Estimating basic capabilities: a structural equation model applied to Bolivia

Krishnakumar, Jaya; Ballon, Paola

How to cite

KRISHNAKUMAR, Jaya, BALLON, Paola. Estimating basic capabilities: a structural equation model applied to Bolivia. In: World development, 2008, vol. 36, n° 6, p. 992–1010. doi: 10.1016/j.worlddev.2007.10.006

This publication URL: https://archive-ouverte.unige.ch/unige:41665

Publication DOI: 10.1016/j.worlddev.2007.10.006

World Development Vol. 36, No. 6, pp. 992–1010, 2008

0305-750X/$ - see front matter
www.elsevier.com/locate/worlddev
doi:10.1016/j.worlddev.2007.10.006

Estimating Basic Capabilities: A Structural Equation Model Applied to Bolivia

JAYA KRISHNAKUMAR and PAOLA BALLON *
University of Geneva, Geneva, Switzerland

Summary. — This paper proposes a suitable theoretical framework for operationalizing the capability approach using the latent variable methodology. A structural equation model is specified to account for the unobservable and multidimensional aspects characterizing the concept of human development and to capture the mutual influence among different capabilities. The model is applied to Bolivian data for studying two "basic" capability domains relating to children: knowledge and living conditions. Individual capability indices are constructed from the estimation results and their empirical distributions analyzed. Our results show a strong interdependence between the above capabilities and confirm the role of exogenous factors in their determination.

Key words — capability approach, structural equation model (SEM), human development, education, living conditions, Bolivia

1. INTRODUCTION

Amartya Sen's capability approach constitutes one of the greatest contributions to the socio-economic debate on well-being, quality of life, and poverty (*cf.* Sen, 1985, 1987, 1992, 1993, 1999). By defining human development as the enlargement of an individual's choices in life, it places the human being in the core of the discussion and goes beyond the space of achievements. Thus, the real opportunities that people face (capability sets) play a fundamental role in this freedom-based approach. This definition of well-being makes the capability approach a richer but at the same time a more demanding one at an informational and methodological level compared to standard income-based approaches, thus challenging its operationalization and its empirical applicability.

This paper proposes a suitable theoretical framework for operationalizing the capability approach using a latent variable methodology. A structural equation model (SEM) is specified to take into account the unobservable and multidimensional aspects characterizing the concept of human development based on the capability approach. This framework is a formal attempt to provide an explanatory model for capability levels in different dimensions through a coherent system of causes, effects, and interactions incorporating social, institutional, and individual factors.

The model is applied in an empirical context to study two "basic" capability dimensions in Bolivia: knowledge (being able to be educated) and living conditions (being able to be adequately sheltered). The number of capability dimensions (two) considered in our empirical model is entirely dictated by data availability; however, we believe that it constitutes a first step toward a more complex and complete model of capabilities where the issue of defining a "list of relevant capability dimensions" will also have to be dealt with. [1] By taking more than one dimension, we can allow the model to capture the interdependent nature of capability dimensions. Our data relate to the 2002 MECOVI (*Programa de Mejoramiento de las Condiciones de Vida*) program, a national household survey conducted by the Bolivian

* The authors would like to thank two anonymous referees for the useful comments and suggestions which enabled them to improve the structure and presentation of the paper. Final revision accepted: October 5, 2007.

National Institute of Statistics (INE) with the support of the World Bank. The survey contains information on socio-economic variables for each member of the household aged 7 or more.

There is no need to motivate the choice of dimensions, namely knowledge and living conditions; however, if one were still to do it may it suffice to say that they are given top priorities in the development agenda of all nations and are an important part of the UNDP's Millennium Development Goals (MDG).[2] Using a latent variable approach for the estimation of the above two basic capability dimensions, this paper attempts to identify the determinants of the range of choice faced by individuals or groups in addition to those of their actual achievements. From this perspective, we hope our results will provide a complementary tool for taking into account the enlargement of freedom of choice in public policy making.

The paper is organized as follows. The next section brings out the important theoretical features of the capability approach and presents the latent variable model proposed for its operationalization. Section 3 is devoted to the empirical application and its results. It begins by matching the literature on the determinants of education and living conditions with the capability approach. Next, the data and the empirical model are described. Finally, model results are discussed; capability indices (CI) are derived and analyzed. Conclusions are drawn in Section 4.

2. THEORETICAL BACKGROUND

(a) *Capabilities, standard of living and measurement issues*

This section reviews the conceptual issues of Sen's Capability Approach that need to be understood in order to formulate an appropriate theoretical framework for the assessment of personal well-being.[3] Traditionally, the assessment of standard of living has taken either the opulence or the utility approach. In the opulence view, the standard of living is judged by the command over resources—wealth, income, while in the utility view[4] it is judged by the consumption of commodities. Sen (1985, 1987, 1992) proposes a *capability-functioning approach* for the evaluation of living standard where the notions of utility and opulence are conceived in a non-traditional setting and the notion of *intrinsic freedom* is at the center of discussion. Thus, the standard of living is conceived as a matter of the life that one wants to lead rather than of the resources and means that one has to lead a life.

In this approach, capabilities refer to the real *choices* that a person has to lead the life he/she wants to lead and hence constitute a broader and "richer" concept than his/her actual lifestyle. While functionings focus on achievements—what the person manages to do or to be, that is, his/her states of doings and beings (being sheltered, being educated), capabilities refer to what he/she can *choose* to do or to achieve, that is, the ability to achieve—being able to be sheltered or educated (*cf.* Sen, 1987).

Sen (1987) gives the following formal framework for his approach. Denote by $\mathbf{z}_i$ the commodity vector[5] possessed by any individual i. These commodities in turn have certain characteristics $c(\mathbf{z}_i)$ that the individual makes use of to achieve certain "beings" and "doings" denoted by $\mathbf{b}_i = \mathbf{f}_i(c(\mathbf{z}_i))$ where $\mathbf{f}_i$ characterizes the "making use of" or the "utilization" of the commodities. Thus, the capability set is the set of all possible $\mathbf{b}_i$'s that a person can achieve using any one of the possible $\mathbf{f}_i$'s that he/she can choose from.

Now, one can notice that many elements are unobservable in this framework: the particular characteristics $c(\mathbf{z}_i)$ that enable any person to convert commodities into functionings, the conversion function $\mathbf{f}_i$ which is particular to each individual, the set of possible conversion functions that any individual can choose from and thus the capability set itself. The only input that is observed, and only partially, is the vector of commodities possessed by the individual apart from his/her actual achievements. We say only partially because one can only observe or measure the material or physical commodities and not the intellectual or social inputs that one combines to "make use of" commodities in a satisfactory way, which are in no way to be neglected in this approach. In our opinion, the term "commodities" has to be interpreted in a large sense, including tangible and non-tangible commodities. Thus, unless it becomes possible to observe all possible functionings that an individual *can achieve* and not only the one that *is actually achieved*, it is not feasible to infer a whole set by only observing one element of it.

One therefore needs to follow a different path in trying to represent the freedom content of a capability set (the set of all possible functionings) from which one particular choice is made.

