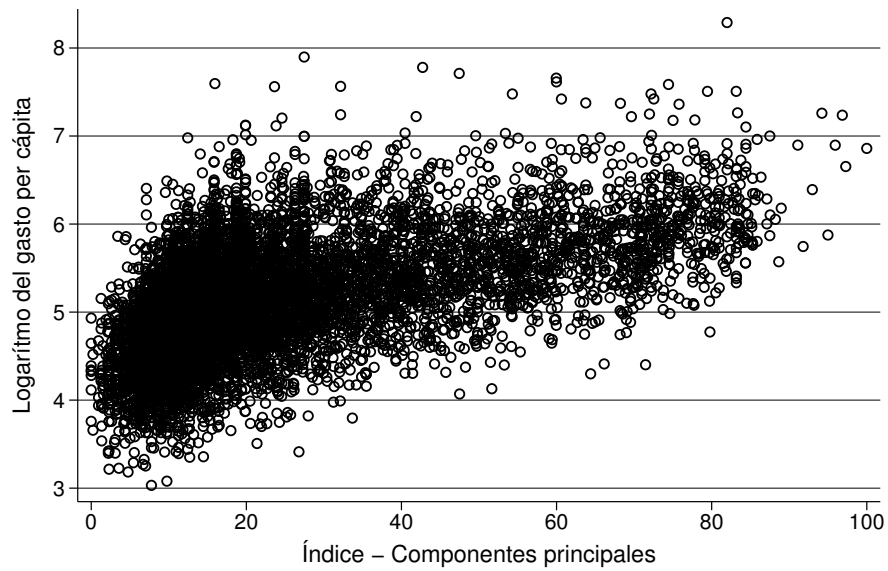
	<b>Total</b>	<b>Aporte a la varianza explicada</b>
Número de artefactos	0.859	24%
TV a color	0.578	16%
Plancha eléctrica	0.574	16%
Cocina a gas	0.536	15%
Pisos	0.378	10%
Afiliados a algún SS	0.249	7%
Piso de tierra en la sierra	0.145	4%
Personas/cuarto	0.123	3%
Mujeres adultas analfabetas	0.110	3%
Menores de 6 años	0.069	2%
Tasa de ocupados	0.027	1%
Varianza explicada	3.649	100%
Varianza total	7.351	
Porcentaje de la varianza explicada	50%	
Alfa de Cronbach	80%	

4. Validación externa

**Figura B5:** Perú urbano 2007. Asociación entre el gasto per cápita y el índice de bienestar



**Figura B6:** Perú rural 2007. Asociación entre el gasto per cápita y el índice de bienestar





## Anexo C: Representación gráfica de la determinación de los umbrales

Figura C1: Costa Urbana (No incluye Lima Metropolitana)

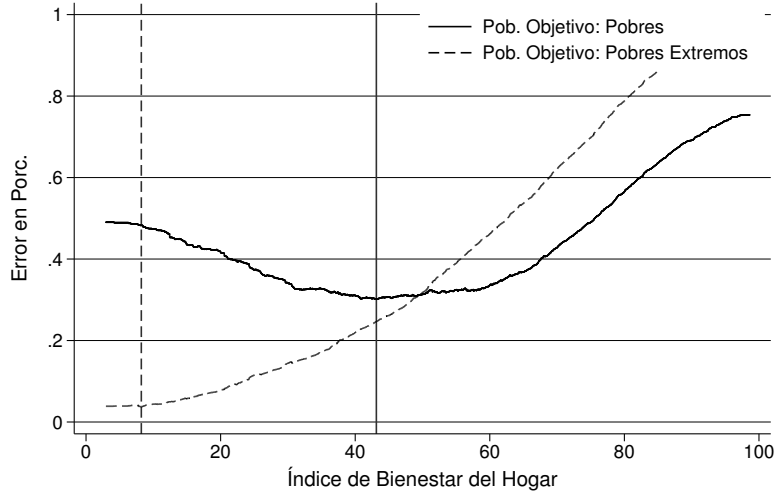
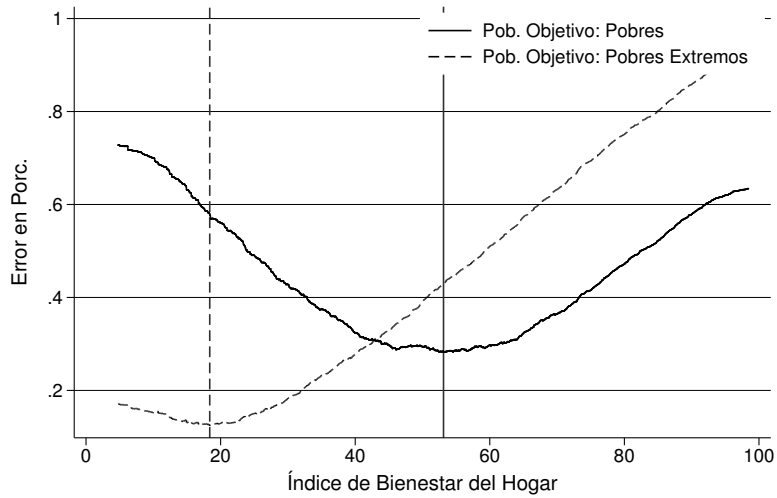
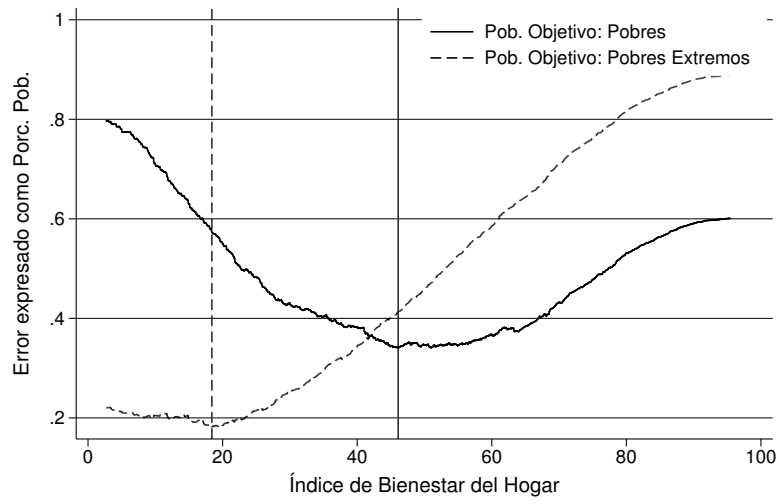


