

Una interpretación económica del Sistema de Focalización de Programas Sociales: el caso Sisben en Colombia

Carlos Eduardo Vélez E.¹

Elkin Castaño V.²

Ruthanne Deutsch D.³

I. Introducción

Este trabajo realiza una interpretación económica del Sisben -un índice aproximado de recursos usado para seleccionar beneficiarios de programas sociales en Colombia desde 1994⁴. Sirve como un indicador de bienestar económico de los hogares y se basa en la aplicación del Algoritmo de Cuantificación Óptima a una muestra de corte transversal de hogares colombianos seleccionados en 1993⁵. El índice Sisben es función de un conjunto de características demográficas y de variables relacionadas con el consumo de bienes durables, dotación de capital humano e ingreso corriente. El Sisben fue creado con la intención de simplificar y reducir el costo de focalización individual de beneficiarios de programas sociales en los distintos niveles de gobierno. El sistema ha sido exitoso y ha generado

alta demanda en municipios y departamentos, debido a sus bajos costos de operación y a los beneficios sociales que ha traído a las comunidades participantes. Actualmente, el Sisben es usado en varios programas sociales subsidiados, particularmente para los subsidios de salud establecidos por la Ley de Seguridad Social Colombiana y los subsidios de vivienda del Inurbe.

Nuestra pregunta central es si el Algoritmo de Cuantificación Óptima usado para derivar los puntajes de cada variable del índice Sisben, es consistente con los principios económicos que contempla el *índice aproximado de recursos* como una medida de bienestar. El análisis lleva a las siguientes conclusiones: *i)* aunque la Cuantificación Óptima es un algoritmo estadístico poco común, encontramos que el resultado de su aplicación, el índice Sisben, tiene un sólido

¹ Senior Economist, World Bank.

² Consultor Misión Social-DNP- e Investigador del Centro de Investigaciones Económicas de la Universidad de Antioquia.

³ Senior Economist, Banco Interamericano de Desarrollo (BID).

⁴ El Sisben fue desarrollado en la Misión Social, Departamento Nacional de Planeación, Colombia. Ver Castaño y Moreno (1994) para una presentación de los resultados y el algoritmo de construcción.

⁵ Encuesta Casen, Dane.

respaldo en la teoría del índice de bienestar. Y en este sentido puede ser visto, o bien como un índice de utilidad métrico cuantitativo o como un índice compuesto de formas alternativas de la función de utilidad de los hogares. Y, *ii*) el algoritmo usado para su construcción, soluciona los retos ocasionados por la naturaleza del problema y los datos disponibles.

La *no observabilidad* de la utilidad de los hogares excluye la aplicación de los métodos tradicionales de regresión para construir el índice. Adicionalmente, la inclusión de información *categórica* detallada sobre consumo de bienes durables de los hogares impide el uso de análisis de componentes principales lineales estándar. Por lo tanto, se requiere un algoritmo alternativo. El método de "Mínimos Cuadrados Alternantes y Cuantificación y Óptima-ALS-OS" resulta apropiado porque: *i*) impone una medida para las categorías de las variables discriminantes que son usadas para construir el índice, de forma tal que se maximice su contribución a un modelo de Componentes Principales; *ii*) en el caso del Sisben, el algoritmo es orientado a concentrar los ajustes de la cuantificación óptima de acuerdo a la contribución de cada categoría a la primera componente principal de la información. En este caso, la primera componente principal de las canastas de consumo observadas es la dimensión de máxima diferenciación o variabilidad entre hogares, y representa su nivel de utilidad. Como resultado, las diferentes categorías de todas las variables discriminantes están dadas en un orden, signo y valor consistentes con la intuición económica sobre consumo y pobreza.

Aunque a primera vista parece redundante el uso de variables determinadas simultánea y endógenamente -tales como consumo, ingreso

y dotación de capital humano-, realmente resulta necesario y útil para la convergencia del algoritmo específico implementado.

La siguiente sección del trabajo describe el Sisben: sus principales objetivos, alcance, cobertura, papel en la determinación de la asignación de recursos, costos, y factores que contaron para su rápida y exitosa implementación en Colombia. La sección 3 muestra un análisis económico del Sisben, *i*) examina su concordancia con la teoría económica sobre índices de bienestar *ii*) evalúa la pertinencia del método estadístico del Algoritmo ALS-OS para seleccionar el mejor conjunto de variables discriminantes y medir el número de puntos con los cuales ellas contribuyen al índice de bienestar total. La sección 4 describe el algoritmo correspondiente al Análisis de Componentes Principales Cualitativas usado para derivar el índice Sisben y detalla una serie de temas metodológicos que fueron usados durante la preparación del Sisben. Además, revisa los resultados sobre la cuantificación y las ponderaciones y realiza comparaciones para áreas rurales y urbanas. La sección final resume los principales resultados y conclusiones.

II. ¿Qué es el Sisben?

A. Objetivos

El Sisben -*Sistema de Selección de Beneficiarios para Programas Sociales*- fue diseñado para proveer a los gobiernos locales de un instrumento técnico para focalizar el gasto social hacia los hogares más pobres y vulnerables. Diseñado dentro del contexto de un esfuerzo de descentralización general en los sectores sociales que siguió a la Reforma Constitucional de 1991, el objetivo del Sisben no es solamente lograr preci-

sión en la focalización, sino también facilitar la implementación en los gobiernos locales con limitada capacidad administrativa.

El objetivo general es obtener información socioeconómica confiable sobre grupos pobres y vulnerables de la población colombiana, que pueda ser empleada en el desarrollo de programas sociales, particularmente en el programa subsidiado de la nueva reforma del sistema de salud pública.

Objetivos específicos incluidos:

- Facilitar la localización de beneficiarios para programas sociales subsidiados en una forma transparente, empleando un criterio uniforme, técnico y equitativo.
- Apoyar la coordinación inter-institucional a nivel municipal con el fin de mejorar el impacto del gasto social y eliminar la duplicación en beneficios del programa.
- Permitir la evaluación de las metas de gasto focalizado fijadas a nivel departamental, distrital y municipal, con el fin de lograr un impacto significativo en la lucha contra la pobreza en Colombia.
- Proveer de insumos para la preparación de estudios de diagnóstico socioeconómico detallado y perfiles de pobreza utilizables para la elaboración de planes municipales de desarrollo social y el diseño e implementación de programas específicos enfocados a los más necesitados; y
- Contribuir al fortalecimiento institucional a nivel municipal, introduciendo un moderno sistema de manejo de la información y proveyendo capacitación para su implementación.

B. ¿Cómo funciona el Sisben?

Esencialmente, el Sisben opera como un *indicador aproximado de recursos* de los hogares que son beneficiarios potenciales de programas sociales. Para implementar el Sisben se requieren dos pasos: modelización estadística y trabajo de campo. El ejercicio de modelización involucra seleccionar el conjunto de variables de los hogares que mejor prediga su bienestar; el establecimiento de un número de puntos que cada variable adiciona al índice de bienestar; y la determinación de los puntos de corte del índice de bienestar, con el fin de separar de la mejor manera posible los pobres de los no pobres. El trabajo de campo consiste en recolectar los datos para todos los beneficiarios potenciales sobre el conjunto de variables seleccionado, y ordenarlos de acuerdo con su índice de Sisben.

El Sisben es un índice de bienestar de los hogares de variables socioeconómicas cualitativas y cuantitativas, el cual fue derivado de la aplicación del algoritmo estadístico de Componentes Principales Cualitativas o Prinqual, a una encuesta de 25,000 hogares (Casen 1993, encuesta de hogares rural-urbana). El subconjunto final de variables seleccionado por el algoritmo, fue determinado como el más eficiente para predecir las diferencias en bienestar de hogares y pobreza⁶.

⁶ El conjunto final de variables discriminantes fue más pequeño que el conjunto con el cual se inició el proceso. Las incluidas fueron sugeridas por los expertos en pobreza en Colombia. El algoritmo seleccionó el subconjunto final de acuerdo a su contribución a la función objetivo del proceso.

Los datos de variables del Sisben para todos los beneficiarios potenciales de subsidios gubernamentales fueron recolectados fácilmente en una corta entrevista, aplicando el cuestionario *Ficha de Caracterización Socioeconómica*. Dada la información sobre las características de cada hogar y la contribución de cada variable en el índice de bienestar, la puntuación del Sisben para un hogar es calculada con la ayuda de un software personalizado. Posteriormente se obtiene un *ordenamiento* y una *estratificación* de beneficiarios potenciales⁷. Los Cuadros 1A y 1B describen las variables recolectadas en el cuestionario del Sisben y el número alternativo de puntos asignados para cada categoría respondida, con el fin de calcular el puntaje del Sisben para el hogar (entre 0 y 100).

Las variables discriminantes recopiladas por la Ficha de Caracterización Socioeconómica pueden agruparse en cuatro subconjuntos: Vivienda y electrodomésticos del hogar; servicios públicos; dotación de capital humano y seguridad social; y, finalmente, un cuarto grupo que incluye variables demográficas, desempleo, tasa de dependencia e ingreso per cápita⁸. En cada grupo hay tres o cuatro variables, cada una de las cuales contiene diferentes categorías y puede ser fácilmente observada o confirmada por la entrevista.

Los valores relativos son asignados a cada una de las (mutuamente excluyentes) catego-

rias observadas con cada subgrupo -estos valores son distintos para áreas rurales y urbanas-. Por ejemplo, en el área urbana hay una ganancia de cinco puntos asociados con la seguridad social y el tipo de empleo, cuando el principal perceptor del hogar trabaja en una empresa que da seguridad social y tiene más de diez trabajadores en vez de trabajar en un empresa pequeña que no da seguridad social⁹. También hay una ganancia de casi 12 doce puntos adicionales cuando el principal perceptor del hogar tiene educación superior en vez de tener solamente primaria incompleta. Si la proporción de miembros del hogar empleados está entre el 30% y el 90%, se obtienen tres puntos adicionales. Obtener el agua por acueducto da una ganancia de casi seis puntos comparado con obtenerla desde un pila pública. Poseer al menos tres electrodomésticos básicos del hogar y una lavadora, incrementa 2,5 puntos. Con este nivel de detalle, es posible calcular el puntaje Sisben final de los hogares¹⁰.

El índice final Sisben está dividido en estratos de acuerdo a la severidad de la pobreza: nivel 1 para los más pobres a los cuales corresponde la definición de extrema pobreza o Indigencia; nivel 2 para la definición tradicional de línea de pobreza; y, nivel 3 para tres veces la línea de pobreza extrema (ver Cuadro 2). Usando el ingreso y las Necesidades Básicas Insatisfechas como indicadores de pobreza, el punto de corte

⁷ En la mayoría de los casos, las comunidades locales permitieron revisar la selección de beneficiarios con el fin de detectar la clasificación errada de familias no pobres.

⁸ Las razones para esta aplicación serán explicadas en la sección 3 y en los anexos.

⁹ En una escala de 0 a 100 puntos de mínimo a máximo.

¹⁰ Ver Vélez, Castaño y Correal (1995).