We hasten to add that this does not at all imply that one should not go further in formalizing the problem in greater depth, for instance, by ranking opportunity sets. [6] In this paper, we propose a different methodology that suits the capability framework and at the same time can be practically implemented. This approach is appealing because of two characteristics: it assumes that (a) the capability set or the freedom to choose is not directly observable but manifests itself in many observable indicators; (b) any single indicator can only be a partial measure of the underlying concept. Factor analysis, multiple indicators multiple causes (MIMIC) models and structural equation models (SEM) are all latent variable models that fit into this line of reasoning.

Other non-statistical techniques have also been proposed in this context such as aggregation and scaling of functionings, and fuzzy sets theory. Most of these techniques address the proper measurement of functionings rather than capabilities. Scaling (i.e., a projection of each variable onto a 0–1 range) was employed in the first major operationalization of the capability approach namely the human development index (HDI). We will not go deeper into the alternate approaches in this paper. Among the notable contributions in the field of fuzzy measures one can cite, for instance, Cerioli and Zani (1990), Qizilbash (2002), Cheli and Lemmi (1995), and Chiappero Martinetti (2000). Here, we will only concentrate on the latent variable approach.

The simplest latent variable model is the factor analysis model in which the observed outcomes (functionings) are postulated to be (linear) functions of a certain (fewer) number of latent factors (capabilities). The MIMIC model adds exogenous causes for the latent factors thus providing an explanation of our capabilities. Klasen (2000), Lelli (2001), Kuklys (2005), and Di Tommaso (2007) provide empirical applications of the use of principal components, factor analysis, and MIMIC models in the context of capability approach.

The structural equation model goes beyond MIMIC and one way causal relationships by specifying interdependencies among different capability dimensions while also including exogenous causes. Thus, SEM constitutes the most appropriate framework for the "estimation" of capabilities as it accounts for their simultaneous determination, their dependence on external causes as well as the impossibility of their direct measurement.

One may ask why simultaneity? Indeed, capabilities do influence one another: for instance, no one can deny that enhancing knowledge capability helps in enhancing health capability or living conditions capability. In other words, better choice in terms of education may lead to better awareness of opportunities and hence an enlargement of the capability space in other dimensions. Whether this actually results in higher outcomes or not depends on the social and institutional environment that one lives in, represented by our exogenous influences. Similarly, better health capability may lead to better knowledge capability by giving access to a new range of otherwise inaccessible opportunities. [7]

(b) *A simultaneous latent variable model*

One of the key elements of our operationalization approach is the specification of capabilities as latent (unobservable) variables. We propose a simultaneous latent variable model in order to take into account the interdependent nature of capabilities as well as the presence of exogenous influences. Achievements or functionings in each of the capability dimensions are measured using proper indicators (generally multiple indicators for each dimension). Hence, the structural relationships are completed by adding "conversion" functions which say how capabilities are transformed into functionings and how external influences come into play in this conversion. These conversion functions, which in fact give us the set of $\mathbf{b}_i$'s starting from $\mathbf{z}_i$'s, that is $\mathbf{f}_i(c(\mathbf{z}_i))$ in our above notation, are called "measurement equations."

Thus our general theoretical framework consists of the following features:

(a) Capabilities (y^*) are *latent, unobservable* and interdependent, and are the key *endogenous*variables of our model.

(b) Capabilities are also influenced by a set of observable external causes (x) or *exogenous variables* (social, political, and institutional factors). [8]

(c) Achievements or functionings (y) are measurable and are linked to the underlying capabilities through a set of measurement equations.

(d) These measurement equations also contain *exogenous* elements w (i.e., individual characteristics).

Table 1. *Notations of our general theoretical framework*

Symbol	Dimension	Definition
Variables		
$\mathbf{y}^*$	$m \times 1$	Vector of latent/unobserved endogenous capabilities
$\mathbf{y}$	$p \times 1$	Vector of observed indicators/functionings
$\tilde{\mathbf{y}}$	$p \times 1$	Vector of latent response variables
$\mathbf{x}$	$q \times 1$	Vector of exogenous causes of $\mathbf{y}^*$
$\mathbf{w}$	$s \times 1$	Vector of exogenous factors in the measurement equations
Coefficients		
τ	$(C - 1) \times 1$	Threshold vector ($C - 1$ thresholds for C categories)
Λ	$p \times m$	Matrix of measurement slopes or loadings, relating $\mathbf{y}$ to $\mathbf{y}^*$
D	$p \times s$	Coefficient matrix of exogenous factors
Γ	$m \times m$	Coefficient matrix for latent endogenous capabilities
B	$m \times q$	Coefficient matrix of exogenous causes
Covariance matrices		
Φ	$p \times p$	Covariance matrix for the residuals in the measurement equations
Ψ	$m \times m$	Covariance matrix for the residuals in the latent variable equations

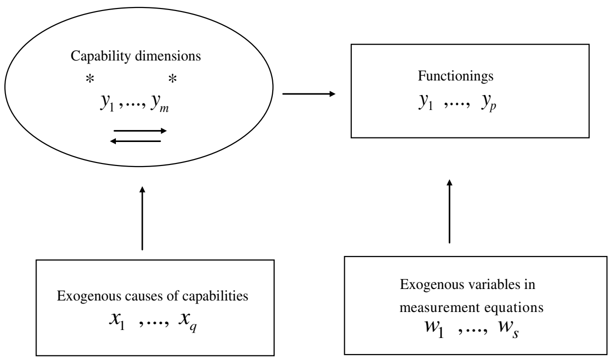


Diagram 1. *Structure of the general theoretical framework.*

Table 1 and Diagram 1 formalize this framework. All the variables are expressed as elements of corresponding vectors, the observed variables are enclosed in boxes and the unobserved or latent variables are circled. Straight single-headed arrows represent causal relations between the variables connected by arrows in the direction shown by the arrow. Two straight single-headed arrows connecting two variables signify feedback relation or simultaneous influence. [9]

(c) *The econometric model*

The above general theoretical framework leads to a general mixed (latent and observed) simultaneous equation model that can operationalize the capability approach. This generalized structural equation model consists of a *structural part* that shows the influence of latent variables on one another and that of exogenous variables on them, and a *measurement part* that specifies the relationships between functionings (observed variables) and capabilities (latent variables). Structural equation models have been extensively used in psychometrics (*cf.* for instance, Bollen, 1989; Muthén, 2002; Skrondal & Rabe-Hesketh, 2004) and more recently in econometrics (see e.g., Di Tommaso, 2007; Di Tommaso, Raiser, & Weeks, 2007; Krishnakumar, 2007).

Following Bollen (1989) and Muthén (1984, 1998–2004), our framework is formalized by the following two sets of equations. The first set of equations represents the latent variable model or the structural model (1) and the second set of equations (2) forms the measurement model. We specify the model as

$$\Gamma \mathbf{y}_i^* + B\mathbf{x}_i + \boldsymbol{\varepsilon}_i = 0, \tag{1}$$

$$\mathbf{y}_i = h(\mathbf{y}_i^*, \mathbf{w}_i) + \boldsymbol{\zeta}_i, \tag{2}$$

where i denotes the individual, $\mathbf{y}_i^*$ is a $(m \times 1)$ vector of capability dimensions, $\mathbf{y}_i$ is a $(p \times 1)$ vector of functionings or indicators, and $\mathbf{x}_i(q \times 1)$, $\mathbf{w}_i(s \times 1)$ are vectors of exogenous variables.

