

# **An Economic Interpretation of Colombia's SISBEN: A Composite Welfare Index Derived from the Optimal Scaling Algorithm**

**By Carlos Eduardo Vélez, Elkin Castaño and Ruthanne Deutsch**

November, 1998

## **ABSTRACT**

SISBEN is a proxy means test index widely used as a targeting system for social programs in Colombia. It serves as an indicator of households' economic well-being and is based on a cross-section sample of Colombian households. The SISBEN index is a function of a set of variables related to the consumption of durable goods, human capital endowment and current income. SISBEN was created with the purpose of simplifying, expediting and reducing the cost of targeting individual beneficiaries of social programs at the various government levels. The system has been very successful and has generated high demand both in municipalities and departments, presumably due to its low operation costs and also to the social benefits being brought to participating communities. SISBEN has expanded quite rapidly. It is currently used in a wide range of subsidized social programs, most heavily by those that benefit from health subsidies established by Colombia's Social Security Law.

This article examines the conformity between the SISBEN index and the degree of economic well-being of beneficiary households, and evaluates -ex post- the adequacy of the specific statistical methods used in terms of that objective. Of particular relevance is the treatment given to an important set of categorical information about household consumption of durable goods.

This analysis allows the authors to conclude that SISBEN represents a composite index of alternative forms of the household utility function or welfare level. The use of "redundant" information on consumption, income and human capital endowments -which is endogenously determined- is necessary and useful, given that the utility of households is unobservable and traditional statistical methods of model fitting do not apply when categorical information is heavily used. The method used, "Optimal Scaling and Alternate Least Squares" solves the statistical challenges as it imposes a metric to the categorical variables that are used to compose the index in order to maximize their contribution to a model of principal components. In addition, in the case of SISBEN the ALS-OS algorithm was forced to concentrate the optimal scaling adjustments according to their contribution to the *first* principal component, which precisely represents the utility index of every household. For the data set -from CASEN household survey, the ALS-OS algorithm converges and is stable to different starting configurations. The procedure selects a relevant set of variables to identify the poor and the signs, values and the order that this algorithm assigns to the different categories of all variables are consistent with economic intuition on consumption and poverty.

**An Economic Interpretation of Colombia's SISBEN :  
A Composite Household Welfare Index  
Derived from the Optimal Scaling Algorithm**

**By Carlos Eduardo Vélez, Elkin Castaño and Ruthanne Deutsch\***

**Introduction**

This paper provides an economic interpretation of SISBEN – a proxy means test index widely used to select beneficiaries of social programs in Colombia since 1994.<sup>1</sup> It serves as an indicator of households' economic well-being and was based on a cross-section sample of Colombian households collected in 1993.<sup>2</sup> The SISBEN index is a function of a set of variables related to the consumption of durable goods, human capital endowments and current income. SISBEN was created with the intention of simplifying, expediting and reducing the cost of targeting individual beneficiaries of social programs at the various government levels. The system has been very successful and has generated high demand both in municipalities and departments, presumably due to its low cost of operation and also to the social benefits being brought to participating communities. SISBEN has expanded quite rapidly and it is currently used in a wide range of

---

\* Velez and Deutsch, from the Poverty and Inequality Advisory Unit at the Interamerican Development Bank and Castaño from Universidad de Antioquia and Misión Social-National Planning Department, Colombia. The authors want to thank Nancy Birdsall for her valuable comments and suggestions. We also express our gratitude to Maria Teresa Zarta from Misión Social for providing valuable data on the implementation of the Sisben program and to Luis Tejerina at the Poverty Unit- IDB for his valuable assistance. .

<sup>1</sup> SISBEN was developed at Misión Social, Departamento Nacional de Planeación, Colombia. See Castaño and Moreno (1994) for a presentation of the results and the algorithm of construction.

<sup>2</sup> CASEN.

subsidized social programs, particularly by those that benefit from health subsidies established by Colombia's Social Security Law and housing subsidies from INURBE.

Our main question is whether the Optimal Scaling algorithm used to derive the SISBEN Index is consistent with the economic principles underpinning a proxy means test welfare index. Our analysis leads to the conclusion that SISBEN represents a composite index of alternative forms of the household utility function –a measure of welfare- and the algorithm used for its construction solves the challenges posted by the data. The unobservability of utility does not allow for the application of traditional methods of regression to construct the index. In addition, the inclusion of detailed categorical information on household consumption of durable goods precludes the use of linear principal component analysis and therefore alternative algorithms are required. The method used, “Alternate Least Squares and Optimal Scaling” –ALS OS- solves the difficulties with the data form and the type of problem: it imposes a metric to the variables that are used to compose the index in order to maximize their contribution to a model of Principal Components. In addition, in the case of SISBEN the algorithm was forced to concentrate the optimal scaling adjustments according to their contribution to the *first* principal component, which is the dimension of maximum variability and represents the postulated composite welfare index. The use of information on consumption, income and human capital endowment –which are endogenously linked- while at first glance appearing to be “redundant”, actually turns out to be necessary and useful, given the type algorithm implemented. The signs, values and the order that this algorithm assigns to the different categories of all variables are consistent with economic intuition on consumption and poverty.

The following section of the paper describes SISBEN – its principal objectives, scope and coverage, role in determining the allocation of program resources, costs, and factors that account for its rapid and successful implementation in Colombia. Section 2 undertakes an economic analysis of SISBEN, (i) examining the consistency of alternative welfare indexes derived from consumer theory and the data set used to derive this specific welfare index and (ii)

evaluating -ex post- the adequacy of the Optimal Scaling Algorithm used to derive the true welfare index. The next section describes the algorithm of Qualitative Principal Components' Analysis –that includes Alternating Least Squares and Optimal Scaling- used to derive the SISBEN index, and details a series of methodological issues that were addressed during the preparation of SISBEN. It also reviews the results on scaling and weights and provides comparisons for urban and rural areas. Finally we summarize the main results and offer some concluding thoughts.

## 1. What is SISBEN?

### *Objectives*

SISBEN – *El Sistema de Selección de Beneficiarios para Programas Sociales* –(System for the Selection of Beneficiaries of Social Programs) was designed to provide local governments with a cost-effective technical instrument that could be used for the targeting of social expenditures to the most poor and vulnerable groups. Designed within the context of an overall decentralization effort within the social sectors following upon the Constitutional Reform of 1991, SISBEN aims to achieve not only accuracy in targeting, but also seeks ease of implementation for local governments with limited administrative capacity.

The overall objective was to obtain reliable socio-economic information on poor and vulnerable groups of the Colombian population to be used in the targeting of social programs -- particularly the newly reformed subsidized public health system.

Specific objectives include:<sup>3</sup>

- Expedite the targeting of beneficiaries for subsidized social programs in a transparent fashion, based on uniform, technical and equitable criteria.
- Support interinstitutional coordination at the municipal level in order to improve the impact of social spending and eliminate duplication in program benefits.

---

<sup>3</sup> See DNP-UNDP (1996).

- Allow for the evaluation of targeted spending goals at the departmental, district, and municipal level in order to make a significant impact in the fight against poverty in Colombia.
- Provide inputs for the preparation of detailed socioeconomic diagnostic studies and poverty profiles usable for the elaboration of municipal social development plans and the design and implementation of specific programs and projects targeted to the neediest; and
- Contribute to institutional strengthening at the municipal level, through the start-up and training in the implementation of a modern management information system;

#### ***How SISBEN Works***

SISBEN functions essentially as a proxy means test for households that are potential beneficiaries of social programs. A household welfare index of qualitative and quantitative socioeconomic variables was derived using the statistical algorithm of Qualitative Principal Components or PRINQUAL -outlined in Section 3-, working with the CASEN survey of 25,000 households. The final subset of variables selected by the algorithm were determined as most efficient in predicting household welfare differences and poverty.<sup>4</sup> They are easily collected in a short interview by applying *la Ficha de Caracterización Socioeconómica* –Socioeconomic Classifications Form. With the aid of a customized software, the SISBEN score for a household is readily calculated and a ranking and a stratification of potential beneficiaries was obtained.<sup>5</sup> Table 1 provides a description of the variables collected in the SISBEN questionnaire, and the alternative number of points assigned to each answer category in order to calculate the SISBEN score of the household .

---

<sup>4</sup> The final set of variables was smaller than the larger set of variables with which the process started. This set included all the suggestions by poverty experts in Colombia. The algorithm selected the final subset according to their contribution to the objective function of the process (see section 3 below).

<sup>5</sup> In most cases, local communities were allowed to review the selection of beneficiaries in order to detect non poor families mistakenly classified.

Table 1.a Score per answer category in each variable. SISBEN, Urban, 1993

VARIABLES CATEGORY DESCRIPTION	SCORE(*)
<b>HUMAN CAPITAL, SOCIAL SECURITY AND FIRM SIZE FACTOR</b>	
<b>Schooling of oldest wage earner</b>	
No education	0
Some elementary	1.6239
Complete elementary	3.4435
Some secondary	5.0039
Complete secondary	7.3434
Some higher education	9.7833
Complete higher education	11.5460
Graduate studies	12.4806
<b>Mean schooling for people 12 years old and older</b>	
0	0
4	1.6570
5	2.9947
10	4.9690
11	7.6387
15	9.4425
16	10.6900
More than 16	11.1396
<b>Social Security and plant size</b>	
Without Social Security and works alone or doesn't work	0
Without Social Security and works in a plant with 2 to 9 employees	1.1660
Without Social Security and works in a plant with 10 or more employees	2.6545
With Social Security and works alone or doesn't work	3.9539
With Social Security and works in a plant with 2 to 9 employees	5.8427
With Social Security and works in a plant with 10 or more employees	6.9718
<b>DEMOGRAPHIC, UNEMPLOYMENT, INCOME AND ROOM CROWDING</b>	
<b>Proportion of children below 6 years old</b>	
Greater than 0.65	0
From 0 to .65	0.2237
Cero	1.4761
<b>Proportion of family members employed</b>	
Less than 0.30	0
From 0.30 to 0.60	0.6717
From 0.60 to 0.90	1.7390
Greater than 0.90	4.0149
<b>Room Crowding (Number of rooms per person)</b>	
Less than 0.20	0
0.20 to 0.30	0.5584
0.30 to 0.40	1.6535
0.40 to 0.70	2.5727
0.70 to 1.00	4.3886
1.00 to 4.00	6.0042
Greater than 4.00	8.3828
<b>Per capita income in minimum wage units</b>	
Less than 0.15	0
From 0.15 to 0.25	0.8476
From 0.25 to 0.35	2.1828
From 0.35 to 0.50	3.5362
From 0.50 to 0.75	5.3636
From 0.75 to 1.00	7.0827
From 1.00 to 1.25	8.2489
From 1.25 to 1.50	9.4853
From 1.50 to 2.00	10.2098
From 2.00 to 3.00	11.3999
From 3.00 to 4.00	13.0872
Greater than 4.00	13.7378