Cuadro 1A
PUNTAJE PARA LAS CATEGORÍAS DE CADA VARIABLE
Sisben urbano 1993

Descripción de las categorías de las variables	Puntaje	Descripción de las categorías de las variables	Puntaje
Capital humano, seguridad social		Ingreso per cápita familiar en unidades de salario mínimo	
Escolaridad del mayor receptor		(0.50, 0.75]	5,3636
Sin Educación	0	(0.75, 1.00]	7,0827
Primaria incompleta	1,6239	(1.00, 1.25]	8,2489
Primaria completa	3,4435	(1.25, 1.50]	9,4853
Secundaria incompleta	5,0039	(1.50, 2.00]	10,2098
Secundaria completa	7,3434	(2.00, 3.00]	11,3999
Superior Incompleta	9,7833	(3.00, 4.00]	13,0872
Superior completa	11,546	Más de 4.0	17,7378
Posgrados y Doctorados	12,4806		
Escolaridad promedia personas de 12 y más años		Vivienda	
Sin educación	0	Material de las paredes	
(0, 4]	1,6570	Sin paredes o guadua, caña, u otros materiales vegetales	0
(4, 5]	2,9947	Zinc, tela, cartón, latas o deshechos	0,2473
(5, 10]	4,9690	Madera burda	2,0207
(10, 11]	7,6387	Bahareque	4,8586
(11, 15]	9,4425	Tapia pisada o adobe	6,2845
(15, 16]	10,6900	Bloque, ladrillo, piedra, material prefabricado, madera pulida	7,7321
más de 16	11,1396	Material techos	
Seguridad social mayor receptor		Paja o palma	0
Sin seguridad social y trabaja sólo o no trabaja	0	Desechos (Cartón, lata, sacos, etc.)	2,1043
Sin seguridad social y trabaja en empresa de 2-9 empleados	1,1660	Zinc, asbesto, cemento, sin cielo raso	3,7779
Sin seguridad social y trabaja en emp. de más de 10 empleados	2,6545	Teja, losa, zinc, cemento, con cielo raso	5,0973
Con seguridad social y trabaja sólo o no trabaja	3,9539	Material pisos	
Con seguridad social y trabaja en empresa de 2-9 empleados	5,8427	Tierra	0
Con seguridad social y trabaja en emp. de más de 10 empleados	6,9718	Madera burda, tabla o tablón	2,9037
		Cemento	3,6967
Demográfico, ingreso y educación		Baldosa, vinilo, tableta o ladrillo	5,8712
Proporción de menores de 6 años		Alfombra de pared a pared, mármol, madera	6,8915
Más de 0.65	0	Equipamiento	
(0.0, .65]	0,2237	Ningún básico	0
Sin niños	1,4761	Hasta tres básicos sin lavadora	2,1435
Proporción de ocupados		Cuatro básicos sin lavadora	3,0763
Menos de 0.30	0	Tres o cuatro básicos con lavadora	4,7194
(0.30, 0.60]	0,6717		
(0.60, 0.90]	1,7390	Servicios	
Más de 0.90	4,0149	Abastecimiento de agua	
Hacinamiento		Río o manantial	0
Número de cuartos/personas familia		Pila Pública u otra fuente	1,1601
Menos de 0.20	0	Pozo sin bomba, algibe, barreno o agua lluvia	2,6497
(0.20, 0.30]	0,5584	Pozo con bomba	4,6037
(0.30, 0.40]	1,6535	carrotanque	6,1693
(0.40, 0.70]	2,5727	Acueducto	7,2554
(0.70, 1.00]	4,3886	Eliminación de excretas	
(1.00, 4.00]	6,0042	No tiene servicio sanitario	0
Más de 4.0	8,3828	Letrina	2,4519
Ingreso per cápita familiar en unidades de salario mínimo		Inodoro sin conexión a alcantarillado o pozo séptico	3,3323
Menos de 0.15	0	Inodoro conectado a pozo séptico	3,9615
(0.15, 0.25]	0,8476	Inodoro conectado a alcantarillado	6,8306
(0.25, 0.35]	2,1828	Recolección de basuras	
(0.35, 0.50]	3,5362	La tiran a un patio, lote, zanja, río o quebrada o laguna	0
		La llevan a un contenedor cercano o basurero público	2,1291
		La recogen los servicios de aseo	3,2701

Fuente: Castaño y Moreno, (1994).

Cuadro 1B

PUNTAJE PARA LAS CATEGORÍAS DE CADA VARIABLE

Sisben rural 1993

Descripción de las categorías de las variables	Puntaje	Descripción de las categorías de las variables	Puntaje
Vivienda		Servicios	
Material de las paredes		Abastecimiento de agua	
Guadua, caña u otros materiales vegetales	0	carrotanque	0
Zinc, tela, cartón, latas o deshechos	3,2042	Pila pública u otra fuente	1,0523
Tapia pisada, adobe o bahareque	3,4319	Pozo sin bomba, algibe, barro o agua lluvia	1,6591
Madera burda	4,5588	Río o manantial	2,264
Bloque, ladrillo, piedra, material prefabricado, madera pulida	7,0780	Pozo con bomba	3,5759
Material de los techos		Acueducto	7,2438
Paja o palma	0	Eliminación de excretas	
Desechos (cartón, lata, sacos, etc)	1,1312	No tiene servicio sanitario	0
Zinc, asbesto, cemento, sin cielo raso	3,7615	Letrina	1,4398
Teja, losa, zinc, asbesto, cemento, con cielo raso	4,8771	Inodoro sin conexión a alcantarillado o pozo séptico	3,0718
Material pisos		Inodoro conectado a pozo séptico	4,211
Tierra	0	Inodoro conectado a alcantarillado	7,3137
Madera burda, tabla o tablón	2,4628	Recolección de basura	
Cemento	3,7474	La tiran a un patio, lote, zanja, río, quebrada o laguna	0
Baldosa, vinilo, tableta o ladrillo	5,4726	La llevan a un contenedor cercano o basurero público	1,5414
Alfombra de pared a pared, mármol, madera	5,7495	La recogen los servicio de aseo	2,6758
Equipamiento familiar		Demográfico, ingresos y ocupación	
Ningún básico	0	Proporción de niños menores de 6 años	
Hasta tres básicos	1,6865	Más de 0.65	0
Más de tres básicos	2,708	(0.0, 0.65]	0,2181
Capital humano y seguridad social		Sin niños	1,1626
Escolaridad del mayor receptor		Proporción de ocupados	
Sin educación	0	Menos de 0.30	0
Primaria incompleta	0,2528	(0.30, 0.60]	1,0806
Primaria completa	1,5723	(0.60, 0.90]	1,8668
Secundaria incompleta	3,7663	Más de 0.90	3,1957
Secundaria completa	5,1876	Hacinamiento	
Superior incompleta	11,4263	(Número de cuartos/personas familia)	
Superior completa	13,0990	Menos de 0.30	0
Postgrados y doctorados	18,3860	(0.30, 0.60]	0,8956
Escolaridad promedio de personas de 12 y más años		(0.60, 1.00]	1,8988
Sin educación	0	(1.00, 4.00]	2,9379
(0, 4]	0,3269	Más de 4.00	4,9313
(4, 5]	1,5793	Ingreso per cápita en unidades de salario mínimo	
(5, 10]	3,2931	Menos de 0.15	0
(10, 11]	6,3560	(0.15, 0.25]	1,1079
(11, 15]	12,3633	(0.25, 0.35]	1,9561
(15, 16]	12,5863	(0.35, 0.50]	2,9685
Más de 16	18,0414	(0.50, 0.75]	3,9781
Seguridad social del mayor receptor		(0.75, 1.00]	4,9210
Sin seguridad social y trabaja sólo o en emp. de 2 a 9 empleados	0	(1.00, 1.25]	5,6862
Sin seguridad social y trabaja en emp. de 10 o más empleados	1,4320	(1.25, 2.00]	5,6862
Con seguridad social y trabaja sólo	2,6097	(2.00, 3.00]	7,7840
Con seguridad social y trabaja en empresa de 2 a 9 empleados	3,6514	(3.00, 4.00]	8,5781
Con seguridad social y trabaja en emp. de 10 o más empleados	4,5259	Más de 4.00	9,3504

Fuente: Castaño y Moreno, (1994).

para los primeros tres estratos está determinado por la aplicación de un modelo probit con el objetivo de minimizar la subcobertura o el error de excluir a los "verdaderos" pobres¹¹.

Hay seis niveles de clasificación de Sisben, con puntaje total entre 0 y 100. La mayoría de los programas sociales nacionales están dirigidos a la población clasificada en los niveles 1 y 2 de pobreza, algunos incluyen el nivel 3. Para elegir la población objetivo para el programa de salud pública subsidiado, el punto de corte es 47 en áreas urbanas y 30 en rurales. Adicionalmente, de acuerdo a su restricción presupuestaria, algunos programas pueden ajustar el punto de corte, y, dado que cada hogar tiene un puntaje Sisben absoluto que ordena los hogares de acuerdo a la intensidad de la pobreza, el sistema es útil para asignar los recursos hacia los más necesitados.

La pobreza rural relativa medida por Sisben, parece ser más intensa que la medida por ingreso (Cuadro 2). Usando ingreso, línea de pobreza extrema, la tasa de pobreza rural es aproximadamente 33%, 3,3 veces la tasa urbana. Sin embargo, usando la estratificación Sisben, la pobreza rural extrema es cercana al 23%, 4 veces la tasa urbana¹².

C. Cobertura y tasas de implementación

El Sisben fue implementado principalmente entre 1994 y 1995. Se aplicó en más del 90% de

Cuadro 2
ESTRATOS SISBEN: PUNTOS DE CORTE Y DISTRIBUCIÓN DE LA POBLACIÓN

Estrato Sisben	% de población por estrato			% pobre por ingreso, 1992 ^a
	Puntos máximos	Predicha	Acumulada	
Urbano				
Pobreza extrema	35,8	4,9	4,9	9,9
Pobreza	47,4	15,5	20,5	37
Tres veces LP extrema	58,5	31,4	46,9	
Rural				
Pobreza extrema	18,0	23,0	23,0	32,7
Pobreza	29,9	32,7	55,7	65
Tres veces LP extrema	45,4	34,2	66,9	

^a Dane, Banco Mundial, (1995).

Fuente: Encuesta Casen, Castaño (1995).

los municipios en Colombia, gracias a una amplia campaña de difusión, asistencia técnica, y entrenamiento diseñado y coordinado desde el nivel central por la Misión Social del Departamento Nacional de Planeación, e implementado en asocio con los departamentos administrativos de planeación a nivel local.

El Cuadro 3 provee más detalles sobre la cobertura del Sisben, que en 1996 fue aproximadamente del 40% de la población total de Colombia¹³. El 60% de los hogares encuestados estaban ubicados en áreas urbanas y el 40% restante en rurales¹⁴. Debido a la concentración de población en las áreas urbanas, menos del 30% fue cubierta por el Sisben y un poco menos del 40% en las áreas rurales. En áreas urbanas, cerca de dos terceras partes de la población encuestada

¹¹ Llamado error tipo I. Ver Castaño (1995).

¹² Cuando se usa la línea de pobreza ingreso los resultados son similares, aunque las diferencias relativas son más pequeñas.

¹³ A junio de 1997, el 98% de los municipios de Colombia habían participado en el Sisben.

¹⁴ Estas proporciones no están muy lejos de la distribución de la pobreza por criterio de ingreso: 44% urbano y 56% rural, de acuerdo al Banco Mundial (1995), cap.1.

Cuadro 3
POBLACIÓN TOTAL Y SISBEN
Urbana y rural por estratos pobres

	Total población	Población Sisben/96	% cubierto	Afiliación a seguro subsidiado de salud		% de población encuestada por estratos				
				Potencial=estratos 1+2 %	Real 1997	Estrato 1	Estrato 2	Estrato 3	Resto	
Total nacional	38.318.448	12.208.122	32	8.354.415	68	7.026.689	29,2	39,2	24,8	6,7
Total urbano	27.751.869	7.735.114	28	4.851.053	63	-	24,6	38,2	27,9	9,4
Total rural	11.758.788	4.473.008	38	3.503.362	78	-	37,2	41,1	19,6	2,1
Rural centros poblados	-	-	-	-	-	-	24,6	29,6	14,4	1,2
Rural dispersa	-	-	-	-	-	-	12,7	11,4	5,2	0,8

Fuente: Informe de las oficinas de Planeación Departamental sobre estado de avance Sisben. Dane: Población estimada 1996, a partir de Censo de Población 1993.

demonstró niveles 1 y 2 de pobreza de acuerdo con el índice Sisben, haciéndolos potencialmente elegibles para el sistema de salud pública subsidiado. En áreas rurales, el 78% de la población de clasificó en los dos primeros grados de pobreza. En áreas rurales, la implementación del Sisben se orientó a alcanzar más zonas marginales, logrando que las tres cuartas partes de los hogares encuestados residieran fuera de los centros de los pueblos.