As the vector of functionings/indicators $\mathbf{y}_i$ may include different types of indicators *viz.* continuous and ordered categorical, depending on the empirical context, we specify a nonlinear relationship for the measurement part. If all the observed indicators in $\mathbf{y}_i$ are continuous, then the relationship (2) can be written as

$$\mathbf{y}_i = \Lambda \mathbf{y}_i^* + D\mathbf{w}_i + \boldsymbol{\zeta}_i. \tag{3}$$

However, in the presence of qualitative indicators, the nature of the function $h(\cdot)$ depends on the type of indicator—dichotomous or categorical. For simplicity of notations let us just consider a single element, say the jth one, of the $\mathbf{y}_i$ vector and denote it as y_{ij} (j denotes the indicator). Let the corresponding latent variable be y_i^*. In this case, one introduces a corresponding continuous *latent response* variable $\tilde{y}_{ij}$ such that

$$\tilde{y}_{ij} = \lambda_j y_i^* + d_j' \mathbf{w}_i + \zeta_{ij}. \tag{4}$$

Then, this latent response variable $\tilde{y}_{ij}$ is linked to the *observed* indicator y_{ij} as follows:

- For a dichotomous indicator y_{ij}, say literate or not, we have

$$y_{ij} = \begin{cases} 1 & \text{if} \quad \tilde{y}_{ij} \geq 0 \quad \text{(say literate)} \\ 0 & \text{if} \quad \tilde{y}_{ij} < 0 \quad \text{(not literate)} \end{cases} \tag{5}$$

- For an ordered categorical indicator with C categories (say different levels of education), y_{ij} will take the values $0, 1, \ldots, C-1$ according to the interval to which the latent response $\tilde{y}_{ij}$ belongs. We therefore have:

$$y_{ij} = c \quad \text{for } \tau_{c,j} < \tilde{y}_{ij} \leqslant \tau_{c+1,j} \quad \text{with}$$
$$c = 0, 1, \ldots, C-1, \quad \tau_{0,j} = -\infty, \tau_{C-1,j} = \infty. \tag{6}$$

The stochastic assumptions of the model are as follows:

$$E(\boldsymbol{\varepsilon}_i) = 0, \quad E(\boldsymbol{\zeta}_i) = 0,$$
$$V(\boldsymbol{\varepsilon}_i) = E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i') = \Phi,$$
$$V(\boldsymbol{\zeta}_i) = E(\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i') = \Psi,$$
$$\boldsymbol{\zeta}_i \quad \text{uncorrelated with } \boldsymbol{\varepsilon}_i, \Gamma \text{ non-singular.}$$

Table 1 gives a definition of all the notations used in our model. Factor loadings (λ_{ij}) give the magnitude of the expected change in the observed indicator or outcome for one unit change in the latent variable or capability. These coefficients are the regression coefficients for the effects of capabilities on outcomes. The simultaneous nature of capabilities is emphasized by the Γ coefficient matrix. The effects of exogenous causes in structural equations and those of the exogenous variables in the measurement equations are respectively given by the coefficient matrices B and D (or the parameters associated with $\mathbf{w}_i$ in the $h(\cdot)$ function), respectively.

Using the stochastic assumptions, in particular, the variance covariance matrices of the error terms, one can obtain the theoretical expressions of the variance matrix of $\mathbf{y}_i$, $\boldsymbol{\varepsilon}_i$, $\boldsymbol{\zeta}_i$ in terms of Γ, B, Λ, D, Ψ and Φ say

$$\Sigma = \Sigma(\theta), \tag{7}$$

where θ is a vector that contains all the distinct elements of the unknown parameter vectors and matrices of the model, that is, τ, Λ, A, B, Γ, Φ, Ψ. Σ is the population covariance/correlation matrix of the observed variables $\mathbf{y}_i$, $\mathbf{w}_i$, $\mathbf{x}_i$. $\Sigma(\theta)$ is the covariance/correlation matrix written as a function of θ.

In order to estimate the parameters of the model, we must be sure that the model is identified (exactly identified or over-identified). Among the tests for identification, the t-rule is widely applied due to its simplicity. However, this rule is a necessary but not a sufficient condition of identification. The rule compares the number of non-redundant elements in the covariance/correlation matrix of the (latent) response variables with the number of unknown free parameters in θ denoted as t. If the former is greater than or equal to the latter, the model is identified (over-identified or exactly identified, respectively). The difference between these two groups of elements gives the degrees of freedom for the calculation of the χ^2 statistic of model fit. [10]

The t rule for identification is given by:

$$t \leqslant \left(\frac{1}{2}\right)(p+q+s)(p+q+s+1),$$

where p, q, and s are defined in Table 1.

As our general structural model contains two sets of equations it is recommended to apply a two-stage procedure to test for identification. In the first step, we treat the model as a confirmatory factor analysis, that is, we only focus on the measurement equations and we test for identification applying, for example, the t rule. In a second step, we examine the latent variable part and we treat it as if it were a structural equation model in observed variables. In other words we assume that each capability is an observed variable, ignoring the measurement equations. Then, we test for identification using typical identification rules for observed structural equations models, that is, rank and order conditions. If both steps show that the respective parameters are identified then the whole model is identified.

Once the identification conditions have been verified, one can proceed to the estimation of the unknown parameters by minimizing the distance between the theoretical expression of the moments and their empirical counterparts. In doing so, some constraints must be introduced to provide a scale for the latent variables (capabilities). Typically, the scale is given by either setting the variances of the latent variables to be one or giving them a scale in the same units as one of their indicators (outcomes) by constraining one factor loading to be equal to one. In this case, the scale means that *on average* a one unit shift of the latent capability leads to a one unit shift in the corresponding observed outcome.

Assuming a multivariate normal distribution of $\mathbf{y}_i^*$ conditional on $\mathbf{x}_i$, $\mathbf{w}_i$ (so that the first and second order moments suffice), Muthén (1983, 1984) proposes a three–stage procedure using weighted least squares for minimizing the following fitting function:

$$F_{\mathrm{WLS}} = [\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\boldsymbol{\theta})]' G^{-1} [\hat{\boldsymbol{\rho}} - \boldsymbol{\sigma}(\boldsymbol{\theta})], \qquad (8)$$

where $\hat{\boldsymbol{\rho}}$ is a $(1/2)$ $(p+q+s)$ $(p+q+s+1) \times 1$ vector containing sample estimates of the non-redundant two by two correlations between the elements of $\mathbf{y}_i^*$. $\boldsymbol{\sigma}(\boldsymbol{\theta})$ is the corresponding vector for the theoretical covariance matrix. G is the optimal weighting matrix given by a consistent estimator of the asymptotic covariance matrix of $\hat{\boldsymbol{\rho}}$ of order $m \times m$, where $m = p + q + s$.

The three stage procedure can be summarized as follows:

Step 1: Estimation of the different components of ρ: the thresholds $\boldsymbol{\tau}$ (of $\tilde{\mathbf{y}}_i$) and the conditional moments $E(\tilde{\mathbf{y}}_i|\mathbf{x}_i, \mathbf{w}_i)$ and $V(\tilde{\mathbf{y}}_i|\mathbf{x}_i, \mathbf{w}_i)$. [11]

Step 2: Construction and estimation of G.

Step 3: Estimation of model parameters by minimizing the fitting function (8).