<b>HOUSING CHARACTERISTICS AND APPLIANCES FACTOR</b>	
<b>Wall Material</b>	
Without walls, bamboo or other organic materials	0
Zinc, cloth, cardboard, cans	0.2473
Raw wood	2.0207
"Bahareque"	4.8586
Adobe, "Tapia pisada"	6.2845
Block, bricks, stone, prefabricated material, polished wood	7.7321
<b>Predominating roof material</b>	
Straw or palm leaves	0
"Desechos" (cardboard, cans, "sacos", etc.)	2.1043
Zinc, asbestos, cement, without "cielo raso"	3.7779
Clay tile, zinc, asbestos, cement, with "cielo raso"	5.0973
<b>Predominating floor materials</b>	
Dirt	0
Burda wood, boards	2.9037
Cement	3.6967
Floor tile (clay, vinilo), brick or paving tile	5.8712
Wall to wall carpet, marble, polished wood	6.8915
<b>Home appliances</b>	
No basics	0
Up to three basics	2.1435
Four basics without washer	3.0763
Up to three or four basics with laundry machine	4.7194
<b>PUBLIC UTILITY SERVICES FACTOR</b>	
<b>Water supply system</b>	
River or spring	0
Public fountain	1.1601
Well without water pump, container or rain water	2.6497
Well with water pump	4.6037
Container truck	6.1693
Aqueduct	7.2554
<b>Sewage disposal</b>	
No sewage	0
Latrine	2.4519
Toilet without connection to sewer or septic tank	3.3323
Toilet with connection to septic tank	3.9615
Toilet with connection to sewer	6.8306
<b>Garbage disposal</b>	
Yard, lot, river, etc.	0
Local container or public trashcan	2.1291
Picked up by public services	3.2701

SOURCE: Velez y Moreno, 1994

(\*) Summary of weights and scales after typification. Sum of maximum scores per variable is equal to 100.

Table 1.b Score per answer category in each variable. SISBEN, Rural, 1993

VARIABLES CATEGORY DESCRIPTION	SCORE(*)
<b>HOUSING CHARACTERISTICS AND APPLIANCES FACTOR</b>	
<b>Wall Material</b>	
Bamboo, canes or other organic materials	0
Zinc, cloth, cardboard, cans	3.2042
"Bahareque", mud wall cast in woodforms, adobe	3.4319
Raw wood	4.5588
Block, brick, stone, prefabricated material, polished wood	7.0780
<b>Predominating roof material</b>	
Straw or palm leaves	0
"Desechos" (cardboard, cans, "sacos", etc.)	1.1312
Zinc, asbestos, cement, without plastered ceiling	3.7615
Clay tile, "losa", zinc, asbestos, cement, with plastered ceiling	4.8771
<b>Predominating floor materials</b>	
Dirt	0
Raw wood, boards	2.4628
Cement	3.7474
Floor tile (clay, vinilo), brick, paving tile	5.4726
Wall to wall carpet, marble, polished wood	5.7495
<b>Home appliances</b>	
No basics	0
Up to three basics	1.6865
Up to three or four basics with laundry machine	2.7080
<b>HUMAN CAPITAL, SOCIAL SECURITY AND FIRM SIZE FACTOR</b>	
<b>Schooling of oldest wage earner</b>	
No education	0
Some elementary	0.2528
Complete elementary	1.5723
Some secondary	3.7663
Complete secondary	5.1876
Some higher education	11.4263
Complete higher education	13.0990
Graduate studies	18.3860
<b>Mean schooling for people 12 years old and older</b>	
0	0
4	0.3269
5	1.5793
10	3.2931
11	6.3560
15	12.3633
16	12.5863
More than 16	18.0414
<b>Social Security and plant size</b>	
Without Social Security and works alone or works in a plant with 2 to 9 employees	0
Without Social Security and works in a plant with 10 or more employees	1.4320
With Social Security and works alone or doesn't work	2.6097
With Social Security and works in a plant with 2 to 9 employees	3.6514
With Social Security and works in a plant with 10 or more employees	4.5259

<b>PUBLIC UTILITY SERVICES FACTOR</b>	
<b>Water supply system</b>	
River or spring	0
Public fountain	1.0523
Well without water pump, container or rain water	1.6591
Well with water pump	2.2640
Container truck	3.5759
Aqueduct	7.2438
<b>Sewage disposal</b>	
No sewage	0
Latrine	1.4398
Toilet without conection to sewer or septic tank	3.0718
Toilet with conection to septic tank	4.2110
Toilet with conection to sewer	7.3137
<b>Garbage disposal</b>	
Yard, lot, river, etc.	0
Local container or public trashcan	1.5414
Picked up by public services	2.6758
<b>DEMOGRAPHIC, UNEMPLOYMENT, INCOME AND ROOM CROWDING</b>	
<b>Proportion of children below 6 years old</b>	
Greater than 0.65	0
From 0 to .65	0.2181
Cero	1.1626
<b>Proportion of family members employed</b>	
Less than 0.30	0
From 0.30 to 0.60	1.0806
From 0.60 to 0.90	1.8668
Greater than 0.90	3.1957
<b>Room Crowding (Number of rooms per person)</b>	
Less than 0.30	0
0.30 to 0.60	0.8956
0.60 to 1.00	1.8988
1.00 to 4.00	2.9379
Greater than 4.00	4.9313
<b>Per capita income in minimum wage units</b>	
Less than 0.15	0
From 0.15 to 0.25	1.1079
From 0.25 to 0.35	1.9561
From 0.35 to 0.50	2.9685
From 0.50 to 0.75	3.9781
From 0.75 to 1.00	4.9210
From 1.00 to 1.25	5.6862
From 1.25 to 2.00	5.6862
From 2.00 to 3.00	7.7840
From 3.00 to 4.00	8.5781
Greater than 4.00	9.3504

SOURCE: Velez y Moreno, 1994

(\*) Summary of weights and scales after typification. Sum of maximum scores per variable is equal to 100.

<[Table 1A. SISBEN. Score per answer category. Urban. ]>

<[Table 1B. SISBEN. Score per answer category. Rural. ]>

The variables collected by the *Ficha de Caracterización Socioeconómica* to tabulate the SISBEN score that predicts the well-being of the household, can be roughly grouped into four sets: Housing and home appliances, public utility services, human capital endowments and economic risk and, finally, a fourth group that includes family demographics, unemployment, dependency ratio and income per capita.<sup>6</sup> Within each group, there are three or four variables and each of them is recorded in different categories or characteristics, easily observed or confirmed by the interviewer. Relative values are assigned to each of the (mutually exclusive) observable characteristics within each sub-group -these values are distinct for rural and urban areas-. For instance, in the urban area there is a gain of five points associated with economic risk and employment type, when the principal household earner works in a firm with social security and more than ten workers instead of a smaller firm and no social security.<sup>7</sup> Having a principal wage earner with higher education instead of incomplete primary education brings almost 12 additional points. If more than 90% of eligible members of the household are employed, three additional points are given than if less than 30% of economically active household members are employed. Procuring water from aqueduct or pipe produces a gain of almost six points relative to getting it from a public fountain. Having at least three basic home appliances *and* a laundry washing machine increases in 2.5 points the total score for the household. With this level of detail, it is then possible to calculate the household's final SISBEN score.

In some programs the final SISBEN index or score is divided into strata based on the severity of poverty – level 1 being the poorest and roughly correspondent to other definitions in the region of extreme poverty, while level 2 corresponds to more traditional definitions of poverty.<sup>8</sup> There are six levels of SISBEN classification, with absolute scores ranging from 0 to 100. The majority of national

---

<sup>6</sup> The reasons behind this grouping will be explained in section 3 and the appendix.

<sup>7</sup> In a scale of zero to 100 points from minimum to maximum.

targeted social programs restrict eligibility to poverty levels 1 and 2, although there are a few exceptions which also include level three –see Table 3-. The cut-off points for the first three strata were determined with a probit model with the objective of minimizing the error (type I ) of excluding the “true” poor when defined in terms of either income measures or basic needs indicators.<sup>9</sup> The cut-off used for determining eligibility for entry into the subsidized public health care program – the most important user of the SISBEN system corresponds to level 2- with a cut-off score of 47 in urban areas and 30 in rural areas. The cut-off is different for each program which uses the SISBEN results and the system can also serve to allocate program resources to the neediest with the cut-off determined by program budget constraints, given that each household is given an absolute score.

*<Table 2. SISBEN Strata 1 to 3. Cut-off points and comparison with distribution of the poor by income criterion. Urban and Rural, Colombia, 1993.>*

When poverty measures by income and SISBEN are compared (Table 2) , *relative* rural poverty seems to be more intense under the second criterion. Using the *income* extreme poverty line, rural poverty is nearly 33%, which is 3.3 times the urban poverty rate . However, using the SISBEN stratification, extreme rural poverty is nearly 23%, which is more than 4 times the extreme urban poverty rate.<sup>10</sup>

### ***Coverage and Rates of Implementation***

SISBEN was implemented largely during the course of 1994 and 1995. The targeting system was applied in more than 90% of the municipalities in Colombia thanks to a wide-ranging campaign of dissemination, technical assistance, and training which was designed and coordinated at the central level by Mision Social from the National Planning Department and implemented in partnership with the administrative departments of Planning at the Departmental level.

*<Table 3. Total SISBEN Surveyed Population, Urban and Rural by Poverty Strata >*

---

<sup>8</sup> See Velez, Castano and Correal (1995).

<sup>9</sup> See Castano (1995).

<sup>10</sup> When income poverty line is used results are similar, although relative differences are smaller.

Table 3 provides more detail on SISBEN's coverage as of 1996. Nearly 40% of the total population of Colombia – in more than 90% of the municipalities and from all the departments, was covered by the SISBEN survey as of 1996.<sup>11</sup> Sixty percent of the households surveyed were in urban areas and 40% in rural areas.<sup>12</sup> Due to the concentration of population in urban areas, the result was that slightly less than 30% of the population in urban areas was covered by SISBEN and slightly under 40% in rural areas. In urban areas, nearly two-thirds of the population surveyed demonstrated levels 1 and 2 of poverty according to the SISBEN index, allowing them potential eligibility for entry into the subsidized public health system. In rural areas, the corresponding proportion of SISBEN respondents classified in the bottom two tiers of poverty was 78%. In rural areas, the implementation of SISBEN made special efforts to reach more marginalized zones, with three quarters of households surveyed residing outside of town centers. Survey coverage was determined at the local level based on analysis of prior socio-economic information to determine localities where low-income groups were concentrated. A first round large-scale census was undertaken in these areas, accounting for roughly 95% of SISBEN coverage. Subsequently, there was a demand-driven phase of surveying, wherein NGOs and community groups solicited surveying of specific households and areas where there was latent demand for social programs. Furthermore, individual households could also apply for SISBEN surveying if they felt that they were eligible to receive subsidies and had been passed over in earlier survey rounds. Nonetheless, nearly one-third of municipalities studied in a recent evaluation reported problems with undercoverage of the target group – the majority citing difficulties in achieving desired response rates in the first round of surveying as the principal problem.

---

<sup>11</sup> As of June 1997, 98% of Colombia's municipalities had participated in SISBEN

<sup>12</sup> These proportions are not very far from the distribution of the poor by income criterion: 44% urban and 56% rural, according to World Bank (1995), ch.1.