La cobertura de la encuesta fue determinada en el nivel local con base en un análisis de información socioeconómica que permitía determinar las localidades donde estuvieran concentrados los grupos de bajo ingreso. En estas áreas se emprendió una primera ronda de censo a gran escala, acumulando alrededor del 95% de la cobertura del Sisben. Subsecuentemente, hubo una fase de demanda dirigida de la encuesta, donde las ONG y los grupos comunitarios solicitaron que se encuestaran hogares específicos y en áreas donde fuera latente la demanda por programas sociales. Además, los hogares individualmente podían aplicar a la encuesta Sisben, si creían que podían ser elegibles para recibir un subsidio y no habían sido encuestados en las

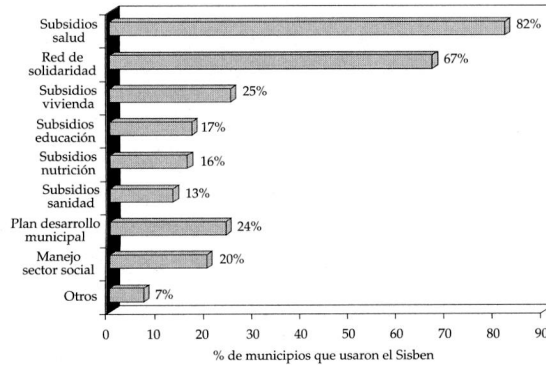
rondas anteriores. A pesar de todo esto, cerca de la tercera parte de los municipios estudiados en una reciente evaluación, reportaron problemas de bajo cubrimiento del grupo objetivo.

D. Utilización

Si bien el Sisben fue usado principalmente para la expansión de programas de salud pública subsidiada (Gráfico 1) - más del 80% de los municipios participantes la usaron para este propósito-, también fue usado para facilitar la asignación de recursos de otras fuentes. El 62% de los municipios usaron los resultados del Sisben como una herramienta para asignar fondos de la Red de Solidaridad. También para asignar subsidios de vivienda (25% de los municipios), planes de desarrollo municipal (24%); y subsidios de educación y nutrición (17% y 16% respectivamente).

El Cuadro 3 muestra que más de dos terceras partes de los 12 millones de los encuestados por el Sisben, 8,3 millones de personas, pertenecían a los estratos 1 y 2, y se convirtieron en beneficiarios potenciales de salud pública subsidiada. Gracias a la extensa cobertura del Sisben, los

Gráfico 1
UTILIZACIÓN DEL SISBEN POR TIPO DE
PROGRAMA. FRECUENCIA POR MUNICIPIO
(Colombia 1996)



Fuente: DNP-UNDP (1996).

municipios lograron incluir fácilmente su población beneficiaria al Sistema General de Seguridad Social en Salud creado en 1993 por la Ley 100. Como resultado, el número de personas cubiertas a finales de 1997 era aproximadamente 7 millones. Esto representa casi el 85% del total de beneficiarios potenciales, y claramente demuestra el éxito de la expansión del nuevo plan de salud, una vez fue vinculado a un método conveniente y relativamente barato de focalización.

E. Los Costos de monitoreo son bajos

Los costos de monitoreo del Sisben son muy moderados. El Cuadro 4 presenta información sobre el promedio de dichos costos por familia Sisben en 453 municipios, que representan alrededor del 90% del total del número de Fichas

aplicadas por la Misión Social en 1996¹⁵. La mediana del costo de aplicación del cuestionario Sisben en 1996 fue 2.395 pesos -aproximadamente US\$2,25 en 1996-. Los costos fueron relativamente heterogéneos, pero la mitad de la muestra cayó entre el 73% y 152% del costo medio reportado por cada familia investigada. Los costos por persona fueron bastante pequeños, en promedio 570 pesos o US\$0,53 por persona. Los costos de monitoreo son una muy pequeña porción -0,5% del costo anual de un plan de salud subsidiado para una familia de cinco personas. Y mucho menor para un subsidio de Vivienda del Inurbe, que tiene un valor dos o tres veces mayor.

Además, la utilización del Sisben para la focalización de más de un programa a la vez, genera economías de escala y reduce aún más el costo de monitoreo. Comparado con otros mecanismos de monitoreo en América Latina, parece ser barato. De acuerdo con Margaret Grosh (1994), un sistema similar de focalización tiene costos entre 0,5% y 1,5% del gasto del programa¹⁶.

Cuadro 4
COSTO PROMEDIO DE APLICACIÓN DEL
CUESTIONARIO SISBEN
Municipios colombianos, 1996

	Costo	Frecuencia acumulada
Primer cuartil	1.752,00	25
Segundo cuartil (mediana)	2.395,00	50
Tercer cuartil	3.657,00	75
Promedio	3.295,31	71

Fuente: Misión Social, DNP.

¹⁵ DNP-UNDP (1996).

¹⁶ Ver capítulo 4 en Grosh (1994). Además, los costos administrativos deben ser comparados; sin embargo, la información actual disponible para Colombia no permite éste cálculo.

F. Factores que contribuyeron al éxito de la implementación

Varios factores políticos e institucionales contribuyeron a la rápida implementación del Sisben en Colombia. En primer lugar, La Constitución Colombiana de 1991 señaló explícitamente la necesidad de focalizar o dirigir el gasto social a los grupos más necesitados del país en el nivel nacional, departamental y municipal. En efecto, más del 60% de las transferencias territoriales son distribuidas en proporción directa al número de residentes con Necesidades Básicas Insatisfechas, como se estipuló en el artículo 357 de la Constitución. Segundo, el artículo 30 de la Ley 60 de 1993 sobre Transferencias Territoriales y Recursos, define la focalización de subsidios públicos como "el proceso por el cual se garantiza que el gasto social sea asignado a los grupos más necesitados".

El Conpes -Consejo para la Política Económica y Social- tiene el mandato de definir cada tres años el criterio para la determinación, identificación y selección de beneficiarios para la asignación de gasto social por los distintos niveles del gobierno en Colombia. El Conpes estipuló dos criterios en 1994: el sistema de estratificación socioeconómica y el Sisben. El primero es usado esencialmente para subsidios focalizados microgeográficos a través de la discriminación de tarifas, y el segundo para identificar beneficiarios.

El segundo estímulo -ciertamente el más poderoso- para la expansión del Sisben, fue la promulgación de la descentralización del Sistema Nacional de Salud, por medio de la cual se asignaron recursos a los municipios para el subsidio a la salud pública sobre la base del número de hogares elegidos en cada municipio -definidos

por su puntaje Sisben. Además, este incentivo fue reforzado por el crecimiento de otros programas sociales que requieren medidas de bienestar individual para focalizar: Fondo Social -Red de Solidaridad Social- y el Programa de subsidios de Vivienda del Inurbe.

Adicionalmente, un diseño cuidadoso y un entrenamiento masivo en el nivel local estimularon la amplia implementación del sistema Sisben. Estos recursos gratis no solamente redujeron los costos monetarios fijos, sino también los costos del tiempo de espera de la implementación del sistema de focalización y de los programas de entrega. La efectiva división del trabajo entre los niveles nacional, departamental, distrital y municipal del gobierno permitió el apoyo técnico necesario para la implementación del programa. El excelente mercadeo social, la educación y el entrenamiento de los funcionarios públicos y representantes de las Organizaciones No Gubernamentales, crearon más conciencia acerca de la utilidad de la información del sistema y demanda para su implementación. El diseño amigable, complementado por una serie de panfletos instructivos y un software personalizado para procesar el puntaje Sisben y el uso de la información para el monitoreo del gasto social en el nivel municipal, incrementó la relevancia y utilidad del sistema para los gobiernos locales.

En resumen, la mayoría de los costos fijos para la implementación fueron absorbidos por el gobierno central, y los costos marginales -monetarios y no monetarios- asumidos por los municipios, fueron relativamente bajos. Las decisiones para crear el Sisben fueron descentralizadas y dieron una garantía política para el retorno del liderazgo local.

III. Una interpretación económica del Sisben

En esta sección mostraremos que el índice Sisben puede ser interpretado como un índice compuesto de indicadores de bienestar que permite ordenar los hogares de acuerdo con su nivel de bienestar. Antes de empezar con el principal tema de esta sección, empezaremos con una amplia revisión de las ventajas y limitaciones de los indicadores de bienestar.

A. Medidas alternativas de bienestar individual

Típicamente, las medidas de bienestar individual son instrumentos que ayudan a la focalización de programas públicos e incrementan su capacidad para reducir la pobreza y la inequidad. Hay varios argumentos, analíticos y empíricos, que muestran que cuanto más y mejor información acerca de niveles de bienestar individual sea usada para la focalización, mayor será la ganancia en equidad¹⁷.

En la elección de medidas de bienestar individual, podemos identificar tres aproximaciones básicas¹⁸: La aproximación *Bienestarista*, basada en la utilidad individual derivada de bienes o ingreso. La aproximación de *Necesidades Básicas*, relacionada con la satisfacción de las necesidades humanas o el nivel de bienestar lo-

grado en términos específicos, tales como: vivienda, servicios básicos, hacinamiento, dependencia económica y acceso a la educación¹⁹. Y tercero, la aproximación de *capacidades* planteada por Sen, que mira los bienes no como finales sino como medios para realizar actividades deseadas²⁰.

Todas las medidas alternativas tienen ventajas y desventajas; la implementación de cada tipo de medida encara un conjunto de complicaciones específicas que restringe la elección de una medida de bienestar.

La aproximación de las *Capacidades* enfrenta el problema de no observabilidad. Los logros son observados pero no las capacidades²¹. La aproximación *Necesidades Básicas* es funcionalmente inflexible en términos de pesos, y es lexicográfica, porque una necesidad básica insatisfecha es suficiente para catalogar un hogar como pobre. En este sentido, el índice de Necesidades Básicas Insatisfechas -NBI- da una señal confusa de pobreza extrema, puesto que ignora el nivel de satisfacción en otras dimensiones de bienestar. De acuerdo al Banco Mundial (1995), en el caso de Colombia, el NBI es un indicador inadecuado de la pobreza por *ingreso*:

Aunque el pobre (definido sobre la base de tres líneas alternativas de pobreza) experimenta una mayor incidencia de las cuatro NBI's, que la población como un

¹⁷ Por ejemplo, Ravallion y Chao (1989), Glewwe y Kanaan (1989).

¹⁸ Seguimos la clasificación sugerida en Lipton y Ravallion (1995), pp. 2566-69 y Deaton (1997) cp. 3.

¹⁹ Ver Streeten et. al. (1981) y Richards et. al. (1982).

²⁰ Sen (1979, 1985, 1987).

²¹ Lipton y Ravallion (1995) p.

conjunto (...) la incidencia de las Necesidades Básicas Insatisfechas (...) está lejos de identificarlo. En áreas urbanas, por ejemplo, solamente uno de tres individuos en el grupo de extrema pobreza, es pobre bajo el criterio NBI. En la medida que los diseñadores de políticas deseen identificar familias con ingreso inferior a la línea de pobreza, el uso del índice NBI involucraría grandes errores de exclusión e inclusión²².

El Ingreso es ideal porque evita una función de utilidad paramétrica, pero es difícil obtener información confiable. El consumo es bueno pero requiere recolección cara de información y la disponibilidad de índices de precios. Adicionalmente, estas dos medidas bienestaristas ignoran el papel del gasto público²³, y por lo tanto pueden sobrestimar la pobreza y la inequidad, cuando el gasto público es progresivo²⁴.

Las medidas de consumo también son criticadas porque ignoran el efecto del ciclo de vida y varias restricciones para suavizar dicho consumo -especialmente crédito para los pobres-²⁵. Si creemos que los pobres tienen mayores limitaciones para acceder al crédito, su acumulación de bienes durables debe seguir aproximadamente el camino de su ahorro; por lo tanto, sus diferencias en consumo durable con respecto del resto de consumidores no restringidos, deben ser más

grandes que lo que sugerirían sus diferencias en ingreso, especialmente para hogares pobres en las etapas más tempranas del ciclo de vida.