Once the parameters of the model are estimated, the latent variable scores, in other words, the capabilities $\mathbf{y}_i^*$ of each individual i can be obtained (for both knowledge and living conditions). The estimates of $\mathbf{y}_i^*$ can be computed either by the empirical Bayes estimation method or by the maximum posterior likelihood method (see, Krishnakumar, 2008; Krishnakumar & Nagar, 2008; Skrondal & Rabe-Hesketh, 2004).

3. APPLICATION

(a) *Determinants of education and living conditions*

Human capital theory has traditionally been the underlying theoretical framework for the study of education achievements and attainments. In this theory, determinants of human capital accumulation are given by household decisions resulting from the maximization of the household's utility function under the constraints given by the household's human capital production function, income, wage, and time allocation (*cf.* Appleton, 2000). The demand for human capital resulting from this maximization process shows the presence of both demand and supply factors as determinants. The demand factors are observed household, individual, and social characteristics as well as unobserved individual characteristics (abilities, tastes, traditions, or preferences). The supply factors refer to policy and infrastructure variables and comprise the availability and quality of educational and health services.

As mentioned earlier, the capability approach stipulates that capability (knowledge in this case) is affected by the command over resources as well as individual and social factors, which are also present in the human capital framework as supply and demand factors, respectively. Thus, we can use the human capital literature to support the choice of our exogenous determinants of knowledge within the capability framework. However, the latter approach is broader than the human capital one which is only limited to the analysis of achievements.

Among the household characteristics that the literature highlights, [12] one finds parental education, family income, being poor, household size, and structure (average number of people in the household, age of brothers and sisters, and their activities—if working or not, if enrolled or not, characteristics of the head of the family), household conditions (availability and quality of services, habitability), and geographical conditions (rural or urban, living in a main city). Individual characteristics comprise age, gender, being indigenous, working status. Community characteristics refer to community participation in policy design. The education indicators commonly used in this literature include enrollment, average years of education, schooling for age, and literacy.

Supply factors include indicators of availability and affordability of educational services. The availability is captured by school accessibility and the quality of the service: classroom facilities, teacher quality, instructional material, and class size; whereas the affordability is represented by user fees. [13]

Turning to living conditions, there is ample evidence in the social (sociological, social psychological) and medical literature showing that living conditions form an integral part of a person's well-being. An important finding of the former literature is that a person's well-being in this dimension is constrained by the demands of the other members of the household. Gove, Hughes, and Galle (1979) found a clear correlation between the number of persons per room and individual's mental and physical health in the United States. The medical literature, for instance, Britten, Davies, and Colley (1987), Rasmussen, Borchsenius, Winslow, and Ostergaard (1978), and Mann, Wadsworth, and Colley (1992), has found that individuals living in overcrowded conditions show respiratory insufficiency and pulmonary problems more frequently than others. Since we do not consider health as a basic capability dimension in our model due to lack of data, the above results allow us to partially capture it in the living conditions component.

The factors affecting household living conditions highlighted by the above two branches of literature can also be classified into supply and demand side determinants. The former includes government's investment in the provision of basic services and infrastructure and the latter essentially household size, composition, and economic status.

The empirical indicators used to measure living conditions cover the degree of overcrowding (e.g., number of persons per room), access to and quality of the provision of basic services and quality of the dwelling structure (Vos, 1992).

(b) *Data*

The data used in this analysis come from the 2002 Bolivia's MECOVI program, a National Household Survey conducted by the National Statistical Institute with the support of the World Bank. Created in 1999, its main goal is to improve data collection on Bolivia's living conditions for the construction of poverty measures and for the design of social policies that contribute to enhancing Bolivia's household welfare.

The 2002 survey covers 5,952 households and 24,933 individuals and contains information at a national and regional level, on education, health, migration, labor, income, household characteristics, and living conditions. Information is collected by interviewing household members who are asked to fill out a multi-thematic questionnaire.

As we are interested in including supply variables (exogenous variables) in our analysis, we complement the information of the MECOVI 2002 survey with information from the National Institute of Statistics (INE) on social investment and school conditions at the municipal level. Our level of analysis being the individual one, information on living conditions and supply variables available at the household and municipal levels, respectively, are projected onto the individual level. [14] Our sample size comprises 5313 enrolled primary school children aged 7–14.

(c) *The empirical model*

We now present the econometric model specified for estimating children's knowledge and living conditions capabilities in Bolivia. The model is given by equations (1) and (2) that are, respectively, the structural and measurement equations described in the theoretical part of the paper. The path diagram of the specified model is represented in Appendix A.1.

The top part of Table 2 below contains the capability dimensions and the indicators used for the measurement model. These are selected according to their informative content on educational and living conditions achievements and data availability in the household survey under investigation. We have three indicators for educational achievements: literacy, level of education and schooling for age (SAGE). The SAGE variable reflects the "lag" or the lack of progress in a child's schooling with reference to a "normal" achievement rate (see Psacharopoulos & Yand, 1991). It is computed

Table 2. *List of variables*

$\mathbf{y}^*$	Capability	$\mathbf{y}$	Functionings—observed indicators
y_1^*	Knowledge	y_1	Literacy
		y_2	Level of education
		y_3	Schooling for age (SAGE)
y_2^*	Living conditions	y_4	Dwelling conditions
		y_5	Habitability conditions
		y_6	Basic services conditions
$\mathbf{x}$	Observed exogenous causes in the structural model	$\mathbf{w}$	Observed exogenous variables in the measurement model
	Supply side factors		Demand side factors
	Municipal characteristics—district data		*Individual characteristics*
x_1	Number of schools	w_1	Age (child)
x_2	Number of classrooms	w_2	Gender (male/female)
x_3	% of social investment	w_3	Working status
x_4	% of agricultural population in the active population	w_4	Being indigenous
x_5	Water coverage	w_5	Being poor
x_6	Access to radio, TV, phone	w_6	Being the oldest
x_7	Access to electricity	w_7	Number of siblings
x_8	Presence of medical facilities	w_8	Number of siblings aged 7–14
		w_9	Number of siblings aged 7–14 enrolled
	Household encouragement "supply"		*Household characteristics*
x_9	Father's level of education	w_{10}	Number of individuals
x_{10}	Mother's level of education	w_{11}	Number of female adults
x_{11}	Use of medical services	w_{12}	Number of male adults
		w_{13}	Number of children
	Household resources	w_{14}	Male household head
x_{12}	Monthly per capita expenditure		
	Household geographic conditions		*Household geographic conditions*
x_{13}	Urban/rural	w_{15}	Urban/rural
x_{14}	Belongs to main cities	w_{16}	Belongs to main cities

To avoid making our paper too lengthy we do not provide the descriptive statistics of the variables. They are available upon request.

using the following formula: $\text{SAGE} = (S/(A - E))$ where S refers to years of completed schooling, A refers to age and E represents the usual school entry age in Bolivia (6 years). Children with a score under 100 (if computed in percentage) are considered as being below normal progress in the school system because of late entry or dropping out and/or re-enrollment. [15] Literacy measures the ability to read and write and is a dichotomous variable. Level of education is an ordered categorical variable (with three categories: no education with a value of 0, primary incomplete with a value of 1 and primary complete with a value of 2).

Living conditions outcomes are measured by the quality of basic services, and the quality of dwelling and habitability conditions enjoyed by the household. All three indicators are measured by an ordered categorical variable with three categories indicating low, middle, and high quality. Appendix A.2 provides detailed explanations of the coding of these variables.