**Table 2. Sisben Strata: cut-off points and population distribution**

Strata Sisben	Maximum Points	% of population by strata		Poor by income 1992(*) %
		predicted	cumulative	
<b>URBAN</b>				
1 (Extreme Poverty)	35.8	4.9	4.9	9.9
2 (Poverty)	47.4	15.5	20.5	37
3 (3 times extreme poverty line)	58.5	31.4	46.9	
<b>RURAL</b>				
1 (Extreme Poverty)	18.0	23.0	23.0	32.7
2 (Poverty)	29.9	32.7	55.7	65
3 (3 times extreme poverty line)	45.4	34.2	66.9	

Source: CASEN Survey and calculations by Castaño (1995)

(\*): Source, DANE, World Bank (1995).

**Table 3. Total SISBEN Population. Urban and Rural by Poverty Strata**

	Total	Population SISBEN/96	% Covered	Affiliation to Health Subsidized Insurance			% of surveyed population by Strata			
	Population			Potential = Strata 1 + 2	Actual 1997	Strata 1	Strata 2	Strata 3	Rest	
National Total	38,318,448	12,208,122	32%	8,354,415	68%	7,026,689	29.2%	39.2%	24.8%	6.7%
Urban Total	27,751,869	7,735,114	28%	4,851,053	63%		24.6%	38.2%	27.9%	9.4%
Rural Total	11,758,788	4,473,008	38%	3,503,362	78%		37.2%	41.1%	19.6%	2.1%
Rural Populated							24.6%	29.6%	14.4%	1.2%
Rural Disperse							12.7%	11.4%	5.2%	0.8%

Source: "Informe de las Oficinas de Planeación Departamental sobre estado de avance SISBEN

DANE: Población estimada 1996, con base en Censo de Población 1993

**Table 4. Average cost of SISBEN Screening. Colombian Municipalities, 1996.**

	Cost	Acumulated frecuency
1st QUARTILE	1752.00	25%
2nd QUARTILE (median)	2395.00	50%
3rd QUARTILE	3657.00	75%
Average	3295.31	71%

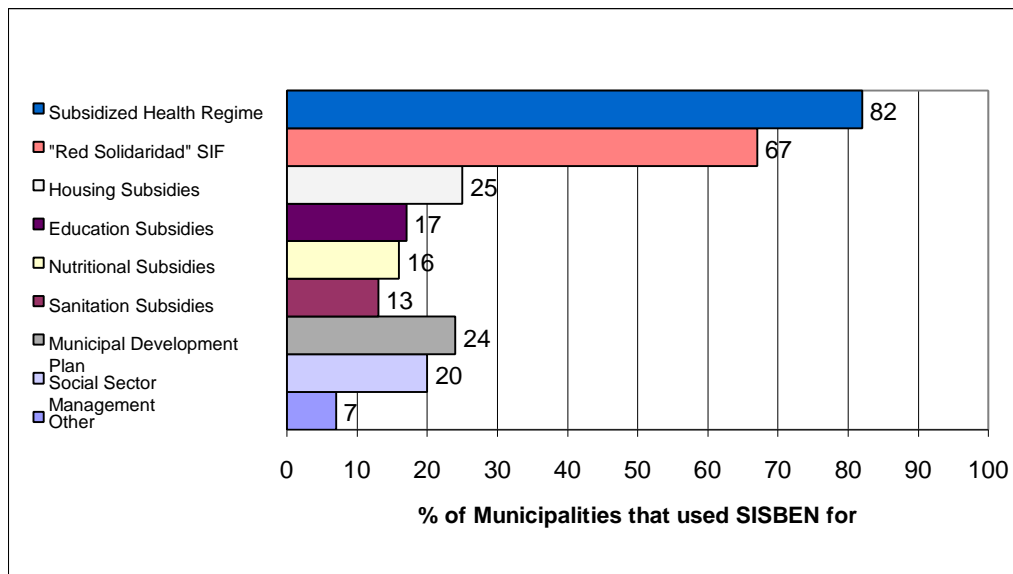
Source: Misión Social, Departamento Nacional de Planeación

### *Utilization*

While SISBEN was most widely used for the expansion of the subsidized public health program (Graph 1) – with more than 80% of participating municipalities using the survey for this purpose-, it was also used to aid in the allocation of program resources from other sources. 62% of municipalities used SISBEN results as an aid in the allocation of funds for Colombia's equivalent of a Social Investment Fund – la Red de Solidaridad. Trailing behind in use of SISBEN information were housing subsidies (25% of municipalities), municipal development plans (24%); and education and nutrition subsidies (17 and 16% respectively). Nonetheless, the main user of SISBEN was without doubt the Ministry of Health, for whom SISBEN provided a key informational component in the design and implementation of a program of decentralized health care subsidies in partnership with the municipalities.

*<Graph 1. SISBEN utilization by program type. Frequency by municipality. Colombia, 1996>*

**Graph 1. Sisben Utilization by program type. Frequency by municipality. Colombia 1996**



Source: DNP-UNDP(1996)

Of the more than 90% *surveyed* population who were not already enrolled in the public health care system, 29% were below SISBEN's extreme poverty level and 39% at poverty level. In other words, more than two thirds of the SISBEN survey respondents, that is 8.3 million persons gained potential access to subsidized public health care. As a result of expanding coverage by SISBEN municipalities gained actual funding to include their population beneficiaries in the newly designed public health care system and the number of people covered at the end of 1997 was nearly 7 million in total. This represents almost 85% of the total of potential beneficiaries, and clearly demonstrates the success of the expansion of the new health care scheme once it was linked to an expedient and relatively inexpensive method of targeting.

#### ***Screening Cost are Low***

Screening cost, for SISBEN are very moderate. Table 4 presents information on the average cost of screening each family by SISBEN system in 453 Colombian municipalities that represent roughly 90% of the total number of applied *Fichas* evaluated by Mision Social in 1996.<sup>13</sup> The median cost of completion of the SISBEN questionnaire in 1996 was 2,395 pesos –roughly \$2.25 in 1996-. Cost were relatively heterogeneous but half of the sample fell between 73% and 152% the reported median cost per family screened. Unit costs per person were quite small, averaging 570 pesos or \$0.53 per person. The cost for screening is a very small portion –half percentage point- of the annual cost of a Subsidized Health Plan for a family of five. And much less for a Housing Subsidy of INURBE that has a value two or three times larger. Furthermore, utilization of SISBEN for targeting more than one program at the same time would bring economies of scale and reduce even further the relative cost of screening. Compared to other screening devices in Latin America, SISBEN thus appears inexpensive. According to Margaret Grosh (1994) similar targeting systems have targeting cost ranging from 0.5% to 1.5% of program spending.<sup>14</sup>

*<Table 4. Average Cost of SISBEN Screening. Colombian Municipalities, 1996>*

---

<sup>13</sup> DNP-UNDP (1996).

### *Factors Contributing Towards Success in Implementation*

There were several political and institutional factors which contributed to SISBEN's rapid implementation in Colombia. First of all, the Colombian constitution of 1991 explicitly states the need to target or direct social spending to the neediest groups in the country – at the national, departmental, and local level. In fact, more than 60% of local government transfers are distributed in direct proportion to the number of residents with Unsatisfied Basic Needs, as stipulated in article 357 of the Political Constitution. Secondly, Article 30 of National Law 60 of 1993 on Intergovernmental Transfers and Resources defines targeting of public subsidies as “the process by which it is guaranteed that social spending is assigned to the neediest groups.”

Towards that end, CONPES –the Council of Economic and Social Policy- has the mandate to define, every three years, the criteria for the determination, identification, and selection of beneficiaries to be used in the allocation of social spending by different levels of government in Colombia. CONPES stipulated two criteria in 1994 – the system of socio-economic stratification, and SISBEN. The former is used essentially for micro-geographic targeted subsidies through public utility tariff discrimination, while the latter is used for beneficiary targeting.

The second principal stimulus –certainly the most powerful- to the expansion of SISBEN was the promulgation of a major decentralization of the national health system, which provided funding for subsidized health care to municipalities on the basis of the number of eligible households -as defined by their SISBEN score- in each municipality.<sup>14</sup> Thus, in addition to the legal requirement for the system, there was a program need for the SISBEN results, with significant resources attached as a prize for implementation. Moreover, this incentive compatible characteristic was reinforced by the growth of other national programs that required individual welfare measures for targeting: for example, the Social Fund -Red de Solidaridad Social- and the INURBE's Housing Subsidy Program.

---

<sup>14</sup> See Chapter 4 in Grosh (1994). Administrative cost should also be compared, however current data available for Colombia does not allow this calculation.

<sup>15</sup> Municipal authorities are democratically elected in Colombia.

Careful user-friendly design and massive training at the local further stimulated the wide-ranging implementation of the SISBEN system. These free resources not only reduced fixed monetary costs, but also time waiting costs for implementation of the targeting system and delivery of programs. Partnerships and effective division of labor between national, departmental, district, and municipal governments allowed for the necessary technical support to accompany implementation. Excellent social marketing, education, and training – not only of government officials but of NGOs representing constituencies who would benefit from the SISBEN system, created additional awareness of the usefulness of the information system and demand for its implementation. The user-friendly design, complete with a series of training pamphlets and a customized software to process SISBEN scores and use SISBEN information for the monitoring of social spending at the municipal level, increased the relevance and utility of the system for local governments – further speeding its implementation.

In summary, most of the fixed costs for targeting implementation were absorbed by the central government and marginal costs –monetary and non-monetary- for municipalities were relatively low. Decisions to implement SISBEN were decentralized and, in addition, incentives to do so were compatible as long as programs promulgated together with SISBEN gave a guaranteed political return for local leadership.

## **2. An economic interpretation of SISBEN:**

In this section we want to show how the SISBEN Index can be interpreted as a composite index of welfare indicators to rank households according to their level of well-being. We will see that SISBEN is a proxy means test index that belongs to the *welfarist* class of welfare measures. Before dealing with the main topic of this section, we start with a brief review of the advantages and limitations of alternative welfare indicators.

### **2.1 Alternative Individual Welfare Measures**

Typically individual welfare measures are instrumental in helping to target public programs in order to increase their capacity to reduce poverty and inequality. There are several arguments, both analytical and empirical, which show that when more and better information about

individual's welfare levels is used for targeting purposes the equity gains out of a fixed government budget can be significantly enlarged.<sup>16</sup>

In the choice of individual welfare measures we may identify three basic approaches:<sup>17</sup> The *Welfarist* approach, based on individual utility derived from commodities or income. The *Basic Needs* approach, which stresses the satisfaction of human needs or the level of well-being achieved in specific terms, such as: housing, basic, housing services, room crowding, economic dependence, access to schooling.<sup>18</sup> Thirdly, there is Amartya Sen's approach that puts emphasis on the *capabilities* to function and therefore looks at commodities not as ends but as means to desired activities.<sup>19</sup> All alternative measures have certain advantages and disadvantages. Moreover, the implementation of each type of measure faces a whole spectrum of specific complications that constrains the choice of a feasible welfare measure.