Figura C2: Sierra Urbana



**Figura C3: Selva Urbana**



**Figura C4: Lima Metropolitana**

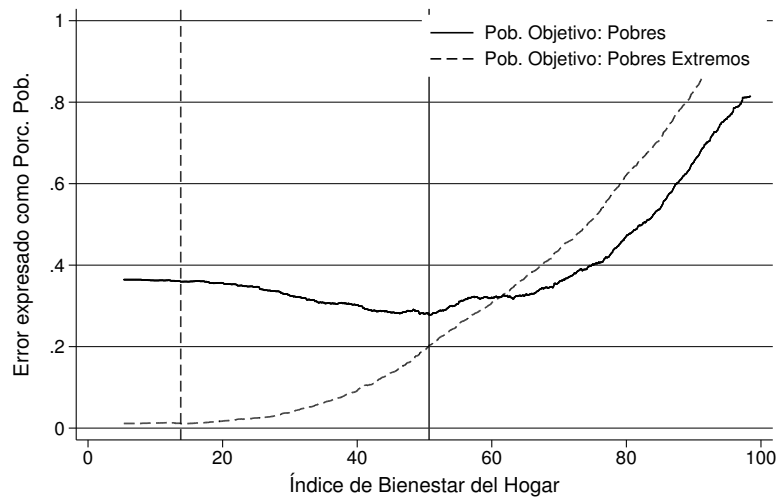


Figura C5: Costa rural

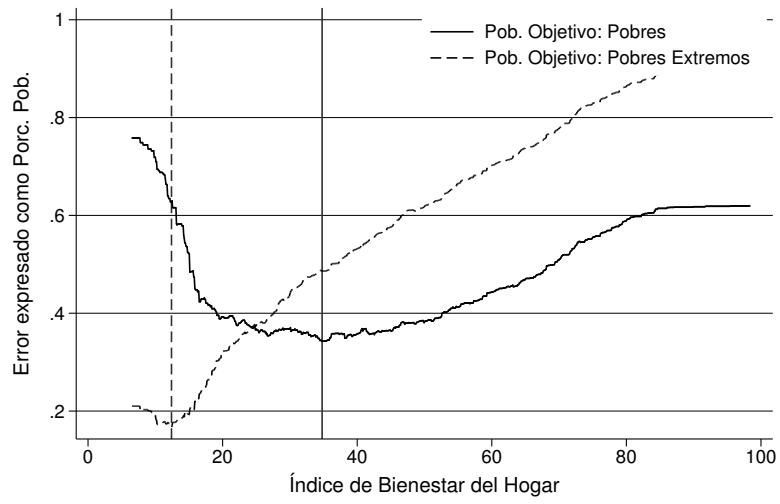


Figura C6: Sierra rural

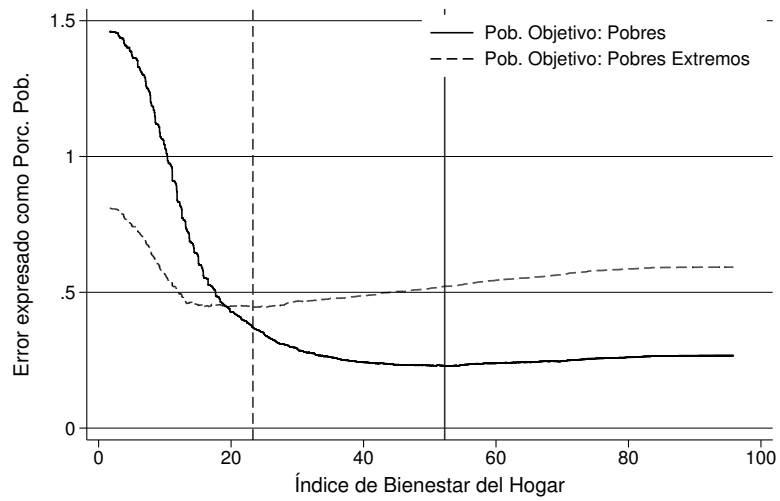


Figura C7: Selva rural