The bottom part of Table 2 gives the exogenous causes that are present in the latent variable model (1) and in the measurement model (2). As the exogenous variables in the structural equations are mainly factors promoting access, increasing choice, and help in enhancing capabilities, we have only included supply factors in these equations. They are listed on the left hand side. Our information on these variables are at the municipal level, for example, number of schools, number of classrooms, social investment (as a percentage of total expenditure), the presence of a hospital, water coverage and so on. The use of medical services variable is considered to be a supply variable as it reflects the parents' decision to give medical treatment to a sick child. Note that we have also included the parental level of education (especially mother's) [16] in the structural model as it is taken as a proxy for the importance attached to the intrinsic value of education in the society. As the main earning member is usually a male, this variable shows how much the family values a woman's education even when it is not required for working. This variable together with the use of medical services could be interpreted as part of the household's "supply" of encouragement to children.

The exogenous variables appearing in the measurement model are mainly demand side variables, that is, individual and household characteristics that are important in explaining how well the choices or capabilities get transformed into achievements and how the same capabilities can lead to different levels of achievements. They are listed on the right hand side of the bottom part of Table 2. Geographic characteristics are included, in both the measurement and the structural equations, to control for urban/rural households and households located in principal cities. [17]

Knowledge capability [18] (y_1^*) is scaled by fixing the literacy factor loading to one, and living conditions capability (y_2^*) by fixing the dwelling conditions loading to one. This means, that *on average* a unit shift in knowledge capability leads to a unit shift in the literacy indicator value. Similarly, a unit shift in living conditions capability leads to a unit shift in the dwelling conditions indicator.

The dimensions of the various vectors/matrices involved are as follows: $\mathbf{y}^*(2 \times 1)$, $\mathbf{y}$ (6×1), $\mathbf{x}$ (14×1), $\mathbf{w}$ (16×1) and hence Γ (2×2), B (2×14), Λ (6×2) and D (6×16). Thus we have $m = 2$, $p = 6$, $q = 14$, and $s = 16$ in our empirical model. [19] Depending on the presence/absence of certain explanatory variables in each equation, certain elements of the coefficient matrices will be zero (the absence of the corresponding variable). The $\boldsymbol{\theta}$ vector in our estimated model contains 68 unknown parameters after taking account of all the parameter restrictions.

(d) *Estimation results*

This section presents and analyzes the estimation results for our generalized latent variable model specified according to the path diagram in Appendix A.1. [20] The model is estimated according to the procedure discussed earlier and heteroscedasticity-consistent standard deviations are calculated for the coefficient estimators. [21] Individual capability indices (CI) are given by the latent variable scores.

The results of the structural model are presented in Table 3. We report both the normal

Table 3. *Structural model results*

Variable		Knowledge capability equation $\mathbf{y}_1^*$			Living conditions capability equation $\mathbf{y}_2^*$		
		Coefficient	Standardized coefficient	Significance	Coefficient	Standardized coefficient	Significance
$\mathbf{y}_1^*$	Knowledge capability	–	–	–	0.078	0.129	***
$\mathbf{y}_2^*$	Living conditions capability	0.124	0.075	***	–	–	–
x_9	Father's level of education				0.074	0.160	***
x_{10}	Mother's level of education	0.141	0.171	***	0.067	0.134	***
x_{14}	Belongs to main cities	0.109	0.044	***	0.128	0.085	***
x_{11}	Use of medical services	0.170	0.023	*			***
x_1	Number of schools	0.001	0.200	***			***
x_2	Number of classrooms	0.000	−0.273	***			***
x_4	% Agricultural population/PEA				−0.045	−0.016	
x_{12}	Monthly per capita expenditure				1.275	0.457	***
$\mathbf{R}^2$		0.065			0.441		

***,* Denote significance at 1% and 10% levels, respectively.

coefficients and the standardized coefficients [22] as only the latter can be compared in size for variables expressed in different units. Knowledge and living conditions capabilities have a positive effect on each other with the former showing a greater impact on the latter (standardized coefficients of 0.13 and 0.08, respectively). [23] Thus, the simultaneous nature of the two capabilities is confirmed by our results.

Among the significant exogenous influences for the knowledge dimension, one finds mother's level of education, living in cities, use of medical services, number of schools and number of classrooms. As explained earlier, mother's level of education is taken as a proxy for the importance given to education by the society and is supposed to reflect the cultural values attached to knowledge and education. This variable has the strongest influence (based on the standardized coefficient). It is also interesting to note that the father's level of education turns out to be insignificant in our results. Urban environment favors knowledge capability and so do access to health services and the amount of infrastructure available in terms of the number of schools. However, the number of classrooms has a negative impact but its unstandardized value is practically zero. Turning to the exogenous factors in the living conditions dimension, we observe that the higher the parents' (father's and mother's) level of education, the better the physical environment of the children. Once again we would like to interpret this coefficient as the effort taken by "educated" parents to ensure a better quality of life. Note that this effect is different from that of income or household resources which is given by our variable "*per capita* consumption" and which has by far the highest standardized coefficient. The proportion of population in agriculture has a negative impact and living in cities provides better possibilities.

Finally, a word on the quality of fit, before turning to the measurement equations. As can be seen from Table 3, the R^2 value is reasonable for living conditions but rather poor for the knowledge dimension. This is mainly due to the fact that, except for SAGE, the indicators for the latter dimension are not fully satisfactory and do not have enough variability. Hence, the range of the latent variable is neither sufficiently captured by these indicators nor adequately explained by the exogenous factors. However, this does not diminish the importance of the exogenous variables present in the knowledge equation. On the contrary, it is to be noted that in spite of the imperfect representation of the latent variable by two out of the three available indicators, our exogenous variables do show a significant impact.

Table 4 reports the estimation results for our measurement model. [24] Standardized factor loadings for the education measurement equation are all significant and show the right sign, positive for literacy, level of education, and SAGE. Note that the parameters are interpreted based on standardized coefficients so that their magnitudes can be compared with one another. A unit change in the knowledge capability will result in an increase of literacy by 0.61 standardized units, of the level of education by 0.60 standardized units and SAGE by 0.66 standardized units. Thus, the effect is more or less the same on all indicators.

According to the theory that we presented earlier (Section (3a)), the exogenous variables of the measurement equations are given by demand side factors namely, the individual/household characteristics. As far as knowledge is concerned, as the number of siblings increases the education achievements decrease except when these siblings are themselves enrolled. Thus, there is mutual benefit among children enrolled at school though this benefit is countered by children who stay at home and probably need to be looked after. Age has a positive influence on literacy and level of education, as expected and a negative influence on SAGE, also anticipated, due to the lag that accumulates over time for underperforming children, hampering their progress even further.

Children belonging to indigenous communities [25] seem to have a lower level of achievement than others. Other negative influences are noticed for male children, male-headed households (everything else being the same) and working status. [26] Though there is no real reason why male children and children in male-headed households should be disadvantaged, the third phenomenon is pretty obvious as work restricts a child's capacity to progress. R^2 values are high for all the three equations of knowledge indicators.