The *Capabilities* approach certainly faces an unobservability problem. At most, achievements are observed but capabilities are not.<sup>20</sup> *Basic Needs* is functionally inflexible in terms of weights and is lexicographic because one unsatisfied Basic Need is enough to label a household as poor. In this sense the Index of Unsatisfied Basic Needs –NBI- gives a confusing signal of extreme poverty, since it ignores the level of satisfaction in other dimensions of well-being. According to the World Bank (1995), in the case of Colombia, NBI is a very poor indicator of *income* poverty incidence

*Although the poor (defined on the basis of three alternative poverty lines) experience higher incidence of the four NBI's than the population as a whole (...) the incidence of unsatisfied basic needs (...) is far from overwhelming. In urban areas, for example, only one on three individuals in the extremely poor group is poor under NBI criteria. To the extent that policymakers wish to reach families with incomes below the poverty line, the use of the NBI index would involve large errors of exclusion and inclusion<sup>21</sup>.*

<sup>16</sup> For example, Ravallion and Chao (1989), Glewwe and Kanaan (1989).

<sup>17</sup> We follow the classifications as suggested in Lipton and Ravallion (1995), pp. 2566-69 and Deaton (1997) ch. 3.

<sup>18</sup> See Streeten *et al* (1981) and Richards *et al* (1982)

<sup>19</sup> Sen (1979, 1985, 1987).

<sup>20</sup> Lipton and Ravallion (1995) p..

<sup>21</sup> World Bank (1995), p.8.

Income is ideal as long as it avoids a parametric utility function, but reliable data is difficult to obtain. Consumption is good but requires expensive data collection and the availability of price indexes. Additionally, these two welfarist measures ignore the role of public spending,<sup>22</sup> and therefore they may overestimate poverty and inequality, when public spending is progressive.<sup>23</sup>

Consumption measures are also criticized because they ignore life cycle effects and different constraints to smooth consumption –especially credit for the poor-.<sup>24</sup> However, if we think that the poor are more likely to be constrained in their access to credit, their accumulation of durable goods should follow closely their savings path. Therefore their differences in durable consumption with respect to the rest of unconstrained consumers should be *larger* than their income differences would suggest, especially for poor households in the earlier stages of the life cycle.

The welfarist approach offers the foundation for the exact Means Test that is typically based on income or consumption. However, these indicators are not practical in developing countries.<sup>25</sup> In addition to the incentives to understate income or consumption welfare levels, verification is costly given the lack of records about social security and taxes or data bases on government payrolls. A preferable alternative is the use of *proxy means tests*, which are weighted indexes on non-income characteristics of the household, including economies of scale or equivalence scales in family composition.<sup>26</sup>

These type of indexes have been used in Latin-American countries –e.g., Chile, Costa Rica and Colombia- and seem to be good predictors of individual welfare. Grosh and Baker

---

<sup>22</sup> Deaton (1997) or Lipton and Ravallion (1995)

<sup>23</sup> For example, Vélez (1995, 1997) shows that Colombian public spending in 1992 reduced the Gini coefficient -income plus imputed subsidies- in almost 9%. And estimates on poverty rate reduction are close to 4 percentage points.

<sup>24</sup> Lipton and Ravallion (1995) p. 2573.

<sup>25</sup> Grosh and Baker (1995).

<sup>26</sup> Note that the Proxy means test remains within the commodities oriented or welfarist approach.

(1995) tested the predictive capabilities of a Proxy Mean Test in Jamaica as a welfare indicator. In this case the index contained information about location, housing, human capital endowments of the family, employment and ownership of durables. Their evaluation shows that these indexes are reasonably good to assess eligibility for social programs. Although they do not eliminate the undercoverage problem completely they reduce leaks and as a result their capacity to reduce poverty is significantly increased. Grosh and Baker (1995) show as well how robust these indexes are to distortions of information by the potential beneficiaries of the programs. According to Grosh (1994), within this class of indexes the Chile's proxy means test has produced the best incidence outcomes to date in Latin America.

## **2.2 A composite index of welfare indicators to rank households according to their level of well-being.**

Given the description offered above, SISBEN is obviously a proxy means test index. It is a weighted index of several characteristics of the household that was constructed with an *ad-hoc* statistical algorithm, namely Optimal Scaling and Alternating Least Squares. In this section we propose an interpretation of the economic foundations implicit in the procedure used to derive the SISBEN index of household's well-being, from the perspective of consumer behavior.

First of all, we notice that this index includes three different sets of variables that are relevant to evaluate the consumer's utility level: (i) consumption of durable goods and services such as housing, public utilities and household appliances (ii) Human capital endowment, and (iii) current income.

### ***Marshallian demands***

From consumer theory we know that the consumer's preference ordering on good bundles can be represented by a utility function, from the goods space to the real line. This cardinal representations of the preference ordering is such that vectors of consumer demand determine

different individual welfare levels: that is,  $U(x_1) > U(x_2)$  if and only if the goods vector  $x_1$  is strictly preferred to  $x_2$ .<sup>27</sup>

In the case of SISBEN, the first set of variables observed for each household  $i$  is the vector of Marshallian demands,  $x^*(p_o, y_i)$ .<sup>28</sup> Given that households *differ in their income levels*, the direct utility derived from the observed Marshallian demands must reflect their well-being differences, that is, their level of utility increases with income:  $U(x^*(p_o, y_i)) \geq U(x^*(p_o, y_j))$  if  $y_i > y_j$ . However, other sources of variation in welfare should come from taste shifters and differences in the price vectors that each household faces -mainly from regional changes-. Thus, to obtain the order of well-being of the households, we are entitled to postulate a utility index  $U_i$  as a function of the vector of observed demands for goods  $x_i$  of each households  $i=1, \dots, n$ ,

$$U_i = F(x_i) + \mathbf{e}_i + \mathbf{n}_i$$

where  $\mathbf{e}_i$  and  $\mathbf{n}_i$  are the error terms that correspond respectively to taste shifters and regional differences in prices vectors. Notice that  $U_i$  can be seen as a quantity metric utility index of an indifference curve, thus it can be represented by *the scale of a vector in the goods space*.<sup>29</sup>

### ***Current Income***

Directly from the definition of the indirect utility function:  $V(p, y) = U(x^*(p, y))$ , it is obvious that  $V(\cdot)$  must reflect the same differences in income distribution that  $U(\cdot)$  captures in terms of optimal consumption. The Indirect Utility Function is nondecreasing in income, therefore  $V(p_o, y_i) \geq V(p_o, y_j)$  if  $y_i > y_j$ .

Even more, we know that (i) the Expenditure Function of the household,  $e(p, U) = p' x_h(p, U)$ , represents the same ordering given by the utility level  $U$  and (ii) It is identically true that  $y = e(p_o, V(p_o, y))$ . Thus income as such represents that same ordering given by the indirect

<sup>27</sup> If there is a preference ordering on the goods space  $X \hat{=} \mathbb{R}n^+$ , and this ordering is Rational, Continuous and strictly Monotonic then there exist a utility function from the goods space to the real line. See Varian (1992) ch. 7 or MasCollé (1997), chs. 1 and 3.

<sup>28</sup> These quantities are the outcome of the optimization of their preferences –i.e.: maximize  $U(x)$ - constrained by their income,  $y$  and the price vector ruling the goods market,  $p_o$ .

utility function  $V(\cdot)$  and the direct utility function  $U(\cdot)$ .<sup>30</sup> Therefore, the household differences in well-being can be represented by a utility index  $VI_i$  as a function of *current income*,  $yc_i$  of different households, subject to additional sources of variation coming from (i) temporary components of income,  $\mathbf{r}_i$  and (ii) the taste shifters and regional price differences, mentioned above.

$$VI_i = G(yc_i) + \mathbf{e}_i + \mathbf{n}_i + \mathbf{r}_i$$

### ***Human Capital Endowments***

Labour income of the household is the realization in the labour market of services derived from their human capital endowment. This process occurs as a joint decision of labour force participation (leisure versus work) and consumption of goods and services. The household optimization process of a typical household is represented by the maximization of a utility function of consumption and leisure subject to income from exogenous sources and from the sale of labour services. Obviously, Marshallian demand functions of consumption  $c^*$  and leisure  $l^*$  (in  $[0, 1]$ ) will be functions of the price vector -including the wage or the price of leisure- and the *full income* that is linear on the human capital endowment of the consumer  $h$  and exogenous income  $Y_o$ :

$$l_i^*(p, wh, + Y_{o_i}) \quad \text{and} \quad c_i^*(p, wh, + Y_{o_i})$$

Where,  $h_i$  is the number of equivalent human capital units of agent  $i$  that have price  $w$  in the labour market. And the corresponding indirect utility function is

$$V_i((p, h_i, Y_{o_i}) \circ U_i(l_i^*(p, h_i, Y_{o_i}), c_i^*(p, h_i, Y_{o_i}))$$

Which implies that the ranking of household welfare can be represented by a utility index  $V2_i$  function of observed household's human capital endowments  $h_i$ .<sup>31</sup> And this index also reflects

<sup>29</sup> Deaton and Muellbauer (1980), p.181.

<sup>30</sup> Deaton (1997), p.148 emphasizes the advantage of using real income or expenditure as welfare indicator "in order to avoid the specification of a parametric utility function".

<sup>31</sup> In a Developing Country like Colombia, this process of labour force participation is obviously more complex and it involves the selection of employment as wage earner or self employed and differences in

the order of preferences represented by  $U_i(\cdot)$  and  $VI_i(\cdot)$ . Obviously, there are additional sources of variations,  $\mathbf{d}$  coming from the (i) regional differences in the price of labour services and (ii) exogenous shocks to demand for labour skills.

$$V2_i = H(h_i) + \mathbf{e}_i + \mathbf{n}_i + \mathbf{d}$$

So far we have shown that for each of the three sets of variables included in SISBEN index we are able to postulate a utility index for the households of the sample. Then a positive linear combination of those three representations of the order of households',  $W_i$ , welfare must reflect that same order,

$$W_i = a U_i + b VI_i + c V2_i = a F(x_i) + b G(y_i) + c H(h_i) + \mathbf{m}$$

where  $a$ ,  $b$  and  $c$  are positive constants. Thus the index  $W_i(x_i, y_i, h_i)$  on the real line of durable goods, current income and human capital endowments should discriminate the households on the sample in the direction of economic well-being.

***Relevant data on consumption, but in categorical form and without a metric***

In the SISBEN index, the data on demands for goods and services is limited: first of all, it does not contain the whole set of goods but a subset of durable goods: housing and home appliances. This choice serves the practical purpose of using information that is not difficult to verify. Secondly, the information available provides a very rich description of the quality or characteristics of this type of goods. In fact, it includes the relevant characteristics of dwellings describing several dimensions of its construction and also the access and quality of public utility services, in various *categories* or quality levels. For example, the house floor is described in categories such as dirt, unpolished wood, cement, bricks, “vinilo or baldosas” and carpet or polished wood or marble. Or the supply of water categories are: obtained by walking to a river or

---

plant sizes. Those choices are bundled to different levels of exposure to risk, e.g.: social security coverage for health and / or retirement. Then, the type of job selected by workers must reflect the level of coverage that workers have against exogenous sources of risk.

spring, collected from a public fountain, obtained from a well by hand or storing rain, or from a well with pump, or from a water tank delivery or, at best, via aqueduct.