La aproximación *bienestarista* ofrece el fundamento para el *Indicador de Recursos Exacto* que típicamente está basado en el ingreso o el consumo. Sin embargo, estos indicadores no son prácticos en países en desarrollo²⁶. Además de los incentivos para subestimar los niveles de bienestar de ingreso o de consumo, la verificación es costosa dada la falta de registros acerca de la seguridad social y los impuestos. Una alternativa favorable es el uso de *indicadores aproximados de recursos*, los cuales son índices ponderados sobre características del hogar distintas al ingreso del hogar, incluyendo economías de escala o escalas equivalentes en la composición familiar²⁷.

Este tipo de índices ha sido usado en países de América Latina, -e.g., Chile, Costa Rica y Colombia- y parecen ser buenos para predecir el bienestar individual. Grosh y Baker (1995) probaron la capacidad predictiva de un *Indicador Aproximado de Recursos* en Jamaica, como un indicador de bienestar. En este caso, el índice contenía información acerca de ubicación, vivienda, dotación de capital humano de la familia, empleo y posesión de bienes durables. Su eva-

²² World Bank (1995), p.8.

²³ Deaton (1997) o Lipton y Ravallion (1995)

²⁴ Por ejemplo, Vélez (1995, 1997) muestra que el gasto público colombiano en 1992 disminuyó el coeficiente de Gini-ingreso más subsidios imputados- en casi 9%. Y la estimación de tasa de pobreza se redujo cerca de 4 puntos porcentuales.

²⁵ Lipton y Ravallion (1995).

²⁶ Grosh y Baker (1995).

²⁷ Note que el *Indicador Aproximado de Recursos* pertenece a la clase de indicadores basados en bienes o de aproximación *bienestarista*.

luación muestra que éstos índices son razonablemente buenos para determinar elegibilidad para programas sociales. Aunque no elimina el problema del error de exclusión completamente, reduce el error de inclusión y, como resultado, su capacidad para reducir la pobreza se incrementa significativamente.

B. Ordenamiento de los hogares de acuerdo a su nivel de bienestar

Dada la anterior descripción, el Sisben es obviamente un *indicador aproximado de recursos*. Es un índice ponderado de varias características del hogar que fue construido con un algoritmo aparentemente ad-hoc, llamado Cuantificación Óptima y Mínimos Cuadrados Alternantes. En esta sección proponemos una interpretación del fundamento económico implícito en el procedimiento usado para derivar el índice Sisben del bienestar del hogar, desde la perspectiva del comportamiento del consumidor.

Primero que todo se debe observar que este índice incluye tres conjuntos de variables relevantes para evaluar el nivel de utilidad del consumidor: *i)* consumo de bienes durables y servicios tales como vivienda y aparatos del hogar *ii)* dotación de capital humano, *iii)* ingreso corriente.

1. Índice de Bienestar Métrico Cuantitativo

La importancia de usar el consumo real de bienes y servicios para construir un índice de bienestar, puede ser interpretada en el sentido de un *Índice de Bienestar Métrico Cuantitativo*, como fue definido por Deaton y Muellbauer (1980)²⁸. Estos

índices se llaman curvas de indiferencia, no por su costo o relación con el ingreso, sino por las *distancias* del origen con respecto a alguna canasta de referencia (Gráfico 2). Están basados en el concepto de función *distancia*, y su valor es enteramente *ordinal*.

La definición de la función *distancia* es: dada una canasta de consumo q con un nivel de utilidad $v(q)$ y un nivel de utilidad de *referencia* u , la función $d(u, q)$ es el número real por el cual q debe ser dividido para obtener el nivel de utilidad u :

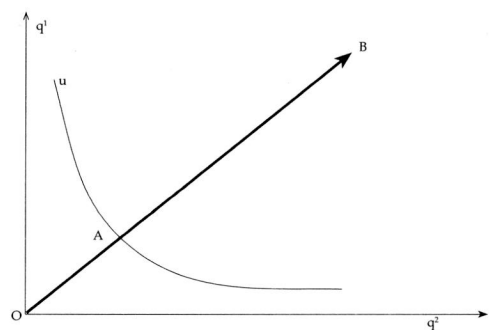
$$v(q / d(u, q)) = u;$$

y en el gráfico 2:

$$d(u, q) = OB/OA$$

Debe notarse que la función *distancia* es proporcional a la función de gasto, si las preferencias

Gráfico 2
CANASTAS DE CONSUMO, NIVELES DE UTILIDAD E ÍNDICE DE BIENESTAR MÉTRICO CUANTITATIVO



Fuente: Deaton y Muellbauer (1980).

²⁸ Ver Deaton y Muellbauer (1980), pp. 54 y 181-182 y Deaton, A. (1979).

son homotéticas y el vector precios es elegido apropiadamente.

¿Cuál es la relación entre la función *distancia* y la información disponible sobre consumo del hogar?. Para simplificar, suponga que para un conjunto de hogares con una distribución de ingresos dada, las preferencias son homotéticas e iguales para todos, igual tamaño de hogar y composiciones demográficas, y no hay variación de precios regionales. Por esto, las canastas observadas de consumo de los hogares con diferentes niveles de ingreso caen *a lo largo de la línea* que pasa a través del origen en el espacio de bienes²⁹. Y la distancia de cada hogar al origen, representa su nivel de bienestar en términos ordinales. Entonces, si aceptamos la heterogeneidad e introducimos cambios en las características demográficas del hogar, en los gustos y precios regionales, deberíamos observar una "nube" de canastas de consumo sobre ambos lados y a lo largo de la misma línea en el espacio de consumo. Por lo tanto, en este contexto, la primera componente principal del conjunto completo de canastas de consumo observadas de los hogares en el espacio de bienes, representaría una aproximación de la función *distancia*.

2. Demandas Marshallianas

De la teoría del consumidor sabemos que sus preferencias pueden ser representadas por una

función de utilidad del espacio de bienes a la recta real. Esta representación cardinal de las preferencias ordenadas es tal que los vectores de demanda del consumidor determinan diferentes niveles de bienestar individual: esto es, $U(x_1) > U(x_2)$, sí y solo sí el vector de bienes x_1 es estrictamente preferido a x_2 ³⁰.

En el caso del Sisben, el primer conjunto de variables observadas para cada hogar i es el vector de demandas Marshallianas $x^*(p_o, y_i)$ ³¹. Dado que los hogares *difieren en sus niveles de ingreso*, la utilidad directa derivada de las demandas Marshallianas observadas deben reflejar sus diferencias en bienestar, esto es, su nivel de utilidad se incrementa con el ingreso:

$$U(x^*(p_o, y_i)) \geq U(x^*(p_o, y_j)) \quad \text{sí} \quad y_i > y_j$$

Sin embargo, otras fuentes de variación en el bienestar deberían provenir de cambios en los gustos y diferencias en los vectores de precios que cada hogar enfrenta -principalmente de cambios regionales.

Entonces, para obtener el orden de bienestar de los hogares, postulamos un índice de utilidad U_i como una función del vector de demandas observadas para bienes x_i de cada hogar $i=1, \dots, n$,

$$U_i = F(x_i) + \varepsilon_i + v_i$$

²⁹ Si los precios relativos son seleccionados apropiadamente, estarían ordenados a lo largo de la raya OAB en el Gráfico 2.

³⁰ Si hay una preferencia ordenada sobre el espacio de bienes $X \subset \mathbb{R}^n$, y este ordenamiento es racional, continuo y estrictamente monótonico, entonces existe una función de utilidad del espacio de bienes sobre la recta real. Ver Varian (1992) cap. 7 ó MasCollé (1997), caps. 1 y 3.

³¹ Estas cantidades son el producto de la optimización de sus preferencias -i.e.: maximizar $U(x)$ - restringido a su ingreso, y el vector precio que rige el mercado de bienes, p_o .

Donde ε_i y v_i son los términos de error que corresponden a cambios en los gustos y diferencias regionales en los vectores de precios respectivamente. De nuevo nótese que $F(x_i)$ puede ser visto como un *índice de utilidad métrico cuantitativo* de una curva de indiferencia, y por tanto puede representarse por la escala de un vector en el espacio de bienes³².

3. Ingreso corriente

Según la función de utilidad indirecta: $V(p, y) \equiv U(x^*(p, y))$, es obvio que $V(\cdot)$ debe reflejar las mismas diferencias en la distribución del ingreso que $U(\cdot)$ captura en términos de consumo óptimo. La Función de Utilidad Indirecta es no-decreciente en el ingreso, por lo tanto:

$$V(p_0, y_i) \geq V(p_0, y_j) \quad \text{sí} \quad y_i > y_j$$

Aún más, sabemos que *i*) la función de gasto del hogar, $e(p, U) \equiv p' x_h(p, U)$, representa el mismo ordenamiento dado por el nivel de utilidad U y *ii*) es idénticamente cierto que $y \equiv e(p_0, V(p_0, y))$. Entonces, el ingreso como tal representa el mismo ordenamiento dado por la función de utilidad indirecta $V(\cdot)$ y la función de utilidad directa $U(\cdot)$ ³³. Por lo tanto, las diferencias en bienestar de los hogares pueden ser representadas por un índice utilidad VI_i como una función del ingreso corriente, yc_i de diferentes hogares, sujeto a fuentes adicionales de variación provenientes de *i*) componentes temporales de ingreso, ρ_i y *ii*) el cambio en los gustos y la diferencia en los precios regionales, mencionados antes.

$$VI_i = G(yc_i) + \varepsilon_i + v_i + \rho_i$$

4. Dotación de Capital Humano

El ingreso laboral de los hogares es la realización de servicios, derivado de la dotación de capital humano, en el mercado laboral. Este proceso ocurre como una decisión conjunta de participación de la fuerza laboral (ocio versus trabajo) y consumo de bienes y servicios. El proceso de optimización de un hogar típico está representado por la maximización de una función de utilidad de consumo y ocio sujeta al ingreso de fuentes exógenas y de la venta de servicios laborales. Obviamente, las funciones de demanda de consumo Marshallianas c^* y ocio l^* (en $[0, 1]$) serán funciones del vector de precios -incluyendo el salario o el precio del ocio- y el ingreso total que es lineal a la dotación de capital humano del consumidor h y el ingreso exógeno Y_0 :

$$l_i^*(p, wh, + Y_0) \quad \text{y} \quad c_i^*(p, wh, + Y_0)$$

Donde, h_i es el número de unidades de capital humano equivalentes de un agente i que tienen precio w en el mercado laboral. Y la correspondiente función de utilidad indirecta es:

$$V_i((p, h_i, Y_0) \equiv U_i(l_i^*(p, h_i, Y_0), c_i^*(p, h_i, Y_0))$$

Lo cual implica que el ordenamiento del bienestar de los hogares puede ser representado por un índice de utilidad $V2_i$, función de dotaciones de capital humano observado del hogar (h_i)³⁴. Y este índice además refleja el orden de preferen-

³² Deaton y Muellbauer (1980), p.181.

³³ Deaton (1997), p.148, enfatiza en las ventajas de usar el ingreso real o el gasto como indicadores de bienestar "con el fin de evitar la especificación de una función de utilidad paramétrica".

cia representado por $U_i(\cdot)$ y $V1_i(\cdot)$. Obviamente, hay fuentes adicionales de variaciones, δ_i , que vienen de *i*) diferencias regionales en los precios de los servicios laborales y *ii*) choques exógenos de demanda de trabajadores calificados.

$$V2_i = H(h_i) + \varepsilon_i + v_i + \delta_i$$

Así hemos mostrado que por cada uno de los tres conjuntos de variables incluidas en el índice Sisben postulamos un índice de utilidad para los hogares de la muestra. Entonces, una combinación lineal positiva de estas tres representaciones del orden de los hogares, W_i , el bienestar debe reflejar ese mismo orden,

$$W_i = a U_i + b V1_i + c V2_i = a F(x_i) + b G(y_i) + c H(h_i) + \mu_i$$

Donde a, b y c son constantes positivas. Entonces el índice $W_i(x_i, y_i, h_i)$ sobre la línea real de bienes durables, ingreso corriente y dotación de capital humano, debe discriminar los hogares sobre la muestra en la dirección del bienestar económico.