The results of the measurement equation for living conditions capability are also coherent with the theoretical predictions. Factor loadings of dwelling, habitability, and basic services are positive and significant, all of them of more or less equal magnitude. They indicate an increase of 0.50 standardized units in all

Table 4. *Knowledge measurement equation results (panel A) and living conditions measurement equation results (panel B)*

Variable		Literacy		Level of education		Sage	
		y_1		y_2		y_3	
		Standardized coefficient	Significance	Standardized coefficient	Significance	Standardized coefficient	Significance
Panel A							
$\mathbf{y}_1^*$	Knowledge capability	0.609	–	0.599	***	0.655	***
w_7	Number of siblings	−0.106	***	−0.029	**	−0.068	*
w_8	Number of siblings aged 7–14	−0.091	***	−0.045	***	−0.027	***
w_9	Number of siblings aged 7–14 enrolled	0.453	***	0.166	***	0.189	***
w_1	Age	0.534	***	0.447	***	−0.156	***
w_4	Being indigenous	−0.041	*	−0.029	***	−0.082	***
w_2	Male			−0.033	***	−0.028	***
w_{14}	Male household head	−0.228	***	−0.097	***	−0.111	***
w_3	Working status			−0.039	***	−0.017	*
R^2		0.759		0.567		0.507	
		Dwelling		Habitability		Basic services	
		y_4		y_5		y_6	
		Standardized coefficient	Significance	Standardized coefficient	Significance	Standardized coefficient	Significance
Panel B							
$\mathbf{y}_2^*$	Living conditions capability	0.532	–	0.546	***	0.532	***
w_4	Being indigenous					−0.031	***
w_5	Being poor	−0.052	***	−0.112	***	−0.065	***
w_{14}	Male household head	−0.048	***	−0.050	***	−0.031	***
w_{11}	Number of female adults	0.036	***	0.028	**	0.076	***
w_{13}	Number of children	−0.080	***	−0.233	***	−0.078	***
w_{15}	Urban	0.278	***	−0.076	***	0.436	***
$\mathbf{R}^2$		0.500		0.446		0.695	

***, **,* Denote significance at 1%, 5%, and 10% levels respectively.

achievements resulting from a standardized unit variation of the capability level (Table 4).

The exogenous determinants selected for living conditions indicators also turn out to be highly significant. Being part of the indigenous community and/or poor negatively affects living conditions outcomes, especially the basic services. Once again male-headed households fare worse and the number of female adults seems to improve living conditions achievements. Number of children has a negative impact showing the pressure exerted by extra members on housing amenities. Urban milieu deteriorates habitability conditions whereas it improves dwelling and basic services. This can be expected as urban houses tend to be stronger in terms of construction but at the same time lodge more persons per square meter than rural ones. The R^2 values of all three indicators are high.

Fit indices indicate a reasonable fit of the model. The CFI is 0.968, TLI 0.944, [27] and the root mean square error of approximation (RMSEA) is smaller than 0.06 (0.033 in our case).

Following the model estimation, factor scores or individual capability scores $\hat{y}^*_{il}$ are computed (l denotes the capability dimension and i the individual). Note that our factor scores are normalized ($\underline{\hat{y}}^*_{il}$) following the same procedure applied by the HDI, that is [28]

$$\underline{\hat{y}}^*_{il} = \frac{\hat{y}^*_{il} - \min}{\max - \min}. \tag{9}$$

Though, the minimum and maximum values correspond to those in the sample and will thus change from sample to sample, this issue is not relevant here as we are not making comparisons across samples or over time. However, this problem will have to be solved and suitable solutions found if one wants to make such comparisons.

Figure 1 below shows the histograms of the normalized scores. The distribution of individual knowledge capability estimates is asymmetric with a big peak in the middle and two small peaks on each side. The mean normalized value (0.40) is not far from the median (0.39) and the values are distributed fairly similarly on either side of the center though the graph is somewhat skewed to the right (a skewness coefficient of 0.34). The living conditions distribution is much more asymmetric and skewed to the right (a skewness coefficient of 1.74), with most of the values below 0.40 and a mean normalized value of 0.15.

Thus, one can say that the spread across the population in Bolivia is better in the knowledge dimension in which the majority of individuals are farther away from the minimum than in the case of living conditions in which the individuals are much more concentrated toward the minimum. If we compare the cumulative distribution functions of normalized scores in the two dimensions (Figure 2), then again we find that knowledge dominates living conditions (greater well-being).

From the policy angle, one can get a rough idea of the effect of some exogenous determinants (individual factors and especially supply variables that the government has control of or can influence) on individual capability scores (indices) by looking at some selected plots (Figures 3 and 4). The enhancement of children's

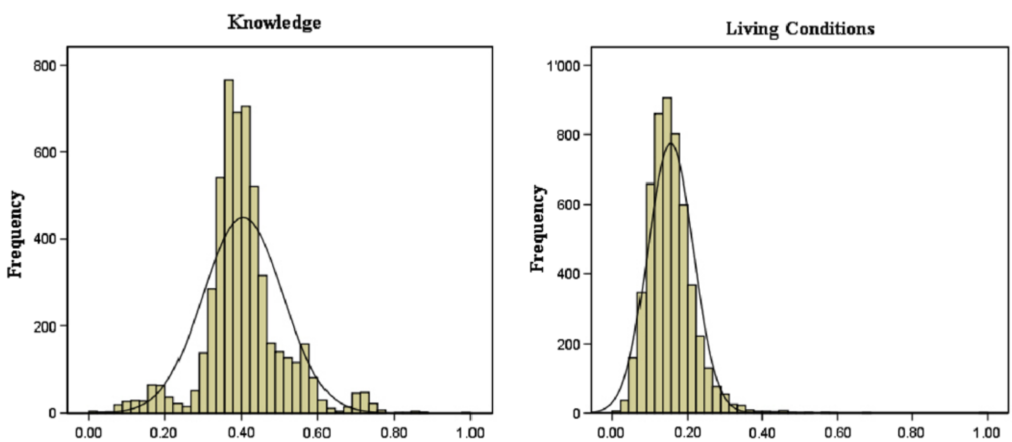


Figure 1. *Histogram of normalized capability scores.*

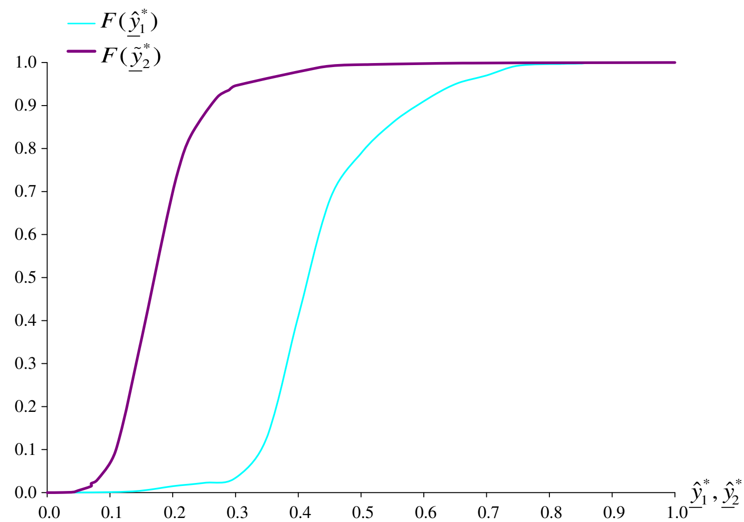


Figure 2. *Knowledge and living conditions—normalized capability estimates cumulative distribution functions.*