Nevertheless, the availability of relevant information about consumption in *categorical form –polycothomus-* is a mixed blessing because it brings additional complications or demands to the algorithm used to derive the utility index: the statistical algorithm used must provide *a metric or a distance* for the quality categories included in the sample. Even more, for those variables for which we have a metric –e.g. current income, years of schooling- it is not desirable to impose *ex-ante* a *linear* relation to the utility index, therefore more flexibility should be allowed to this functional relation.<sup>32</sup>

### **3. The construction of the utility index by the statistical algorithm of Optimal Scaling: A solution to the weight problem and the use of valuable categorical information on consumption of durables**

The construction of the utility index proposed in the previous section cannot be done with regression analysis methods. The utility levels represented by  $U$ ,  $V1$  and  $V2$  are not observable. Second, most of the variables of the consumption of housing and other durables are observed as categories without a metric, thus linear principal components do not apply. However, from consumer theory as summarized in the preceding subsection, we know that there exist an index on the real line that represents the mapping to the utility of households from Marshallian demands, or from current income or from human capital endowments.

We claim that the algorithm of Alternating Least Squares and Optimal Scaling –ALS-OS- (Qualitative Principal Components) finds that index by (i) providing a metric to the different categories of the variables and (ii) giving them weights according to their contribution to the *first* principal component of the system. Precisely that first component represents the dimension of maximal variability in the data, which is the unobservable utility index that must be constructed. Note that this claim rest upon the assumption that the main source of variability in the sample

---

<sup>32</sup> For example, by entering several range values of the variable with metric and allowing different slopes, spline forms of the index might result.

data is related to differences in income and human wealth and not to taste shifters or regional price changes.

The use of simultaneously or endogenously determined variables like current income, consumption and human capital could be considered a problem of “redundancy” and a source of multicollinearity in a regression exercise. However in this case it is convenient and advantageous given that the implicit consumer’s optimization process links the different subsets of data and increases the constraints on the algorithm to *converge* to the true *utility index*. We know *a priori* that there is a covariance structure between these three main sets of variables –consumption, income and human capital endowments- and that is precisely the force that drives the weight assignment of weights in the principal component part of the process.

The algorithm used involves several steps, the first and most important procedure is called Optimal Scaling, a procedure that produces non-linear transformations of qualitative variables in order to maximize the fit of the data to a model of principal components. Notice that this specific characteristic of the transformation procedure imposes the objective of producing an ordering of households in the context of a principal component model. More specifically the ALS-OS algorithm was programmed to privilege the component with highest variance in the system, that is the first principal component. This is consistent to the objective of building the postulated welfare index as long as consumer theory anticipates that the maximum variance among households, subject to differences in the distribution of income, will be in the direction of utility or well-being.

In the next subsection we describe the technical steps involved in the algorithm of Qualitative Principal Components and then, we finish this section with a discussion of the main results of SISBEN for the urban and rural areas.

### **3.1 Qualitative Principal Components: PRINQUAL and ALS-OS<sup>33</sup>**

---

<sup>33</sup> In this subsection we describe the Algorithm of construction along the lines presented in Castano and Moreno (1994).

The PRINQUAL procedure uses an algorithm based on Alternating Least Squares (ALS) and Optimal Scaling (OS) to obtain non linear transformations of a set of qualitative variables in order to maximize the fit of the data to a *model* of linear principal components. The principle of OS considers each observation as a category and represents each category by a parameter *of the data*.

The ALSOS principle divides the set of parameters into two mutually exclusive and exhaustive subsets: the first one consists of the parameters of the model *-weights-* and the second contains the parameters of the data, which are called *optimal scaling parameters*. In turn, each subset can contain several mutually exclusive and exhaustive subsets. The optimization processes is performed by finding the least square estimations of the parameters in one subset, fixing the parameters in the others.

These estimations are called *conditional* least squares estimations, because its least square nature is conditional on the parameter's values of the others subsets. When the conditional least squares estimations are obtained, the old estimators are replaced by the new estimations. Then, the processes continues for the next subset and their conditional least square estimations are calculated. Alternating, the procedure continues one time on the model's parameters, another on the data's parameters, until convergence is obtained.<sup>34</sup>

### ***Principal Components***

The Principal Components are new variables which are generated as linear combinations (weighted sums) of the original variables. The weights are functions of the covariance structure of the variables, and have finite variance. Specifically, for each set of weights, its sum of squares is constrained to be one. The first principal component is the original variables weighted sum that has the largest variance. The second principal component is a weighted sum of the original variables that has the second largest variance and that is uncorrelated with the first one. The next component is unrelated to the former and its variance is less or equal to the variance of the

---

<sup>34</sup> See Exhibit A.1 in the Appendix below.

former components. The sum of the variances of the original variables -called total variation- is equal to the variance sum of the principal components.

The Principal Components model for media centered data is:

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

for  $i=1,2,\dots, n$   $y j=1,2,\dots,m$ , where  $n$  is number of observations and  $m$  is the number of variables,  $x_{ij}$  is the  $i$ -th individual value on  $j$ -th variable,  $w_{kj}$  is the weight of the  $j$ -th variable on the  $k$ -th component ( $k=1,2,\dots,c$ ), where the number of components is  $c$ .

The first principal component variable can be used in place of the full set of original variables. It provides a more parsimonious representation of the data, and explains the largest amount of total variation of the system.

There are  $n \times m$  equations like (1), one for each observation. A brief notational simplification can be done by use of matrices. In this case, we can represent (1) by

$$X = FW' + E$$

where  $X$  is a  $n \times m$  data centered matrix. We suppose that the variables (columns) have the same unit of measurement. In the Principal Component Analysis there are the same number of variables as principal components.  $W$  is the matrix of  $c \times m$  weights and  $E$  is the matrix of  $n \times m$  errors. When  $c$  is equal to  $m$  (there are as many components as variables), the matrix  $E$  is a null matrix.  $F$  is a  $n \times c$  matrix containing the values for the  $n$  subjects on the  $c$  principal components. If the variables are measured in different units, the analysis will be dominated by the variables with larger variance. In that case, the variables are standardized. Equivalently, the correlation structure of the  $X$  matrix can be used in place of the covariance structure

Note that principal components are optimal in the following sense: In a truncated model – e.g. with only the first principal component, as in the SISBEN case- the total variance of the component is maximal given the restrictions. That is, there is no other component that accumulates a higher proportion of the total variance of the original variables. In the case of SISBEN the value of  $c$  was restricted to one and the model had only one principal component.

The variability captured by the *first principal component represented 49% of the total variance, in both the rural and urban models.*<sup>35</sup>

### ***Qualitative Principal Components***

The PRINQUAL procedure (Qualitative Principal Components) extends the ordinary Principal Component Analysis to a more general context where we can use both qualitative and quantitative variables. These analysis uses non-linear transformations of the qualitative variables, preserving the variable measuring level. Ordinal variables are transformed in monotonic way. For nominal variables, the transformation the pertinence of the category of each observation. The procedure is based on Kruskal and Shepard (1974), Young, Takane and de Leeuw (1978) and Winsberg and Ramsey (1983). As we mentioned above, PRINQUAL uses nonlinear transformations of the qualitatives variables to maximize the fit of the data to the model of Linear Principal Components.

### ***The Maximum Total Variance –MTV- Method***

Based on Young, Takane y de Leeuw (1978), the Principal Component Model and attempts to maximize the sum of the first  $c$  eigen values of the covariance matrix. This method transforms variables to be (in a least squares sense) as similar to linear combinations of  $c$  principal component score variables as possible, where  $c$  can be much smaller than the number of variables. This maximizes the total variance of the first  $c$  principal components. In the SISBEN construction  $c$  was constrained to one, i.e. the first principal component. On each iteration, the MTV algorithm alternates classical Principal Component Analysis (Hotelling, 1933) with Optimal Scaling (Young, 1981).<sup>36</sup>

### ***Optimal Scaling in PRINQUAL***

PRINQUAL uses the OPSCAL PROC MATRIX function (SAS Institute, 1983) to do the optimal scaling. The OPSCAL function allows for optimally scaling nominal variables using Fisher's

---

<sup>35</sup> Castano and Moreno (1995), p.42, 74.

<sup>36</sup> Kuhfeld, Sarle y Young (1985) present a description of the method.

(1938) optimal scoring transformations method. Ordinal variables are rescaled using Kruskal and Shepard (1974) least squares monotonic transformation. Interval variables can be rescaled using splines transformations (de Boor, 1978) or using non optimal transformations like powers transformations.

### ***Stability and Convergence***

The algorithm for the SISBEN derivation was satisfactorily tested for convergence under different initializations of the variables.

After performing Optimal Scaling the set of variables remains *multidimensional*. Therefore *Oblique Factor Analysis* was performed in order to find if variables were associated in groups and if this groups could be ranked in terms of their contribution to total variance. Detailed explanation about this procedure is given in the Appendix below.

### **3.2 Results: adequate order of categories, changing importance of groups of variables in rural versus urban areas and relevant information about the poor.**

The signs, values and order that the algorithm assigns to the different categories of all variables are consistent with economic intuition on consumption and poverty. Tables 5A and 5B show the Optimal Scaling of the categories of variables included in the urban and rural SISBEN Index. In the urban area, the *average schooling of oldest wage earner* between complete elementary and some middle school and was given 3.6 points. However, those points could increase up to 7.7 if that same individual had graduate studies. *Average schooling of households members older than 12 years* was similar and the relative scale between maximal and minimal category had a comparable range. *Employment, social risk coverage and plant size for the oldest wage earner* was given values in a relative scale of 1 to 5.5. *Room Crowding* went from a minimum value of 1.7 when the number of rooms per person where less than 0.2 to almost 5 when the number of rooms per person was between 0.7 and 1.0. *Current Income* changes in a scale of 1 to 35 were given SISBEN point increments in a scale of 1 to 10, and furthermore the functional relation had

---

a spline concave form.<sup>37</sup> In contrast with this result the proportion of employed family members was in a smaller relative scale of 1 to 3.6 but the spline relation appeared to be convex. For all variables on housing characteristics and services, home appliances and employment category and insurance the *order of categories* reflects economic rationality and does not contradict the established opinion of sectorial experts.

<Table 5.A Optimal Scaling of SISBEN categories in each variable. Urban area. Colombia>

After convergence, the PRINQUAL process produced both the *optimal scaling* and the *weights* of the variables on the *first principal component*.<sup>38</sup> Finally the points used for the calculation of the SISBEN score are the result of multiplying the scaling of each variable's category by the corresponding weight of the variable. That result plus the typification of those values -to add up from 0 as a minimum to 100 as a maximum- are the scores initially presented in Tables 1.A and 1.B. As a group human capital and employment category counted for nearly 30% of total potential differences in SISBEN score points. Income per capita alone accounted for almost 14% of potential differences and together with the other three variables in its group accounted for almost 30%. Largest ranges are for income and education differences followed by bedroom crowding –8.4 points-, followed by a group of five variables with a similar range values: wall and floor type, water supply system, sewage disposal and employment type, social security and firm size for the oldest wage earner. Finally with smaller range values are –in descending

---

<sup>37</sup> Diminishing marginal impact of current income on the SISBEN welfare index. One could be tempted to interpret this result as diminishing marginal utility of income, however we should avoid it since this model's aim is only predictive.