5. Información relevante sobre consumo, pero en forma categórica y sin una métrica

En el índice Sisben, la información sobre demanda de bienes y servicios es limitada: primero que todo, no puede contener el total del conjunto de bienes, pero sí un subconjunto de bienes durables: vivienda y electrodomésticos del hogar. Esta elección sirve al práctico propósito de usar información que no es difícil de verifi-

car. Segundo, la información disponible da una descripción muy rica de las cualidades o características de este tipo de bienes. En efecto, incluye características relevantes de la vivienda, describiendo varias dimensiones sobre su construcción y el acceso y calidad de los servicios públicos, en varias *categorías* o niveles de calidad.

Sin embargo, la disponibilidad de información relevante sobre consumo en *forma multicategorica* no es del todo una bendición, puesto que trae complicaciones adicionales en el algoritmo usado para derivar el índice de utilidad: el algoritmo estadístico usado debe proveer una *métrica* o una *distancia* para las categorías cualitativas incluídas en la muestra. Además, para aquellas variables para las cuales tenemos una medida -e.g. ingreso corriente, años de escolaridad- no es deseable imponer ex-ante una relación lineal al índice de utilidad; por lo tanto debe permitirse más flexibilidad a esta relación funcional³⁵.

IV. La construcción del Índice de Utilidad por el Algoritmo Estadístico de Escala Optima: una solución al problema de la ponderación y el uso de información categórica sobre el consumo de bienes durables

La construcción del índice de utilidad propuesta en la sección anterior, no puede hacerse con Métodos de Análisis de Regresión tradicionales. Primero, los niveles de utilidad representa-

³⁴ En un país como Colombia, este proceso de participación de la fuerza laboral es más complejo e implica la selección de empleo como asalariado o como cuenta propia y con diferencias en los tamaños de planta. Estas diferencias están atadas a distintos niveles de exposición al riesgo, e.g.: cobertura de seguridad social en salud y / o pensiones.

³⁵ Por ejemplo, incluyendo varios rangos de valores de la variable, se producen diferentes formas del índice.

dos por U , V_1 y V_2 no son observables. Segundo, la mayoría de las variables de consumo de vivienda y otros bienes durables, son observados como categorías sin una métrica, entonces el análisis lineal de componentes principales lineal no aplica. Sin embargo, de la teoría del consumidor resumida en la subsección precedente, sabemos que existe un índice sobre la recta real que representa la utilidad de los hogares desde las demandas Marshallianas, o de ingreso corriente o de dotación de capital humano.

Afirmamos que el algoritmo de Mínimos Cuadrados Alternantes y Cuantificación Óptima -ALS-OS- (Componentes Principales Cualitativas) encuentra el índice de forma tal que (i) proporciona una métrica para diferentes categorías de las variables y (ii) les da un peso de acuerdo a su contribución a la *primera* componente principal del sistema. Precisamente, esa primera componente representa la dimensión de variabilidad máxima en los datos, que es el índice de utilidad no observable que debe ser construido. Observe que lo anterior descansa en el supuesto de que la principal fuente de variabilidad en la muestra está relacionada con las diferencias en ingreso y salud, y no sobre cambios en los gustos o cambios en los precios regionales.

El uso de variables determinadas simultáneamente o endógenamente como el ingreso corriente, el consumo y el capital humano, podría ser considerado un problema de "redundancia" y una fuente de multicolinealidad en una regresión. Sin embargo, en este caso es conveniente y ventajoso, dado que el proceso de optimización implícito del consumidor, vincula los diferentes subconjuntos de datos e incrementa las restricciones sobre el algoritmo para que converja al verdadero índice de utilidad.

Sabemos a priori que hay una estructura de covarianza entre estos tres conjuntos principales de variables -consumo, ingreso y dotación de capital humano- y eso es precisamente la fuerza que guía la asignación de la ponderación en el proceso de componentes principales.

El algoritmo usado involucra varios pasos, el primero y más importante procedimiento es llamado Cuantificación Óptima, el cual produce transformaciones no lineales de variables cualitativas, con el fin de maximizar el ajuste de los datos para un Modelo de Componentes Principales. Estas características específicas del procedimiento de transformación imponen el objetivo de producir un ordenamiento de los hogares en el contexto de un modelo de componente principal. Más específicamente, en la construcción del Sisben, El algoritmo ALS-OS fue restringido a privilegiar la componente con la varianza más alta en el sistema, es decir, la primera componente principal. Esto es consistente con el objetivo de construir el índice de bienestar postulado en la teoría del consumidor, que anticipa que la varianza máxima entre los hogares, sujeto a diferencias en la distribución del ingreso, está en la dirección de la utilidad o bienestar.

En la siguiente subsección describiremos los pasos técnicos involucrados con el algoritmo de Componentes Principales Cualitativas y terminaremos esta sección con una discusión de los resultados principales del Sisben para las áreas rurales y urbanas.

A. Componentes Principales Cualitativas: Prinqual y ALS-OS³⁶

El proceso Prinqual emplea un algoritmo basado en Mínimos Cuadrados Alternantes (ALS) y

Cuantificación Óptima (OS), para obtener transformaciones no lineales de un conjunto de variables cualitativas con el fin de maximizar el ajuste de los datos a un modelo de componentes principales lineales. El principio de OS considera cada observación como una categoría y representa cada categoría por un parámetro *de los datos*.

El principio ALS-OS divide el conjunto de parámetros en subconjuntos mutuamente excluyentes y exhaustivos: el primero consiste en parámetros del modelo de principales componentes -ponderaciones- y el segundo contiene los parámetros de los datos, los cuales son llamados *parámetros de cuantificación óptima*. A su vez, cada subconjunto puede contener varios subconjuntos mutuamente excluyentes y exhaustivos. El proceso de optimización se realiza encontrando las estimaciones de Mínimos Cuadrados de los parámetros en un subconjunto, fijando los parámetros en el otro.

Estas estimaciones son llamadas estimaciones de *Mínimos Cuadrados Condicionales*, ya que su naturaleza mínima cuadrática es condicional sobre los valores de los parámetros en el otro subconjunto. Cuando se obtienen las estimaciones de los Mínimos Cuadrados Condicionales -ya sea de las ponderaciones o de los parámetros de cuantificación óptima- los viejos estimadores son reemplazados por los nuevos. A continuación, el proceso continua para el siguiente subconjunto y son calculados sus estimaciones de Mínimos Cuadrados Condicionales. El proceso continua alternándose una vez en los parámetros

del modelo y otra en los parámetros de los datos, hasta que es obtenida la convergencia³⁷. Las fuerzas que conducen el proceso hacia la convergencia están dadas por la relación estructural entre una utilidad observable -representada por la primera componente principal- y la fortaleza de la contribución por las diferentes categorías en cada variable discriminatoria a la utilidad del hogar.

1. Componentes principales

Las componentes principales son nuevas variables generadas por combinaciones lineales (sumas ponderadas) de las variables originales. Las ponderaciones son funciones de la estructura de covarianza de las variables, y tienen varianza finita. Específicamente, para cada conjunto de ponderaciones, la suma de sus cuadrados es restringido a uno. La primera componente principal es la suma ponderada de las variables originales que tienen la mayor varianza. La segunda componente principal es una suma ponderada de las variables originales que tienen la segunda mayor varianza y que está correlacionada con la primera. La siguiente componente está correlacionada con la anterior y su varianza es menor o igual a la varianza de la anterior componente. La suma de las varianzas de las variables originales -llamada variación total- es igual a la suma de varianzas de los componentes principales.

Las Componentes Principales del modelo para los datos centrados en la media son:

³⁶ En esta subsección describiremos el Algoritmo de construcción de acuerdo a lo presentado en Castaño y Moreno (1994).

³⁷ Ver Anexos.

$$x_{ij} = f_{i1}w_{1j} + f_{i2}w_{2j} + \dots + f_{ic}w_{cj} + e_{ij} \quad (1)$$

para $i=1,2,\dots,n$ y $j=1,2,\dots,m$, donde n es el número de observaciones y m es el número de variables, x_{ij} es el i -ésimo valor individual en la j -ésima variable, w_{kj} es la ponderación de la j -ésima variable en el k -ésimo componente ($k=1,2,\dots,c$), donde el número del componente es c .

La primera variable componente principal puede ser empleada en lugar del conjunto completo de variables originales. Esto provee una representación más parsimoniosa de los datos y explica la mayor cantidad de la variación total del sistema, el índice utilidad.

Hay $n \times m$ ecuaciones como (1), una para cada observación. Se puede hacer una simplificación notacional empleando matrices. En este caso se puede representar (1) por:

$$X = FW' + E$$

Donde X es una matriz centrada de $n \times m$. Suponemos que las variables (columnas) tienen la misma unidad de medición. En el Análisis de Componentes Principales hay tanto número de variables como componentes principales. W es la matriz de $c \times m$ ponderaciones y E es la matriz de $n \times m$ errores. Cuando c es igual a m (hay tantos componentes como variables), la matriz E es una matriz nula. F es una matriz $n \times c$ que contiene los valores para los n sujetos sobre las c componentes principales. Si las variables son medidas en diferentes unidades, el análisis estará dominado por las variables con mayor varianza. En ese caso, las variables están estandarizadas.

De manera equivalente, la estructura de correlación de la matriz X puede ser empleada en lugar de la estructura de covarianza.

Note que las componentes principales son óptimas en el siguiente sentido. En un modelo truncado; i.e. un modelo con solo la primera componente principal, la varianza total de la componente está maximizada dadas las restricciones. Esto es, no existe otra componente que acumule una mayor proporción de la varianza total de las variables originales. En el caso de Sisben, el valor de c fue restringido a uno, y el modelo tenía solamente una componente principal. La variabilidad capturada por la *primera componente principal* representa el 49% de la varianza total, tanto en el modelo rural como en el urbano³⁸.

2. Componentes principales cualitativas

El proceso Prinqual (Componentes Principales Cualitativas) extiende el Análisis del Componente Principal ordinario a un contexto más general, en donde se puede emplear tanto variables cualitativas como cuantitativas. Este análisis emplea transformaciones no-lineales de las variables cualitativas, preservando el nivel de medición de la variable. Las variables ordinales son transformadas en forma monótona. Las variables nominales son transformadas de forma tal que se preserve la pertenencia de las observaciones en cada categoría de la variable. El proceso está basado en Kruskal y Shepard (1974), Young, Takane y de Leeuw (1978) y Winsberg y Ramsey (1983). Como se mencionó anteriormente, Prinqual emplea transformaciones no-lineales de las variables cualitativas para maxi-

³⁸ Castaño y Moreno (1995), p.42, 74.

mizar el ajuste de los datos al modelo de Componentes Principales Lineales.

3. Estabilidad y convergencia

El algoritmo para la derivación del sistema Sisben mostró una prueba satisfactoria de convergencia bajo diferentes niveles iniciales de la variable. Después de realizar escala óptima, el conjunto de variables sigue siendo multidimensional. Por lo tanto, se aplicó un Análisis de Factor Oblicuo para determinar si las variables estaban asociadas en grupos y si estos grupos pueden ser ordenados dependiendo de su contribución a la varianza total. Una explicación más detallada de éste procedimiento se encuentra en los Anexos.

B. Resultados: orden adecuado de categorías, cambio en la importancia de los grupos de variables en áreas rurales versus áreas urbanas

Los signos, valores y orden que el algoritmo asigna a las diferentes categorías de todas las variables, son consistentes con la intuición económica sobre consumo y pobreza. Los Cuadros 5A y 5B muestran la Cuantificación Óptima de las categorías de variables incluidas en el índice rural y urbano del Sisben. En el área urbana, la *escolaridad promedio del asalariado más viejo* entre primaria completa y secundaria incompleta obtuvo 3,8 puntos. Estos puntos podrían incremen-

tarse a 7,7 si ese mismo individuo tiene estudios superiores. La *escolaridad promedio para miembros del hogar con edad superior a los 12 años* fue similar y la escala relativa entre la máxima categoría y la mínima, tenían un rango comparable. Al *empleo, la cobertura de riesgo social y el tamaño de la planta del asalariado más viejo del hogar*, les fue otorgados valores en una escala re-lativa de 1 a 5,5.