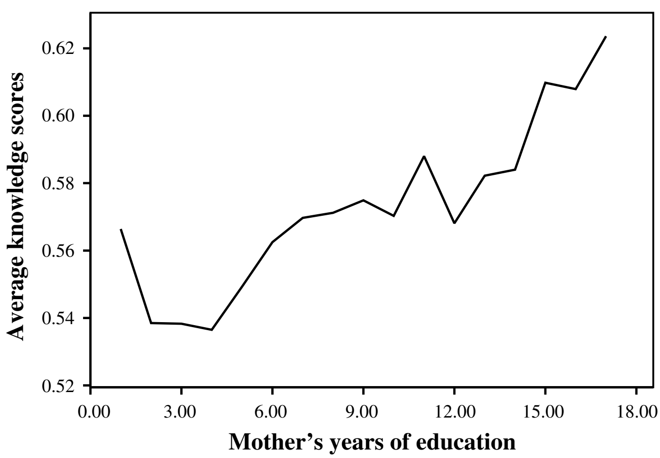


Figure 3. *Children's knowledge scores versus mother's education.*

knowledge capability with the mother's years of education is clearly noted by the positive "slope" that can be detected in the corresponding scatter plot in Figure 3. In this figure, we have plotted the sample mean capability value for each year of education. The initial decrease shows that the positive effect starts only when the mother is at least literate (about 4 years of education).

Similarly, the positive trend of living conditions capability with respect to income (monthly per capita expenditure) shows the positive impact of economic resources for better conditions of living, up to a certain level of income—here 200 bolivianos—which is around the monthly poverty line, beyond which the basic needs in terms of housing seem to be met for most of the households (Figure 4).

Finally, for each dimension, we aggregate the normalized capability scores (indices) for children within a same department, using the multipliers, to obtain an aggregate index at the department level. More specifically we compute a weighted average of the children's scores in a

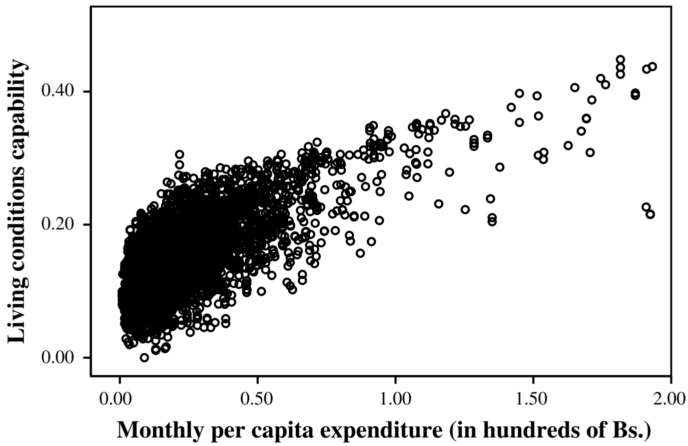


Figure 4. *Children's living conditions scores versus household expenditure.*

Table 5. *Ranking of departments*

Department (alphabetical order)	Rank	Rank	Rank	Rank	Differences	
	Edu. Achiev.	Knowledge	LC Achiev.	LC Cap.	Education	Liv. Cond.
	A1	C1	A2	C2	C1–A1	C2–A2
Beni	8	6	9	8	−2	−1
Chuquisaca	3	4	7	5	1	−2
Cochabamba	4	3	4	2	−1	−2
La Paz	7	5	3	3	−2	0
Oruro	2	2	5	4	0	−1
Pando	5	6	8	6	1	−2
Potosí	9	5	6	7	−4	1
Santa Cruz	1	1	1	1	0	0
Tarija	6	5	2	3	−1	1

department, using the inverse of the survey's selection probabilities (multipliers) as weights. The aggregation exercise is purely carried out to be able to identify zones with poorer "average" performance. As the capability *values* have no intrinsic meaning, the department average has no such meaning either; thus, rankings based on average values should be interpreted under the clause "assuming every individual of the group has a capability level equal to the group average." Here it should be stressed that we are not attempting any inter-personal comparisons which are excluded given the purely ordinal nature of our latent variables.

In Table 5, we present the ranking of the nine departments in the two capability dimensions (denoted as C1 and C2) along with those in terms of functionings (achievements) given by means of aggregate indices of normalized outcome indicators (A1 and A2, respectively).

Looking at the knowledge dimension, the first two departments are Santa Cruz and Oruro in terms of both criteria. Cochabamba and Chuquisaca share the third and fourth positions in the capability dimension but the order is reversed when we look at achievements. Cochabamba is better off in the former and Chuquisaca in the latter. The ranks of La Paz, Potosi, and Tarija are equal in terms of capabilities (their confidence intervals overlap) whereas they differ in terms of achievements. Beni and Pando are the last in terms of capability while Potosi is the last in terms of achievements. This department has a lower rank in achievements than what would be expected based on its capability (a difference of −4). Except for Chuquisaca and Pando, all other departments are in a similar situation though the difference between the two ranks is not as high as Potosi's.

Turning to living conditions, Santa Cruz and Cochabamba occupy the first and second positions in terms of capability followed by Tarija and La Paz of equal rank. These are in turn followed by Oruro, Chuquisaca, Pando, Potosí, and Beni. In terms of achievements the rankings are similar. Most departments have negative differences between the two ranks implying that their achievements are slightly below their capability estimates, on a purely ordinal comparison. La Paz and Santa Cruz fare equally well in both achievements and capability. Chuquisaca, Cochabamba, and Pando have a difference of -2 which is the biggest in this dimension.

To end this section, we would like to add a note on the interpretation of positive and negative differences in ranks from a capability perspective. When a department's achievement rank is lower than the capability rank (negative difference) it might mean that on average, individuals in the department do not fully utilize the range of opportunities available to them (represented by the latent variable scores) and there is scope for improving their achievements. On the other hand, a positive difference means that even though the range of choice is possibly narrower, the actual choice made by individuals leads to functionings of a higher rank. In our framework, the exogenous variables play a crucial role in explaining these differences.

4. CONCLUSIONS

In this paper, we propose a new theoretical framework to operationalize the capability approach through a structural equation model taking into account the interactions and causal factors determining the level of capabilities. The model is applied in an empirical context to study two basic capabilities—knowledge and living conditions—of Bolivian children.

Our results show a strong interdependence between the above two capabilities and highlight the major role played by supply factors such as the availability of schools, social investment, and family support for children. They also confirm the importance of exogenous demand factors such as ethnic belonging, living in rural or urban areas and siblings structure in the "conversion" of capabilities into achievements. Thus, structural equation modeling is seen to be a useful tool in the explanation and "measurement" of individual capabilities.

Regarding policy implications, the "capability indices (CI)" constructed at the department level indicate that Santa Cruz, Oruro, and Cochabamba perform relatively well in the knowledge dimension and Santa Cruz, Cochabamba, and La Paz fare well in the living conditions dimension. La Paz is at the bottom end in the first dimension along with Potosi and Beni, whereas only the latter two remain at the bottom in the second dimension.

One can also attempt similar comparisons by considering other criteria of interest to policy makers, for instance, by comparing average performances across different socio-economic groups or for evaluating gender and indigenous gaps.