<sup>38</sup> The result is identical to the one obtained by Factor Analysis after Optimal Scaling and then applying first Principal Component within each group, followed by Principal Component for all four groups.

Table 5A OPTIMAL SCALING OF SISBEN CATEGORIES IN EACH VARIABLE.

URBAN AREA			
VARIABLES CATEGORY DESCRIPTION	Optimal scaling	Mean	Times the lowest category scale
<b>HUMAN CAPITAL, SOCIAL SECURITY AND FIRM SIZE FACTOR</b>			
<b>Schooling of oldest wage earner</b>			
No education	1.187	3.57	1.00
Some elementary	2.0371		1.72
Complete elementary	2.9897		2.52
Some secondary	3.8066		3.21
Complete secondary	5.0313		4.24
Some higher education	6.3086		5.31
Complete higher education	7.2314		6.09
Graduate studies	7.7207		6.50
<b>Mean schooling for people 12 years old and older</b>			
0	1.26	3.62	1.00
4	2.1382		1.70
5	2.8472		2.26
10	3.8936		3.09
11	5.3086		4.21
15	6.2646		4.97
16	6.9258		5.50
More than 16	7.1641		5.69
<b>Social Security and plant size</b>			
Without Social Security and works alone or doesn't work	1.0869	2.93	1.00
Without Social Security and works in a plant with 2 to 9 employees	1.9058		1.75
Without Social Security and works in a plant with 10 or more employees	2.9512		2.72
With Social Security and works alone or doesn't work	3.8638		3.55
With Social Security and works in a plant with 2 to 9 employees	5.1904		4.78
With Social Security and works in a plant with 10 or more employees	5.9834		5.51
<b>DEMOGRAPHIC, UNEMPLOYMENT, INCOME AND ROOM CROWDING</b>			
<b>Proportion of children below 6 years old</b>			
Greater than 0.65	1.8042	2.51	1.00
From 0 to .65	1.9861		1.10
Cero	3.0044		1.67
<b>Proportion of family members employed</b>			
Less than 0.30	1.252	1.94	1.00
From 0.30 to 0.60	1.7908		1.43
From 0.60 to 0.90	2.347		1.87
Greater than 0.90	4.4727		3.57
<b>Room Crowding (Number of rooms per person)</b>			
Less than 0.20	1.7537	4.34	1.00
0.20 to 0.30	2.1675		1.24
0.30 to 0.40	2.979		1.70
0.40 to 0.70	3.6602		2.09
0.70 to 1.00	5.0059		2.85
1.00 to 4.00	6.2031		3.54
Greater than 4.00	7.9658		4.54
<b>Per capita income in minimum wage units</b>			
Less than 0.15	1.196	4.99	1.00
From 0.15 to 0.25	1.822		1.52
From 0.25 to 0.35	2.8081		2.35
From 0.35 to 0.50	3.8076		3.18
From 0.50 to 0.75	5.1572		4.31
From 0.75 to 1.00	6.4268		5.37
From 1.00 to 1.25	7.2881		6.09
From 1.25 to 1.50	8.2012		6.86
From 1.50 to 2.00	8.7363		7.30
From 2.00 to 3.00	9.6152		8.04
From 3.00 to 4.00	10.8613		9.08
Greater than 4.00	11.3418		9.48

<b>HOUSING FACTOR</b>			
<b>Wall Material</b>			
Without walls, bamboo or other organic materials	1.4536		1
Zinc, cloth, cardboard, cans	1.5986		1.10
Burda wood	2.6382		1.81
"Bahareque"	4.3018		2.96
"Tapia pisada" or "adobe"	5.1377	5.61	3.53
Block, bricks, stone, prefabricated material, polished wood	5.9863		4.12
<b>Predominating material for roofs</b>			
Wheat or palm	0.6656		1.00
"Desechos" (cardboard, cans, "sacos", etc.)	2.0254		3.04
Zinc, asbestos, cement, WITHOUT "cielo raso"	3.1069	3.55	4.67
"Teja", "losa", zinc, asbestos, cement, with "cielo raso"	3.9595		5.95
<b>Predominating floor materials</b>			
Ground	0.9191		1.00
Burda wood, boards	2.4727		2.69
Cement	2.897	3.17	3.15
"Baldosa, vinili, tablera o ladrillo"	4.0605		4.42
Wall to wall carpet or tapestry, alabaster, polished wood	4.6064		5.01
<b>Home appliances</b>			
No basics	0.6547		1
Up to three basics	2.1504	2.2447	3.28
Four basics without washer	2.8013		4.28
Up to three or four basics with washer	3.9478		6.03
<b>SERVICES FACTOR</b>			
<b>Water supply system</b>			
River or fountain	1.2078		1.00
Public supply	1.9734		1.63
Well without water pump, container or rain water	2.9565		2.45
well with water pump	4.2461		3.52
Container truck	5.2793	5.6	4.37
Water pipe	5.9961		4.96
<b>Sewage disposal</b>			
No sewage	1.157		1.00
Latrine	2.5537		2.21
Toilet without connection to sewer or septic tank	3.0552		2.64
Toilet with connection to septic tank	3.4136	4.45	2.95
Toilet with connection to sewer	5.0479		4.36
<b>Garbage disposal</b>			
Yard, lot, river, etc.	0.9752		1.00
Local container or public trashcan	2.2881	2.59	2.35
Picked up by public services	2.9917		3.07

SOURCE: Velez y Moreno, 1994

Table 5B. OPTIMAL SCALING OF SISBEN CATEGORIES IN EACH VARIABLE

RURAL AREA			
VARIABLES CATEGORY DESCRIPTION	Optimal scaling	Mean	Times the lowest category scale
<b>HOUSING CHARACTERISTICS AND APPLIANCES FACTOR</b>			
<b>Wall Material</b>			
Bamboo, canes or other organic materials	1.1165		1.00
Zinc, cloth, cardboard, cans	2.9033		2.60
"Bahareque", mud wall cast in woodforms, adobe	3.6587		3.28
Raw wood	3.0303	3.83	2.71
Block, brick, stone, prefabricated material, polished wood	5.0635		4.54
<b>Predominating material for roofs</b>			
Straw or palm leaves	0.9072		1.00
"Desechos" (cardboard, cans, "sacos", etc.)	1.5959	3.039	1.76
Zinc, asbestos, cement, without plastered ceiling	3.1973		3.52
Clay tile, "losa", zinc, asbestos, cement, with plastered ceiling	3.8765		4.27
<b>Predominating floor materials</b>			
Dirt	0.9491		1.00
Raw wood, boards	2.2944	2.248	2.42
Cement	2.9961		3.16
Floor tile (clay, vinilo), brick, paving tile	3.9385		4.15
Wall to wall carpet, marble, polished wood	4.0898		4.31
<b>Home appliances</b>			
No basics	0.9634	1.652	1.00
Up to three basics	2.0688		2.15
Up to three or four basics with laundry machine	2.7383		2.84
<b>HUMAN CAPITAL, SOCIAL SECURITY AND EMPLOYMENT TYPE FACTOR</b>			
<b>Schooling of oldest wage earner</b>			
No education	1.7031		1.00
Some elementary	1.8445	2.358	1.08
Complete elementary	2.5825		1.52
Some secondary	3.8096		2.24
Complete secondary	4.6045		2.70
Some higher education	8.0938		4.75
Complete higher education	9.0293		5.30
Graduate studies	11.9863		7.04
<b>Mean schooling for people 12 years old and older</b>			
0	1.8396		1.00
4	2.0229	2.506	1.10
5	2.7251		1.48
10	3.686		2.00
11	5.4033		2.94
15	8.7715		4.77
16	8.8965		4.84
More than 16	11.9551		6.50
<b>Social Security and plant size</b>			
Without Social Security and works alone or works in a plant with 2 to 9 employees	0.9856	1.418	1.00
Without Social Security and works in a plant with 10 or more employees	2.1992		2.23
With Social Security and works alone or doesn't work	3.1973		3.24
With Social Security and works in a plant with 2 to 9 employees	4.0801		4.14
With Social Security and works in a plant with 10 or more employees	4.8213		4.89

<b>PUBLIC UTILITY SERVICES FACTOR</b>				
<b>Water supply system</b>				
River or fountain		3.4385		1.00
Public supply		2.7891		0.81
Well without water pump, container or rain water		3.1143	4.498	0.91
well with water pump		4.1416		1.20
Container truck		2.2251		0.65
Water pipe		6.1074		1.78
<b>Sewage disposal</b>				
No sewage		1.0781		1.00
Latrine		1.9192	2.445	1.78
Toilet without conection to sewer or septic tank		2.8726		2.66
Toilet with conection to septic tank		3.5381		3.28
Toilet with conection to sewer		5.3506		4.96
<b>Garbage disposal</b>				
Yard, lot, river, etc.		0.9982	1.214	1.00
Local container or public trashcan		2.1401		2.14
Picked up by public services		2.9805		2.99
<b>DEMOGRAPHIC, UNEMPLOYMENT, INCOME AND ROOM CROWDING</b>				
<b>Proportion of children below 6 years old</b>				
greater than 0.65		1.7451		1.00
from 0 to .65		1.9819	2.459	1.14
cero		3.0073		1.72
<b>Proportion of employed family members</b>				
less than 0.30		0.9905	1.976	1.00
from 0.30 to 0.60		2.0356		2.06
from 0.60 to 0.90		2.7959		2.82
greater than 0.90		4.0811		4.12
<b>Number of rooms per person (Hacinamiento)</b>				
Less than 0.30		1.0564		1.00
0.30 to 0.60		1.9651	2.262	1.86
0.60 to 1.00		2.9829		2.82
1.00 to 4.00		4.0371		3.82
greater than 4.00		6.0596		5.74
<b>Per capita income in minimum wage units</b>				
less than 0.15		0.8722		1.00
from 0.15 to 0.25		2.0713		2.37
from 0.25 to 0.35		2.9893	3.26	3.43
from 0.35 to 0.50		4.085		4.68
from 0.50 to 0.75		5.1777		5.94
from 0.75 to 1.00		6.1982		7.11
from 1.00 to 1.25		6.7461		7.73
from 1.25 to 2.00		7.0264		8.06
from 2.00 to 3.00		9.2969		10.66
from 3.00 to 4.00		10.1563		11.64
greater than 4.00		10.9922		12.60

SOURCE: Velez y Moreno, 1994

order- ceiling materials, household appliances, proportion of employed, garbage collection and proportion of kids less than 6 years old.<sup>39</sup>

Are the variables selected for the SISBEN Index appropriate to distinguish the well-being of the poor and the non poor? Previous studies on income distribution and poverty suggest a positive answer. For Latin-American economies Psacharopolous et al (1997) identified Human capital and employment category as the main determinant factors of labour income inequality and poverty. In the Colombian case the World Bank's Poverty Assessment (1995) reports similar findings. Grosh and Baker (1995) in their comparative evaluation of alternative welfare indicators for Jamaica found that the variables about housing characteristics –in dichotomous form- entered early in a stepwise procedure and contributed to reduce the leakage in targeting –type II error-. In addition, they found family characteristics -including human capital endowment- to be extremely powerful in reducing undercoverage or exclusion errors -type I-. In the case of SISBEN, the set of variables on housing characteristics were included with rich categorical information and performed well, both in the rural and urban sectors.