La *densidad de personas en los cuartos* fue de 1,7 cuando el número de cuartos por persona era menor a 0,2, y de 5 cuando el número de cuartos por persona estaba entre 0,7 y 1,0. Los cambios en el *ingreso corriente* en salarios mínimos recibieron puntos adicionales en el Sisben con una escala de 1 a 10, además, la relación funcional es cóncava³⁹. En contraste con este resultado, la *proporción de miembros familiares empleados* tuvo una escala bastante menor, de 1 a 3,6, pero la relación se muestra convexa. Para todas las variables de características de la vivienda y de servicios, electrodomésticos y categorías de empleo y aseguramiento, el orden de las categorías reflejan racionalidad económica y no contradice la opinión de expertos en estos sectores.

Después de convergencia, el proceso Prinqual produjo tanto la cuantificación óptima como las ponderaciones óptimas de las variables sobre la primera componente principal⁴⁰. Finalmente, los puntos utilizados para calcular el puntaje del Sisben, son el resultado de multiplicar la cuantificación de cada categoría de las variables por el

³⁹ Disminuyendo el impacto marginal del ingreso corriente en el índice de bienestar del Sisben. Uno podría estar tentado a interpretar este resultado como utilidad marginal decreciente del ingreso, sin embargo debemos evitarlo, ya que el objetivo de este modelo es solamente predictivo.

⁴⁰ El indicador se obtiene usando el Factor de Análisis Oblicuo después de la Cuantificación Óptima y luego consiguiendo la Primera Componente Principal en cada grupo. Finalmente, se obtiene de la Primera Componente Principal de las primeras componentes de los cuatro grupos.

Cuadro 5A
CUANTIFICACIÓN ÓPTIMA DE LAS CATEGORÍAS DE CADA VARIABLE DEL SISBEN
(Área urbana)

Descripción de las categorías	Cuantificación	Media	No. de veces el mínimo	Descripción de las categorías	Cuantificación	Media	No. de veces el mínimo
Capital humano, seguridad social				Ingreso per cápita familiar en unidades de salario mínimo			
Escolaridad del mayor receptor				salario mínimo			
Sin Educación	1,187	-	1,00	De 0.50 a 0.75	5,1572	-	4,31
Primaria incompleta	2,0371	-	1,72	De 0.75 a 1.00	6,4268	-	5,37
Primaria completa	2,9897	3,57	2,52	De 1.00 a 1.25	7,2881	-	6,09
Secundaria incompleta	3,8066	-	3,21	De 1.25 a 1.50	8,2012	-	6,86
Secundaria completa	5,0313	-	4,24	De 1.50 a 2.00	8,7363	-	7,30
Superior Incompleta	6,3086	-	5,31	De 2.00 a 3.00	9,6152	-	8,04
Superior completa	7,2314	-	6,09	De 3.00 a 4.00	10,8613	-	9,08
Posgrados y Doctorados	7,7207	-	6,50	más de 4.00	11,3418	-	9,48
Escolaridad promedio personas de 12 y más años				Vivienda			
Sin educación	1,26	-	1,00	Material de las paredes			
(0, 4]	2,1382	-	1,70	Sin paredes o guadua, caña, u otros	1,4536	-	1,00
(4, 5]	2,8472	3,62	2,26	Materiales vegetales			
(5, 10]	3,8936	-	3,09	Zinc, tela, cartón, latas o deshechos	1,5986	-	1,10
(10, 11]	5,3086	-	4,21	Madera burda	2,6382	-	1,81
(11, 15]	6,2646	-	4,97	Bahareque	4,3018	-	2,96
(15, 16]	6,9258	-	5,50	Tapia pisada o adobe	5,1377	5,61	3,53
más de 16	7,1641	-	5,69	Bloque, ladrillo, piedra, material prefabricado, madera pulida	5,9863	-	4,12
Seguridad social mayor receptor				Material techos			
Sin seguridad social y trabaja sólo o no trabaja	1,0869	-	1,00	Paja o palma	0,6656	-	1,00
Sin seguridad social y trabaja en empresa de 2-9 empleados	1,9058	2,93	1,75	Deshechos (Cartón, lata, sacos, etc.)	2,0254	-	3,04
Sin seguridad social y trabaja en empresa de más de 10 empleados	2,9512	-	2,72	Zinc, asbesto, cemento, sin cielo raso	3,1069	3,55	4,67
Con seguridad social y trabaja sólo o no trabaja	3,8638	-	3,55	Teja, losa, zinc, cemento, con cielo raso	3,9595	-	5,95
Con seguridad social y trabaja en empresa de 2-9 empleados	5,1904	-	4,78	Material pisos			
Con seguridad social y trabaja en empresa de más de 10 empleados	5,9834	-	5,51	Tierra	0,9191	-	1,00
Demográfico, ingresos y educación				Equipamiento			
Proporción de menores de 6 años				Ningún básico			
Más de 0.65	1,8042	-	1,00	Hasta tres básicos sin lavadora	2,1504	2,2447	3,28
(0.0, .65]	1,9861	2,51	1,10	Cuatro básicos sin lavadora	2,8013	-	4,28
Sin niños	3,0044	-	1,67	Tres o cuatro básicos con lavadora	3,9476	-	6,03
Proporción de ocupados				Servicios			
Menos de 0.30	1,252	-	1,00	Abastecimiento de agua			
(0.30, 0.60]	1,7908	1,94	1,43	Río o manantial	1,2078	-	1,00
(0.60, 0.90]	2,347	-	1,87	Pila Pública u otra fuente	1,9734	-	1,63
Más de 0.90	4,4727	-	3,57	Pozo sin bomba, algibe, barreno o agua lluvia	2,9565	-	2,45
Hacinamiento (número de cuartos/personas familia)				Pozo con bomba			
Menos de 0.20	1,7537	-	1,00	carrotanque	4,2461	-	3,52
(0.20, 0.30]	2,1675	-	1,24	Acueducto	5,2793	5,6	4,37
(0.30, 0.40]	2,979	-	1,70	Eliminación de excretas	5,9961	-	4,96
(0.40, 0.70]	3,6602	4,34	2,09	No tiene servicio sanitario	1,157	-	1,00
(0.70, 1.00]	5,0059	-	2,85	Letrina	2,5537	-	2,21
(1.00, 4.00]	6,2031	-	3,54	Inodoro sin conexión a alcantarillado o pozo séptico	3,0552	-	2,64
Más de 4.0	7,9658	-	4,54	Inodoro conectado a pozo séptico	3,4136	4,45	2,95
Ingreso per cápita familiar en unidades de salario mínimo				Inodoro conectado a alcantarillado			
Menos de 0.15	1,196	-	1,00	Recolección de basura	5,0479	-	4,36
De 0.15 a 0.25	1,822	-	1,52	La tiran a un patio, lote, zanja, río o quebrada o laguna	0,9752	-	1,00
De 0.25 a 0.35	2,8081	-	2,35	La llevan a un contenedor cercano o basurero público	2,2881	2,59	2,35
De 0.35 a 0.50	3,8076	4,99	3,18	La recogen los servicios de aseo	2,9917	-	3,07

Fuente: Castaño y Moreno, (1994).

peso correspondiente de la variable. Este resultado más la tipificación de estas variables, -que suma de 0 como mínimo hasta 100 como máximo- son los puntajes presentados inicialmente en los Cuadros 1A y 1B. Como grupo, el capital humano y la categoría de empleo representan casi el 30% del total de las diferencias potenciales en el puntaje del Sisben. El Ingreso per cápita solo, es aproximadamente el 14% de las diferencias potenciales, y junto con las otras tres variables de su grupo acumulan casi el 30%. Los rangos más amplios son para la educación, el ingreso y el hacinamiento en alcobas; los siguen un grupo de cinco variables con un rango de valores similar: tipo de piso y pared, sistema de suministro de agua, tipo de empleo, seguridad social y tamaño de la empresa para el mayor perceptor del hogar. Finalmente, con un rango menor, los valores son -en orden descendente- materiales de techo, electrodomésticos, proporción de empleados, recolección de basura, y proporción de niños menores de seis años de edad⁴¹.

¿Las variables seleccionadas para el índice del Sisben son adecuadas para distinguir la situación real de los pobres y los no pobres?. Estudios anteriores sobre la distribución del ingreso y pobreza sugieren una respuesta positiva. Para economías latinoamericanas, Psacharopoulos et. al. (1997) identificaron el Capital Humano y la categoría del empleo como determinantes importantes en la determinación del empleo, in-

greso, inequidad y pobreza. En el caso Colombiano, la evaluación del Banco Mundial de 1995 reporta resultados similares. Grosh y Baker (1995) en su evaluación comparada de indicadores alternativos de bienestar para Jamaica, encontraron que las variables sobre características de la vivienda -en forma dicotómica- entran en una etapa más temprana en un proceso de regresión por pasos y contribuyen a reducir el error de inclusión en la focalización -error de tipo II. Adicionalmente, encontraron que las características familiares, incluyendo la dotación de capital humano, son muy poderosas para reducir el error de exclusión -error de tipo I. En el caso de Sisben, el conjunto de variables que caracterizan la vivienda fueron incluidas con información multicategoría, la cual se comportó bien tanto en el sector rural como en el urbano.

Los resultados revelan algunas comparaciones interesantes entre los sectores rurales y urbanos. En primer lugar, el sector rural tiene una variable adicional llamada *Tiempo empleado para la recolección de agua*. Como se explicó en la subsección anterior, el Análisis del Factor Oblicuo crea grupos de variables "cercanas"⁴² y ranquea aquellos grupos de acuerdo a su poder explicativo sobre la variabilidad total de los datos (ver Tablas A1.a y A1.b en Anexos). De hecho, Los grupos generados por el algoritmo son los mismos. Sin embargo, el potencial de estos grupos de variables como predictores de bienestar di-

⁴¹ El lector no debe concluir que las variables con rangos de puntaje mayores tienen necesariamente una mayor contribución a la variabilidad total de los datos o a la primera componente principal. Esta propiedad queda mejor establecida con los resultados del Factor Análisis.

⁴² Dentro del espacio de consumo, los grupos de variables "cercanas" pueden representar variables determinantes de funciones específicas de subutilidad. Y en este sentido, desde la agrupación presentada en las Tablas A1.a y A1.b podemos identificar dos funciones de subutilidad, una de calidad de la vivienda y otra de servicios de la vivienda.

fiere significativamente en ambas áreas. En el área urbana se distinguen claramente⁴³ cuatro grupos y en orden de importancia son: *i*) capital humano y seguridad social y empleo; seguido por, *ii*) Ingreso corriente, hacinamiento, proporción de miembros del hogar empleados y proporción de menores en el hogar; *iii*) materiales de construcción de la vivienda y aparatos domésticos; y, *iv*) la calidad de servicios públicos⁴⁴. Por contraste, en el área rural, la importancia de las características de las viviendas se incrementa y alcanza el primer nivel, mientras que el capital humano et. al. desciende al segundo lugar. Al mismo tiempo, el grupo de ingreso per cápita et. al., pasa del segundo al cuarto lugar, y servicios a las viviendas pasa del cuarto al tercer lugar. Estos resultados son consistentes con las observaciones de analistas rurales: el ingreso está sujeto a más inestabilidad y factores estacionales que pueden introducir algún tipo de ruido y reducir el poder que tiene el ingreso corriente para diferenciar los niveles de bienestar.

Como un índice aproximado de recursos, el índice Sisben pertenece a la clase de índices similares al OLS estimado para Jamaica. Sin embargo, en su algoritmo de construcción se diferencia claramente de los índices de bienestar diseñados explícitamente y ex ante, para maximizar la reducción de la pobreza⁴⁵. Las pruebas comparativas de Grosh y Baker (1995) de los indicadores aproximados de recursos frente a la

otra clase de indicadores fueron prometedoras, pues el indicador aproximado de recursos junto con un punto de corte adecuado, como el caso del Sisben, muestra resultados casi iguales tanto en el error de exclusión como en el de inclusión.

V. Conclusiones

Hemos mostrado que el indicador aproximado de recursos Sisben, se ha extendido rápidamente y es utilizado en muchos programas sociales en la mayoría de los municipios de Colombia. Los principales factores que impulsaron esta exitosa historia de descentralización parecen ser el bajo costo marginal de monitoreo para las autoridades regionales y los evidentes beneficios que los programas sociales financiados por la Nación pueden traer a las comunidades locales.

El algoritmo utilizado para derivar las ponderaciones y la cuantificación de las variables incluidas en el Sisben, Mínimos Cuadrados Alternantes y Cuantificación Óptima, es consistente con el desafío de construir un índice de bienestar para una variable no observable, utilizando datos categóricos. Los resultados mostraron que de la variabilidad total en la muestra, la parte que el índice Sisben puede explicar es un poco menos del 50%.

El Sisben incorporó valiosos datos cualitativos sobre el consumo de bienes durables, como

⁴³ Note que en la Tabla A1.a los valores de la matriz del Factor Análisis que no están en negrilla son relativamente iguales a cero.

⁴⁴ Finalmente note que tan cercano -coeficiente 0.2727- es Hacinamiento al tercer factor que corresponde a las características del hogar.

⁴⁵ Ravallion y Chao (1989).

las características de construcción de las residencias y la disponibilidad de servicios públicos. Los resultados muestran que es útil distinguir varios niveles de calidad dentro de estas variables: sus diferencias métricas acumulan aproximadamente el 40% del total de puntos del puntaje Sisben.

Aunque el Sisben produce conjuntos similares de variables "cercanas" en el sector urbano y rural, el agrupamiento es mucho más claro o muestra menos ruido en el área urbana que en la rural. Adicionalmente, el potencial de este grupo de variables como predictores de bienestar difiere significativamente en ambas áreas. Mientras en el sector urbano el capital humano, el tipo de empleo, estrato, y el ingreso corriente son de gran importancia, en las áreas rurales estas señales generan más ruido que las características de la vivienda. Mas que esto, en las áreas rurales, el grupo que incluye empleo, estrato e ingreso pasa de segundo lugar a cuarto en importancia para explicar la variabilidad total de los datos.

Aunque el ingreso fue incluido en el conjunto inicial del Sisben de acuerdo con un relativo consenso en 1993 entre expertos colombianos en el tema de la pobreza, los fuertes incentivos de los hogares para sub reportar el ingreso para obtener rangos favorables, sugiere que esta variable debería ser removida si se construyera un nuevo índice de bienestar. Además, si el ingreso se utilizara como la medida de oro del bienestar -evitando los supuestos sobre una función de utilidad paramétrica- un método alternativo para construir un índice de bienestar sería en-

tonces construir una función discriminante con el fin de cuantificar todas las variables que entran en forma categórica.

Si aceptamos que el procedimiento de Cuantificación Óptima es adecuado para incorporar datos categóricos relevantes para predecir niveles de bienestar, el siguiente paso lógico sería aprovechar la riqueza de los datos categóricos en los censos en América Latina, para mejorar la calidad y nivel de detalle de la información utilizada en la focalización regional. Por ejemplo, tomando el mismo caso de Colombia, algunas transferencias del nivel nacional al regional se basan en indicadores de bienestar regional. La imposibilidad de obtener información confiable sobre ingresos a nivel regional, obliga a las autoridades a utilizar indicadores como el NBI, obtenido de los datos del censo realizado a los hogares. Sin embargo, con base en el mismo conjunto de datos del censo, un mejor predictor de bienestar puede ser derivado utilizando el Algoritmo de Cuantificación Óptima y, como consecuencia, se pueden obtener mayores ganancias en equidad con una cantidad dada de transferencias compensatorias entre regiones. En el momento, el Departamento Nacional de Planeación en Colombia ha derivado unos indicadores de ese tipo que podrían eventualmente reemplazar el índice NBI en las fórmulas para calcular las transferencias de recursos inter gubernamentales⁴⁶.

En este trabajo hemos evaluado algunos aspectos importantes del sistema Sisben, sin embargo, no hemos estudiado todas las áreas im-

⁴⁶ Ver Sarmiento, et al (1996, 1997) y Castaño (1995). Este tipo de indicadores puede ser utilizado también con objetivos geográficos dentro de las entidades territoriales.

portantes. Es necesario que el Sisben sea evaluado ex-post en términos de la eficiencia en la focalización. Además, hasta el momento, las ganancias asociadas con el procedimiento de Análisis de Componentes Principales Cualitativas no ha sido medido. Por lo tanto, *i)* deben realizarse algunas simulaciones para medir las pér-

didias en el modelo del componente principal cuando la información cualitativa se deteriora, colapsando en dos categorías. *ii)* Adicionalmente, la capacidad del Sisben para predecir pobreza debe ser comparada explícitamente con índices derivados de otros algoritmos, como el índice aproximado de recursos OLS y otros⁴⁷.

⁴⁷ Similar al tests de Grosh y Baker (1994) para el caso de Jamaica.

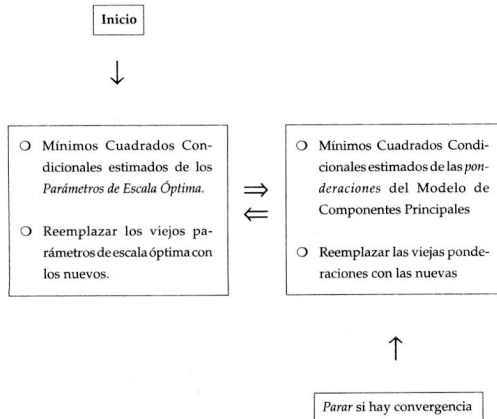
Bibliografía

- Anand, Sudhir y Christopher J. Harris (1994), "Choosing a Welfare Indicator" *AEA Papers and Proceedings*, vol 84 (2) pp.226-231.
- Behrman y Srinivasan, eds.(1995), "Handbook of Development Economics", volume III. Elsevier B.V.
- Castaño, Elkin y Hernando Moreno (1994), "Metodología Estadística del Modelo de ponderaciones del sistema de Selección de Beneficiarios de Programas Sociales (Sisben)". Mimeo, DNP-Misión Social Mayo de 1994.
- Castaño, Elkin (1995), "Metodología para predicción del indicador de pobreza Sisben en términos de las variables observadas en el censo de 1993". Mimeo, DNP-Misión Social, Santafé de Bogotá.
- CEPAL (1995), "Experiencias Recientes de Selección de Beneficiarios en Costa Rica: una Evaluación del Sisben". Mimeo.
- De Boor, C. (1978), "A Practical Guide to Splines", New York: Springer Verlag.
- Deaton, Angus (1979), "The Distance Function in Consumer Behavior with Application to Index Numbers". *Review of Economic Studies*, vol XLVI (3). pp. 391-405.
- _____(1995), "Data and Econometric Tools for Development Analysis". in (eds.) Behrman and Srinivasan.
- _____(1997), "The Analysis of Household Surveys: A microeconomic approach to development policy". World Bank, Washington DC.
- ____y Muellbauer, John (1980), "Economics and Consumer Behavior". Cambridge University Press.
- Departamento Nacional de Planeación (1997), "Focalización del Gasto Social". Documento Conpes. Santafé de Bogotá.
- DNP-Misión Social (1997), "¿Qué es el ICV? Índice de Condiciones de Vida". Mimeo, DNP-Misión Social octubre.
- DNP-UDS (1995), "La Estratificación Socioeconómica: Avances y Retos". Mimeo.
- DNP-UDS-Misión Social (1997), "Focalización del Gasto Social". Mimeo.Bogotá: DNP.
- DNP-UNDP (1996), "Evaluación de la Etapa de Implantación del Sistema de Selección de Beneficiarios para Programas Sociales -Sisben-". Mimeo. Bogotá: PNUD-DNP.
- Fisher, R. (1938), "Statistical Methods for Research Workers", 10th. ed., Edinburg: Oliver and Boyd Press.
- Glewwe, Paul y Jacques van der Gaag (1988), "Confronting Poverty in Developing Countries". *Living Standards Measurement Study Working Paper No. 48*. World Bank, Washington, D.C.
- ____y O. Kanaan (1989), "Targeting Assistance to the Poor: A multivariate Approach Using Hogar Survey Data". *Policy, Planning and Research Working Paper No. 225*, World Bank, Washington, D.C.
- Grosh, Margaret y Judy Baker (1995), "Proxy Means Tests for Targeting Social Programs: Simulations and Speculation" *Living Standard Measurement Study Working Paper No. 118*. World Bank.
- ____(1994), "Administering Targeted Social Programs in Latin America: From Platitudes to Practice". Washington D.C.: The World Bank.
- ____y Glinskaya, Elena (1998), "Proxy means and testing social assistance in Armenia". Mimeo. Washington DC. The world Bank.
- Hotelling, H. (1933), "Analysis of Complex Statistical Variables into Principal Components", *Journal of Educational Psychology*, 24, 498-520.
- Kruskal, J.B. & Shepard, R.N. (1974), "A Nonmetric Variety of Linear Factor Analysis", *Psychometrika*, 38, 123-157.
- Kuhfeld, W.F., Sarle, W.S. & Young, F.W. (1985), "Methods of Generating Model Estimates in the PRINQUAL Macro", *SAS Users Group International Conference Proceedings: Sugi 10*, Cary, NC: SAS Institute, 962-971.
- Mas-Colell Andrew, Whinston Michael y Green Jerry (1995), "Microeconomic Theory". New York: Oxford University Press.
- Psacharopoulos, George ; Morley, Samuel ; Fiszbein, Ariel ; Lee, Haeduck y Wood, Bill (1997), "La pobreza y la distribución de los ingresos en América Latina. Historia del decenio de 1980". Documento Técnico del Banco Mundial, Número 351S. Banco Mundial Washington, D.Cz.
- Ravallion, Martin y Kalvin, Chao (1989), "Targeted Policies for Poverty Alleviation under Imperfect Information: Algorithms and Applications". *The Journal of Policy Modeling*, 11(2):213-224.
- ____, (1992), "Poverty Comparisons: Guide to Concepts and Methods" *Living Standard Measurement Study Working Paper No. 88*. World Bank.

- ____y Lipton, Michael (1995), "Poverty and policy". In (eds.) Behrman and T.N. Srinivasan.
- Richards, P. y Leonor, M., eds. (1982), "Target setting for basic needs". Geneva: International Labor Office.
- Saporta, Gilbert (1983), "Multidimensional Data Analysis and Quantification of Categorical Variables". in J. Janssen and J.-F. Marcotorchino (eds.) *New Trends in Data Analysis and Applications* Elsevier Science Publishers B.V.
- Sarmiento Gómez, Alfredo, Ramírez Gómez, Clara, Alonso, Carlos Eduardo y Acosta Trujillo, Rodrigo (1997), "Tipología Municipal con base en las Condiciones de Vida". Mimeo, DNP-Misión Social. Santafé de Bogotá.
- ____, Ramirez, Molina y Castaño (1996), "Índice de Condiciones de Vida". Mimeo, DNP-Misión Social, Santafé de Bogotá, julio.
- SAS/STAT User Guide (1990), Volume 2, Version 6, Cuarta edición.
- Sen, A. (1979), "Persona al utilities and public judgements: or what's wrong with bienestareconomics?", *The Economic Journal*, 89:537-558.
- ____ (1987), *The Standard of Living*. Cambridge: Cambridge University Press.
- ____ (1985), *Comodities and Capabilities*. Amsterdam: North Holland.
- Streeten, P., Burki, S.J., ul Haq, M., Hicks, N. and Stewart, F. (1981), *first things first: meeting basic Necesidadesin developing countries*. New York: Oxford University Press.
- Tenenhaus, M. & Vachete, J.L. (1977), "Prinqual: A programme d'Analys in Principals D'un Ensemble de Variables Nominales ou Numeriques, *Les Cahiers de Recherche #68*, CESA, Jouy-en-Josas, France.
- Varian, Hal R. (1992), "Microeconomic Analysis". New York: W.W. Norton & Company, Inc.
- Vélez, C.E. (1996), "Gasto Social y Desigualdad : Logros y Extravíos". Bogotá: DNPlaneación- Misión Social.
- Vélez, C.E. (1997), "Eficiencia, Equidad y Reestructuración Sectorial del Gasto Público Social". En Cardenas Mauricio y Nora Lustig (eds.) *Pobreza y Desigualdad en America Latina. Articulos de la conferencia LACES 1997*. Bogotá: Tercer Mundo, 1998.
- Vélez, Carlos Eduardo, Elkin Castaño y Daniel Correal. (1995), "Consideraciones sobre la Aplicación de los puntos de corte Sisben". Mimeo. Junio de 1995.
- Winsberg, S. & Ramsay, J.O. (1983), "Monotone Spline Transformations for Dimension Reduction", *Psychometrika*, 48, 575-595.
- Wold, H. & Litkens, E. (1969), "Nonlinear Interactive Partial Least Squares (NIPALS) Estimation Procedures", *Bulletin ISI*, 43, 29-47.
- Young, F.W. (1975), "Methods for Decribing Ordinal Data with Cardinal Models", *Journal of Mathematical Psychology*, 12, 416-436.
- ____ (1981), "Quantitative Analisis of Qalitative Data", *Psychometrika*, 46, 357-388.
- ____, Takane, Y. & de Leeuw, J. (1978), "The Principal Componente s of Mixed Measurement Nivel Multivariate Data: An Alternating Least Squares Method with Optimal EscalaFigures", *Psychometrika*, 43, 279-281.
- ____, Takane, Y. & de Leeuw, J. (1985), "PROC PRINQUAL- Preliminary Specifications". Manuscrito no publicado, The University of North Carolina Psychometric Laboratory, Chapel Hill NC.
- Zarta, Maria Teresa (1997), "Avance y Cobertura de la Encuesta Sisben por Departamentos y Municipios". Mimeo, DNP-Misión Social julio.