The results reveal some interesting comparisons between the urban and rural sectors. First of all, the rural sector has an additional variable, namely, *Time spent to collect water*. As explained in the previous subsection and the Appendix, Oblique Factor Analysis creates groups of close variables and ranks those groups according to their power to explain the *total* variability of the data. In fact, the groups generated by the algorithm are the same. However, the potential of these groups of variables as welfare predictors differs significantly in both areas. In the urban area four groups were *clearly*<sup>40</sup> distinguished and in order of importance they are (i) human capital and social security and employment; followed by (ii) Current Income, Crowding, proportion employed and minors in the house, (iii) House building materials and household appliances (iv)

---

<sup>39</sup> The reader should not conclude that the variables with larger score ranges have necessarily a greater contribution to the total variability of the data or the first principal component. This property is better established with the results of Factor analysis, below.

the quality of public utility services.<sup>41</sup> By contrast in the rural area Housing characteristics increases its importance and reaches the first level, while human capital *et al* descend to second place. At the same time the group of income per capita *et al*, go from the second to the fourth place and services to the dwellings moves from fourth to third level. These results are consistent with the observation of rural analysts that income is subject to more instability and seasonal factors that may introduce some noise and reduce the power of *current income* to differentiate welfare levels.

As a proxy means test index, SISBEN belongs to the class of indexes similar to the OLS estimated for Jamaica. In their construction algorithm they clearly differ from the welfare indicators designed explicitly and ex-ante to maximize poverty reduction.<sup>42</sup> Grosh and Baker's (1995) comparative tests of Proxy Means Index *vis a vis* the other type algorithms are encouraging, because the former indicator, complemented with a suitable cut-off point for eligibility like the SISBEN case, performs almost identically both in terms of undercoverage and leaks.

#### 4. Conclusions

We have shown that the proxy means test SISBEN has expanded at a fast rate and is used in a variety of social programs in many municipalities in Colombia. The main factors behind this decentralized success story seem to be the low marginal screening cost for the subnational authorities and the certain benefits that nationally financed social programs could bring to local communities.

The algorithm used -Alternating Least Squares and Optimal Scaling- to derive the weights and the scaling of the variables included in SISBEN is consistent with the challenge of

---

<sup>40</sup> Notice that in Table A.1A the Factor analysis matrix values that are not in bold are relatively closer to zero.

<sup>41</sup> Finally notice how close -coefficient 0.2727- is Room Crowding to the third factor that corresponds to house characteristics.

building a welfarist index of an unobservable variable by using a significant set of categorical data about consumption of durables. The results showed that from the total variability in the sample data the SISBEN index is able to explain is little bit less than 50%.

SISBEN incorporated valuable qualitative data on building characteristics of the dwellings and their corresponding availability of public utilities services. Results show that it is necessary to distinguish *many* quality levels within those variables and that their metric could account for nearly 40% of the total number of points in the SISBEN score.

Although SISBEN produces similar sets of “close” variables in the urban and rural sectors, the grouping is much cleaner or shows less noise in the urban than in the rural areas. In addition, the potential of these groups of variables as welfare predictors differs significantly in both areas. While in the urban sector human capital, employment type, status and characteristics and current income are of primary importance, in the rural areas these signals are noisier than housing characteristics. More telling is that in the rural areas the group that includes employment status and current income drops from second place to fourth in importance to explain the total variability of the data.

Although income was included in the initial set of variables of SISBEN to obey a relative consensus among Colombian poverty experts in 1993, considering the strong incentives of households to misrepresent income in order to obtain favorable rankings suggest that this variable could be removed if a new welfare index was to be constructed. In this case the use of categorical information would not necessarily require the Optimal Scaling algorithm. Income could represent welfare –avoiding assumptions about a parametric utility function- and the most convenient method build a welfare index would be to construct a discriminant function in order to quantify all the variables that enter in categorical form.

If we accept the adequacy of these Optimal Scaling procedures to incorporate relevant categorical data to predict welfare levels, the next logical step should be to exploit the wealth of categorical census data in Latin-America to improve the quality and level of detail of information used for regional targeting. For example, in the same case of Colombia some transfers from the national to the subnational level are based on regional welfare indicators. The impossibility of

---

<sup>42</sup> Ravallion and Chao(1989).

obtaining reliable income information at a desegregated subnational level forces the authorities to use indicators like the NBI, obtained from census data on households. However, based on the same census data set, a better predictor of welfare can be derived using the Optimal Scaling algorithm and, as a consequence, larger equity gains could be obtained from a given amount of compensatory cross-regional transfers. At the moment the Department of National Planning in Colombia has derived an indicator of this type that could eventually replace the NBI Index in the formulas to calculate the intergovernmental resource transfers.<sup>43</sup>

In this paper we have evaluated some significant aspects of the SISBEN system, however we have not studied all important areas. It seems necessary that SISBEN be evaluated *ex-post* in terms of targeting efficiency. Furthermore, at the moment, we have not yet measured the gains associated with the procedure of Qualitative Principal Components. Therefore, (i) some simulations should be performed to measure the losses of the principal component model when the qualitative information deteriorates collapsing in two categories. Additionally, the poverty predictive capacity of SISBEN should be explicitly compared for the Colombian data with indexes derived from other algorithms, like the OLS proxy means test Indexes and others.<sup>44</sup>

---

<sup>43</sup> See Sarmiento, et al (1996, 1997) and Castano (1995). This type of indicators could also be used for geographical targeting within municipalities.

<sup>44</sup> Similar to the tests by Grosh and Baker (1994) for the Jamaican case.

## BIBLIOGRAPHY

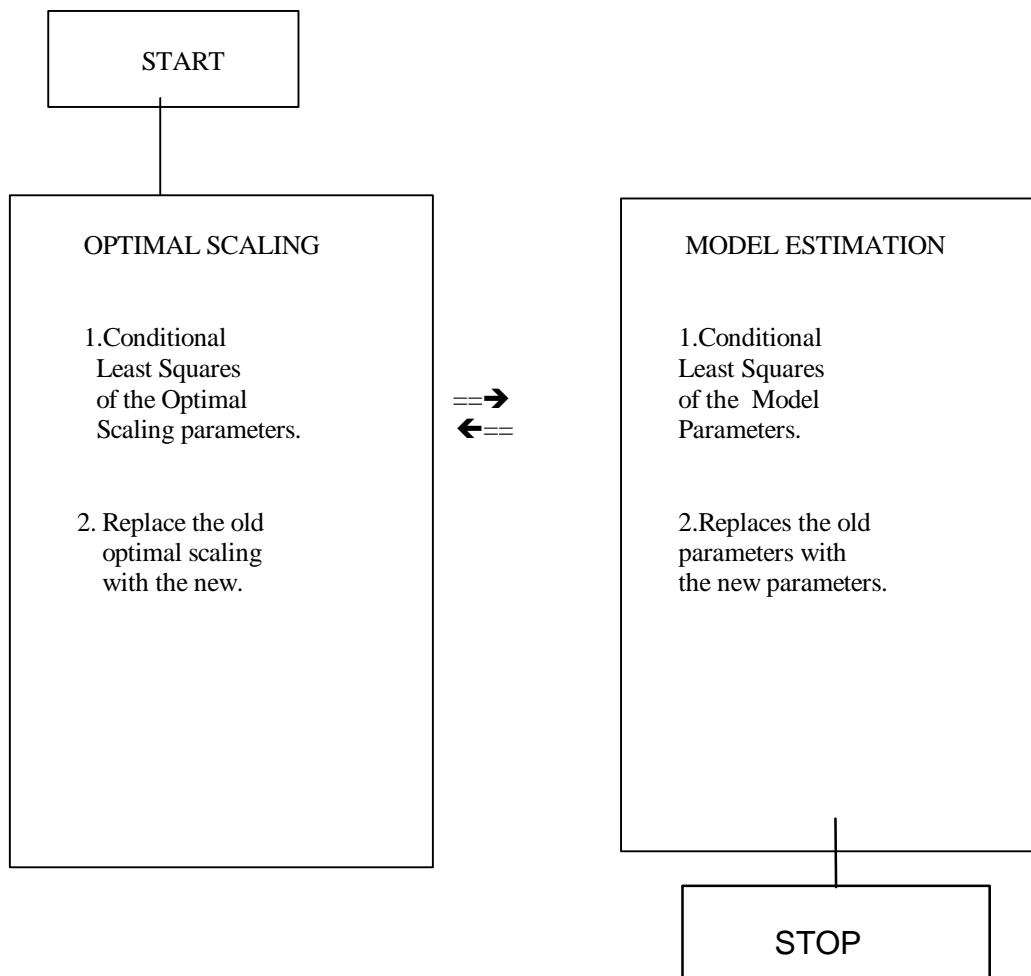
- Anand, Sudhir and Cristopher J. Harris (1994) "Choosing a Welfare Indicator" *AEA Papers and Proceedings*, vol 84 (2) pp.226-231.
- Behrman and Srinivasan, eds.(1995) *Handbook of Development Economics*, volume III. Elsevier B.V.
- Castaño, Elkin y Hernando Moreno (1994) Metodología Estadística del 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, Experiencias Recientes de Selección de Beneficiarios en Costa Rica: una Evaluación del Sisben. Mimeo. CEPAL, abril de 1995.
- De Boor, C. (1978), *A Practical Guide to Splines*, New York: Springer Verlag.
- Deaton, Angus (1997). *The Analysis of Household Surveys: A microeconomic approach to development policy*. World Bank, Washington DC.
- Deaton, Angus, and Muellbauer, John (1980). *Economics and Consumer Behavior*. Cambridge University Press.
- Deaton, Angus. (1995). Data and Econometric Tools for Development Analysis. in (eds.) Behrman and Srinivasan (1995)
- Departamento Nacional de Planeación. (1997). Focalización del Gasto Social. Documento Conpes. Santafé de Bogotá.
- DNP-Misión Social *¿Qué es el ICV? Índice de Condiciones de Vida* Mimeo, DNP-Misión Social Octubre de 1997.
- 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*, 10<sup>th</sup>. ed., Edinburg:Oliver and Boyd Press.
- Glewwe, Paul, and Jacques van der Gaag. 1988. Confronting Poverty in Developing Countries. *Living Standards Measurement Study Working Paper No. 48*. World Bank, Washington, D.C.
- Glewwe, Paul, and O. Kanaan. 1989. Targeting Assistance to the Poor: A multivariate Approach Using Household Survey Data. *Policy, Planning and Research Working Paper No. 225*, World Bank, Washington, D.C.
- Grosh, Margaret and Judy Baker, 1995. "Proxy Means Tests for Targeting Social Programs: Simulations and Speculation " *Living Standard Measurement Study Working Paper No. 118*. World Bank.
- Grosh, Margaret. 1994. *Administering Targeted Social Programs in Latin America: From Platitudes to Practice*. Washington D.C.: The World Bank.
- Grosh Margaret and 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 and 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, and Kalvin Chao. 1989. "Targeted Policies for Poverty Alleviation under Imperfect Information: Algorithms and Applications." *The Journal of Policy Modeling*, 11(2):213-224.
- Ravallion, Martin. 1992. "Poverty Comparisons: Guide to Concepts and Methods" Living Standard Measurement Study Working Paper No. 88. World Bank
- Ravallion Martin and Lipton Michael. (1995). Poverty and policy. In (eds.) Behrman and T.N. Srinivasan.
- Richards, P. and 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 and Acosta Trujillo, Rodrigo (1997). Tipología Municipal con base en las Condiciones de Vida. Mimeo, DNP-Misión Social. Santafé de Bogotá.
- Sarmiento, A., C. Ramirez, C.G. Molina, E. Castaño. (1996) Índice de condiciones de vida. Borrador preliminar para discusión. Departamento Nacional de Planeación. Programa Naciones Unidas para el Desarrollo. Santa Fe de Bogotá.
- Sarmiento, Ramirez, Molina y Castaño (1996) Índice de Condiciones de Vida. Mimeo, DNP-Misión Social, Santafé de Bogotá. Julio de 1996.
- SAS/STAT User Guide (1990), Volume 2, Version 6, Cuarta edición.
- Sen, A. (1979) 'Personal utilities and public judgements: or what's wrong with welfare economics?', *The Economic Journal*, 89:537-558.
- Sen, A. (1987) *The Standard of Living*. Cambridge: Cambridge University Press.
- Sen, A. (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 needs in developing countries*. New York: Oxford University Press.
- Tenenhaus, M. & Vachete, J.L. (1977), "PRINQUAL: Un programme d'Analyse en Composantes Principales 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 : Departamento Nacional de Planeación- 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. Artículos 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 Describing Ordinal Data with Cardinal Models, *Journal of Mathematical Psychology*, 12, 416-436.