Anexos

MÍNIMOS CUADRADOS ALTERNATIVOS Y ESCALA ÓPTIMA



El método de Varianza Total Máxima - VTM

Basado en Young, Takane y Leeuw (1978), el Modelo de Componentes Principales maximiza la suma de los primeros valores c de la matriz de covarianzas. Este método transforma las variables para que sean lo más similares posible a combinaciones lineales de variables puntaje componente principal c , donde c puede ser mucho más pequeño que el número de variables. Esto maximiza la varianza total de las primeras componentes principales c . En la construcción del Sisben, c fue restringido a uno, i.e. la primera componente principal. Sobre cada iteración, el algoritmo VTM alterna el Análisis de Componente Principal clásico (Hotelling, 1933) con el de Escala Óptima (Young, 1981)¹.

Escala Óptima en PRINQUAL

PRINQUAL usa la función OPSCAL PROC MATRIX (Instituto SAS, 1983) para hacer la escala óptima. La función OPSCAL permite escalar de manera óptima las variables nominales usando el Método de transformaciones de puntaje óptimo de Fisher (1938). Las variables ordinales son re-escaladas usando la transformación monotónica Mínimos Cuadrados de Kruskal y Shepard (1974). Las variables Intervalo pueden ser re-escaladas usando transformaciones no óptimas como transformaciones de poderes.

Análisis de Factor Oblicuo

Una vez se ha hecho la cuantificación de las categorías de todas las variables por el procedimiento PRINQUAL, se usa la técnica de Análisis de Factores para identificar subgrupos de variables que son "más cercanas", y al mismo tiempo, para encontrar si existe correlación entre los subgrupos. Este método es llamado Análisis de Factores Oblicuo, que usa el factor de rotación Promax (SAS).

Luego, usando el Análisis de componente principal con cada subgrupo identificado por el Análisis de factores, se extrae su primera componente principal. Subsecuentemente, usando cada uno de los subgrupos de las componente principal, se construye el indicador de bienestar global como primera componente.

Finalmente, el indicador de bienestar SISBEN toma valores entre 0 y 100. La Tabla A3 resume el procedimiento después del primer paso, donde X_k , $k=1,2,\dots,m$, la k -ésima variable discriminante, ha sido transformada en TX_k gracias al uso del ALS-OS por el comando PRINQUAL SAS. Es importante anotar que este procedimiento puede ser simplificado: si no hay interés en detectar subconjuntos de variables discriminantes "cercanas", el proceso determinaría inmediatamente las ponderaciones de cada variable discriminante del primer componente principal. Lo cual es equivalente a saltar del paso 1 al paso 5 en el Anexo 2.

La Tabla A.1A muestra los resultados de aplicar el procedimiento sobre las variables cuantificadas en áreas urbanas. El procedimiento identifica cuatro grupos de variables "cercanas" en el sector urbano. El primer factor incluye educación y variables de seguridad social; el segundo, variables demográficas y de ingreso; el tercer factor incluye variables de vivienda y equipamiento; y el último factor está formado por las variables de servicios de la vivienda.

La Tabla A1.B muestra los resultados de la aplicación del procedimiento para las variables cuantificadas en el sector Rural. Encontramos cuatro grupos de variables. El primero contiene variables de vivienda y equipamiento; el segundo, variables de capital humano y seguridad social; el tercer grupo está conformado por servicios del hogar y el último por variables demográficas y de ingreso.

¹ Kuhfeld, Sarle y Young (1985) presentan una descripción del método.

Tabla A1.a
ESTRUCTURA DE LOS FACTORES
Sisben urbano

Método de Rotación: Promax

Variable	Factor 1	Factor 2	Factor 3	Factor 4
ESMXMAYA	0,90836	-0,03175	-0,4792	-0,00773
ESP12YM	0,88053	-0,01386	0,01269	-0,00341
SEGURO	0,87049	-0,07371	-0,05914	0,09877
PRPO6	-0,25828	0,79564	0,06812	-0,03747
PROPOCUP	0,03787	0,78125	-0,22825	0,03544
HACICUAR	0,09804	0,62166	<u>0,27275</u>	-0,05088
INGPERC	0,48271	0,52543	0,05714	0,01918
MATPAR	-0,09836	-0,05693	0,69726	0,11222
MATTECH	-0,12761	0,01911	0,69601	0,03520
MATPIS	0,13294	0,08420	0,62003	0,11964
EQUIPA	0,33095	-0,11653	0,55854	-0,19687
ABAGUA	-0,03300	-0,05035	-0,07300	0,81253
EXCRET	0,02261	0,02240	0,16428	0,75977
RECOBA	0,08302	0,00346	0,04283	0,73218

Fuente: Castaño y Moreno (1995).

Tabla A1.b
ESTRUCTURA DE LOS FACTORES
Sisben rural

Método de Rotación: Promax

Variable	Factor 1	Factor 2	Factor 3	Factor 4
MATTECH	0,75109	0,11833	0,31109	0,14576
MATPAR	0,79025	0,31929	0,31182	0,11061
MATPIS	0,80067	0,37131	0,42947	0,21265
EQUIPA	0,64206	0,44566	0,34483	0,09350
ESMXMAY	0,25272	0,88245	0,21910	0,11367
ESP12YM	0,31706	0,88886	0,25764	0,13537
SEGURO	0,34258	0,58543	0,24456	0,06216
ABAGUA	0,36576	0,21634	0,90109	0,06815
TIEMAGU	0,38883	0,19393	0,88070	0,10681
EXCRET	0,58471	0,47239	0,70688	0,17092
RECOBA	0,28164	0,43371	0,54987	0,07170
PROPOCU	0,02521	0,00170	-0,00390	0,82308
PROP6	0,01670	-0,02780	0,04722	0,74527
HACICUA	0,32544	0,23509	0,14693	0,74289
INGPERC	0,39136	0,47228	0,23786	0,66320

Fuente: Castaño y Moreno (1995).

Tabla A2.a
PONDERACIÓN DE LA VARIABLES CON RESPECTO A LOS FACTORES Y DE LOS FACTORES CON
RESPECTO AL MODELO (Primera componente principal) (Sisben urbano)

Factor	Código de la variable	Ponderación de la variable	Ponderación	Ponderación total de la variable
Vivienda	EQUIA	0,5961	0,7974	0,4753
	MATPAR	0,7095	-	0,5658
	MATPIS	0,7774	-	0,6199
	MATTECH	0,6437	-	0,5133
Servicios	EXCRET	0,8573	0,6792	0,5823
	ABAGUA	0,7400	-	0,5026
	RECOBA	0,7920	-	0,5379
Capital humano y seguridad social	ESP12YMA	0,8993	0,6959	0,6258
	ESMXMAYA	0,9105	-	0,6336
	SEGURO	0,6786	-	0,4723
Demografía, ingreso y ocupación	PROP60.6792	0,6792	0,6006	0,4079
	PROPOCUP	0,6884	-	0,4135
	HACIGUAR	0,7452	-	0,4476
	INGPERC	0,7478	-	0,4491

Fuente: Castaño y Moreno (1995).

Tabla A2.a
PONDERACIÓN DE LA VARIABLES CON RESPECTO A LOS FACTORES Y DE LOS FACTORES CON
RESPECTO AL MODELO (Primera componente principal) (Sisben rural)

Factor	Código de la variable	Ponderación de la variable	Ponderación	Ponderación total de la variable
Vivienda	EQUIA	0,6775	0,8094	0,5483
	MATPAR	0,7963	-	0,6445
	MATPIS	0,8129	-	0,6579
	MATTECH	0,7293	-	0,5903
Servicios	EXCRET	0,7871	0,7816	0,6152
	ABAGUA	0,8580	-	0,6706
	TIMEAGU	0,8335	-	0,6515
	RECOBA	0,6207	-	0,4851
Capital humano y seguridad social	ESP12YMA	0,9112	0,7035	0,6410
	ESMXMAYA	0,9135	-	0,6426
	SEGURO	0,6028	-	0,4241
Demografía, ingreso y ocupación	PROP60.6792	0,7126	0,4646	0,3310
	PROPOCUP	0,7999	-	0,3716
	HACIGUAR	0,7625	-	0,3542
	INGPERC	0,7148	-	0,3321

Fuente: Castaño y Moreno (1995).

Tabla A3
METODOLOGÍA DE CONSTRUCCIÓN DEL ÍNDICE SISBEN

	Input	Procedimiento	Output
1.	X_k , $k=1, \dots, m$, conjunto completo de variables	Usar PRINQUAL sobre el conjunto completo de variables	<i>Variables transformadas</i> TX_k , $k=1, \dots, m$ cuantificación
2.	<i>Variables transformadas</i> TX_k , $k=1, \dots, m$	Usar FACTOR sobre <i>Variables transformadas</i>	Subgrupos p de variables "cercanas"
4.	subgrupos p de <i>variables transformadas</i> TX_{1j}, \dots, TX_{mj} para cada subgrupo $i, j=1, 2, \dots, p$	<i>Análisis de Componente Principal</i> sobre variables de cada subgrupo	Primera componente Principal CPI_j para cada subgrupo $j=1, 2, \dots, p$
5.	Primer componente Principal CPI_j para cada subgrupo $CPI_{1j}, \dots, CPI_{pj}$	Análisis de Componente principal para el CP de cada subgrupo	Primera componente Principal o indicador de Bienestar Global, CPI_G
6.	CPI_G	Cambio de origen y escala	Sisben INDEX Indicador de Bienestar Global Tipificado

Tabla A4
LISTA DE CÓDIGOS DE VARIABLES

ESMXMAYA	Escolaridad del asalariado más viejo
ESP12YMA	Escolaridad media de las personas de 12 años y más
SEGURO	Seguridad social y tamaño de planta
PROP6	Proporción de niños menores de 6 años
PROPOCUP	Proporción de miembros de la familia empleados
HACICUAR	Número de cuartos por persona
INGPERC	Ingreso per cápita en salarios mínimos
MATPAR	Material de las paredes
MATTECH	Material predominante para el techo
MATPIS	Material predominante en el piso
EQUIPA	Herramientas del hogar
ABAGUA	Sistema de bastecimiento de agua
TIEMAGU	Tiempo gastado en conseguir gas
EXCRET	Manejo de excretas
RECOBA	Manejo de la basura