- Young, F.W. (1981) "Quantitative Analysis of Qualitative Data," *Psychometrika*, 46, 357-388.
- Young, F.W., Takane, Y. & de Leeuw, J. (1978), "The Principal Components of Mixed Measurement Level Multivariate Data: An Alternating Least Squares Method with Optimal Scaling Figures," *Psychometrika*, 43, 279-281.
- Young, F.W., Takane, Y. & de Leeuw, J. (1985), "PROC PRINQUAL-Preliminary Specifications," non published manuscript, The University of North Carolina Psychometric Laboratory, Chapel Hill NC.
- Zarta, Maria Teresa Avance y Cobertura de la Encuesta Sisben por Departamentos y Municipios Mimeo, DNP-Misión Social Julio de 1997.

## Appendix

### Exhibit A1. ALTERNATING LEAST SQUARES AND OPTIMAL SCALING



### *Oblique Factor Analysis*

Once the quantification of the categories has been done by the PRINQUAL procedure, we use the Factor Analysis technique to identify subgroups of variables that are nearer, and at the same time, to find if correlation exists between the subgroups. The method used is called Oblique Factor Analysis and the factor rotation used is the Promax Rotation.

Next, using the Principal Component Analysis over each subgroup identified by the Factor Analysis, we obtained its first principal component for each group. Subsequently, using each of subgroups poverty indicators, we built the global welfare indicator as the first component of them.

Finally, we obtain an indicator with values in the range 0 to 100, by changing the origin and the scale of the original indicator.

Specifically, let  $X_k$ ,  $k=1,2,\dots,m$ , to be the  $k$ -th discriminant variable, and  $TX_k$  the transformation of  $X_k$  using the PRINQUAL technique. Exhibit 1 summarizes the procedure:

Table A.1A shows the results of applying the Factor procedure on the variables quantified in urban areas. The procedure identifies four groups of near variables in the urban sector. The first factor involves education and social security variables; Demographics and income variables remain in the second factor; the third factor consists of housing and equipment variables; and the last factor are conformed by the housing services variables.

Table A1.B shows the result for the application of the Factor procedure to the variables quantified in the Rural sector. We find four groups of variables. The first one consist of housing and equipment variables; second group contain human capital and social security; third group is conformed by housing services and the last group involves demographics and income variables.

Variable code list	
ESMXMAYA	Schooling of oldest wage earner
ESP12YMA	Mean schooling for people 12 years old and older
SEGURO	Social Security and plant size
PROP6	Proportion of children below 6 years old
PROPOCUP	Proportion of family members employed
HACICUAR	Room crowding (Number of rooms per person)
INGPERC	Per capita income in minimum wage units
MATPAR	Wall Material
MATTECH	Predominating material for roofs
MATPIS	Predominating floor materials
EQUIPA	Home appliances
ABAGUA	Water supply system
TIEMAGU	Time spent in getting water
EXCRET	Sewage disposal
RECOBA	Garbage disposal

Table A1.a. Factor Structure SISBEN Urban. Factor Analysis.  
Rotation Method: Promax

	Variable	Factor 1	Factor 2	Factor 3	Factor 4
1	ESMXMAYA	<b>0.90836</b>	-0.03175	-0.4792	-0.00773
2	ESP12YMA	<b>0.88053</b>	-0.01386	0.01269	-0.00341
3	SEGURO	<b>0.87049</b>	-0.07371	-0.05914	0.09877
4	PROP6	-0.25828	<b>0.79564</b>	0.06812	-0.03747
5	PROPOCUP	0.03787	<b>0.78125</b>	-0.2825	0.03544
6	HACICUAR	0.09804	<b>0.62166</b>	<i>0.27275</i>	-0.05088
7	INGPERC	0.48271	<b>0.52543</b>	0.05714	0.01918
8	MATPAR	-0.09836	-0.05693	<b>0.69726</b>	0.11222
9	MATTECH	-0.12761	0.01911	<b>0.69601</b>	0.0352
10	MATPIS	0.13294	0.0842	<b>0.62003</b>	0.11964
11	EQUIPA	0.33095	-0.11653	<b>0.55854</b>	-0.19687
12	ABAGUA	-0.033	-0.05032	-0.073	<b>0.81253</b>
13	EXCRET	0.02261	0.0224	0.16428	<b>0.75977</b>
14	RECOBA	0.08302	0.00346	0.04283	<b>0.73218</b>

Source: Castaño y Moreno (1995).

Table A1.b. Factor Structure, SISBEN Rural  
Rotation Method: Promax

	Variable	Factor 1	Factor 2	Factor 3	Factor 4
1	MATTECH	<b>0.75109</b>	0.11833	0.31109	0.14576
2	MATPAR	<b>0.79025</b>	0.31929	0.31182	0.11061
3	MATPIS	<b>0.80067</b>	0.37131	0.42947	0.21265
4	EQUIPA	<b>0.64206</b>	0.44566	0.34483	0.0935
5	ESMXMAYA	0.25272	<b>0.88245</b>	0.2191	0.11367
6	ESP12YMA	0.31706	<b>0.88886</b>	0.25764	0.13537
7	SEGURO	0.34258	<b>0.58543</b>	0.24456	0.06216
8	ABAGUA	0.36576	0.21634	<b>0.90109</b>	0.06815
9	TIEMAGU	0.38883	0.19393	<b>0.8807</b>	0.10681
10	EXCRET	0.58471	0.47239	<b>0.70668</b>	0.17092
11	RECOBA	0.28164	0.43371	<b>0.54987</b>	0.0717
12	PROPOCUP	0.02521	0.0017	-0.0039	<b>0.82308</b>
13	PROP6	0.0167	-0.0278	0.04722	<b>0.74527</b>
14	HACICUAR	0.32544	0.23509	0.14693	<b>0.74289</b>
15	INGPERC	0.39136	0.47228	0.23786	<b>0.6632</b>

Source: Castaño y Moreno (1995).

**Table A2.a Summary of weights. Urban  
Weight of the Variables with respect to the Factors  
and the Factors with respect to the Model (First principal component)**

Factor	Variable code	Variable Weight	Factor Weight	Variable total weight
Housing	EQUIPA	0.5961	0.7974	0.4753
	MATPAR	0.7095		0.5658
	MATPIS	0.7774		0.6199
	MATTECH	0.6437		0.5133
Services	EXCRET	0.8573	0.6792	0.5823
	ABAGUA	0.7400		0.5026
	RECOBA	0.7920		0.5379
Human Capital and Social Security	ESP12YMA	0.8993	0.6959	0.6258
	ESMXMAYA	0.9105		0.6336
	SEGURO	0.6786		0.4723
Demographic, Income and Occupation	PROP6	0.6792	0.6006	0.4079
	PROPOCUP	0.6884		0.4135
	HACICUAR	0.7452		0.4476
	INGPERC	0.7478		0.4491

Source: Castaño y Moreno (1995).

**Table A2b. SISBEN Rural .  
Weight of the Variables with respect to the Factors  
and the Factors with respect to the Model (First principal component)**

Factor	Variable code	Variable weight	Factor weight	Variable total weight
Housing	EQUIPA	0.6775	0.8094	0.5483
	MATPAR	0.7963		0.6445
	MATPIS	0.8129		0.6579
	MATTECH	0.7293		0.5903
Services	EXCRET	0.7871	0.7816	0.6152
	ABAGUA	0.8580		0.6706
	TIMEAGU	0.8335		0.6515
	RECOBA	0.6207		0.4851
Human Capital and Social Security	ESP12YMA	0.9112	0.7035	0.6410
	ESMXMAYA	0.9135		0.6426
	SEGURO	0.6028		0.4241
Demographic, Income and Occupation	PROP6	0.7126	0.4646	0.3310
	PROPOCUP	0.7999		0.3716
	HACICUAR	0.7625		0.3542
	INGPERC	0.7148		0.3321

Source: Castaño y Moreno (1995).

## Exhibit 2. METODOLOGY OF BUILDING THE SISBEN INDEX

	INPUT	PROCEDURE	OUTPUT
1.	$X_k, k=1, \dots, m$	Use PRINQUAL on the full set of variables	$TX_k, k=1, \dots, m$ quantification
2.	$TX_k, k=1, \dots, m$	Use FACTOR on transformed variables	p subgroups of near variables
4.	$TX_{1j}, \dots, TX_{ni,j}$ $i, j=1, 2, \dots, p$	Principal Component Analysis on variables of each subgroup	First Principal Component $CP1_j$ $j=1, 2, \dots, p$
5.	$CP1_1, \dots, CP1_p$	Principal Component Analysis for the indicators of each subgroup	First Principal Component or Global Welfare Indicator, $CP1_G$
6.	$CP1_G$	Change of origin and scale	SISBEN Typified Global Welfare Indicator