NOTES

1. Here we will use the term "capabilities" or "capability set" to refer to the freedom of choice or the set of choices across dimensions, whereas we will either say "capability" or "capability dimension" while referring to the set of choices within a single dimension.

2. To achieve universal primary education and to ensure environmental sustainability through the reduction of the proportion of people without sustainable access to safe drinking water and the improvement in lives of slum dwellers, constitute two of the eight MDG's.

3. Although standard of living and personal well-being are not equivalent (see Sen, 1987), our operationalization approach, based on a latent variable, allows us to use both terms interchangeably.

4. Understood as happiness, pleasure, or desire fulfillment.

5. Variables in bold refer to vectors.

6. There are studies (Pattanaik & Xu, 2000; Xu, 2002) which propose to measure the freedom contained in a capability set by the size and quality of elements in the set. However, we feel that there are many unresolved practical issues in this method, to cite a few (a) the comparison of the qualitative characteristics of the

choices involved and (b) the relevance of the choices, for instance, whether they are feasible, superfluous, or affordable.

7. Further examples of such interdependencies can be found in Krishnakumar (2007).

8. Some of these social, political and institutional factors may in turn be influenced by capabilities themselves in which case they should be treated as observed endogenous variables.

9. For examples of the different sets of variables in the diagram the reader is referred to the path diagram of our empirical model in Appendix A.1.

10. For mixed models (observed continuous and categorical indicators) a "robust" χ^2 goodness-of-fit statistic can be obtained and the degrees of freedom adjusted accordingly (see Muthén (1998–2004)).

11. For continuous indicators, $\boldsymbol{\rho}$ contains the sample covariance matrix of the sample mean vector and the sample covariance matrix. For categorical variables, $\boldsymbol{\rho}$ includes probit thresholds, slopes, and residual correlations computed by a set of probit regressions of each pair of y variables on all exogenous variables and a set of $p(p-1)/2$ bivariate probit regressions of each pair of y variables on all exogenous variables.

12. One can cite, for instance, Patrinos and Psacharopoulos (1995, 1997), Psacharopoulos (1997), Psacharopoulos and Yand (1991), and Drèze and Sen (2002).

13. See the references cited in footnote 12.

14. This implies that all household members face the same living conditions and that all children belonging to the same municipality face the same supply variables.

15. Note, that given the structure of the variable and the possibility of early entrants in the school system the survey exhibits SAGE values greater than 100.

16. The parent's level of education corresponds to an ordered categorical variable with eight possible values: no education, primary incomplete, primary complete, secondary incomplete, secondary complete, higher non-university, higher university, and other.

17. Reflects the household's location in the main-city axe: La Paz, Cochabamba, or Santa Cruz.

18. For simplicity of notation we omit the subscript i in the empirical section.

19. Recall that m, p, q, and s denote, respectively, the number of latent variables, indicators, exogenous causes in the structural model, and exogenous factors in the measurement equations.

20. We only report significant coefficients in our tables of results.

21. These calculations were carried out using the software MPLUS.

22. A standardized coefficient gives the change in units of standard deviations of y for one standard deviation unit change in x. The standardization is done by multiplying the estimated coefficient by the ratio of the standard deviation of the explanatory variable to the standard deviation of the explained variable.

23. All coefficients are rounded to two decimals in the text.

24. We only report the standardized coefficients for comparison purposes.

25. According to the survey, a child is considered to belong to the indigenous population if he/she speaks a native language (e.g., quechua, aymara, or guaraní) or identifies himself/herself as indigenous. For our purposes, we followed Bolivia's 1992 and 2001 Census where the competence in an indigenous language is used to denote the indigenous status.

26. According to the survey, a child is considered to work if he/she has worked at least one hour, inside or outside the household, in the week preceding the survey. Inside household work refers to a child's activity aimed for the household's benefit such as cooking, sewing, knitting or agricultural related tasks.

27. CFI is the comparative fit index of Bentler (1990) and TLI is the Tucker–Lewis index of Tucker and Lewis (1973). The cutoff value for both indices for "good" models is 0.95, see Hu and Bentler (1999).

28. Our latent variables are purely ordinal and hence their values do not have any interpretations as such. Note that the order is invariant to any monotonic transformation.

REFERENCES

Appleton, S. (2000). Education and health at the household level in Sub-Saharan Africa. *Center for International Development Working Paper 33*, University of Harvard. <www.cid.harvard.edu/cidwp/033.htm>.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.

Britten, N., Davies, J., & Colley, J. (1987). Early respiratory experience and subsequent cough and peak expiratory flow rate in 36 year old men and women. *British Medical Journal, 294*(6583), 1317–1320.

Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Cerioli, A., & Zani, S. (1990). A fuzzy approach to the measurement of poverty. In C. Dagum, & M. Zenga (Eds.), *Income and wealth distribution, inequality and poverty* (pp. 272–284). Berlin: Springer-Verlag.

Cheli, B., & Lemmi, A. (1995). A "totally" fuzzy and relative approach to the multidimensional analysis of poverty. *Economic Notes, 24*(1), 115–134.

Chiappero Martinetti, E. (2000). A multidimensional assessment of well-being based on Sen's functioning theory. *Rivista Internationale di Scienzie Sociali, 108*(2), 207–231.

Di Tommaso, M. (2007). Children capabilities: A structural equation model for India. *The Journal of Socio-Economics, 36*(3), 436–450.

Di Tommaso, M., Raiser, M., & Weeks, M. (2007). Home grown or imported? Initial conditions, external anchors, and the determinants of institutional reform in the transition economies. *The Economic Journal, 117*(520), 858–881.

Drèze, J., & Sen, A. (2002). *India: Development and participation* (2nd ed.). Oxford: Oxford University Press.

Gove, W., Hughes, M., & Galle, O. (1979). Overcrowding in the home: An empirical investigation of its possible pathological consequences. *American Sociological Review, 44*(1), 59–80.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Klasen, S. (2000). Measuring poverty and deprivation in South Africa. *Review of Income and Wealth, 46*(1), 33–58.

Krishnakumar, J. (2007). Going beyond functionings to capabilities: An econometric model to explain and estimate capabilities. *Journal of Human Development, 8*(1), 39–63.

Krishnakumar, J. (2008). Multidimensional measures of poverty and well-being based on latent variable models. In N. Kakwani, J. Silber (Eds.), *Quantitative approaches to multidimensional poverty*. Palgrave—Macmillan.

Krishnakumar, J., & Nagar, A. (2008). On exact statistical properties of multidimensional indices based on principal components, factor analysis, MIMIC and structural equation models. *Social Indicators Research, 87*(3).

Kuklys, W. (2005). *Amartya Sen's capability approach: Theoretical insights and empirical applications*. Berlin, Germany: Springer-Verlag.

Lelli, S. (2001). Factor analysis vs. fuzzy sets theory: Assessing the influence of different techniques on Sen's functioning approach. *Center of Economic Studies Discussion Paper*, KU Leuven. <www.econ.kuleuven.be/eng/ew/discussionpapers/Dps01/Dps0121.pdf>.

Mann, S., Wadsworth, M., & Colley, J. (1992). Accumulation of factors influencing respiratory illness in members of a national birth cohort and their offspring. *Journal of Epidemiology and Community Health, 46*(3), 286–292.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22*(1–2), 43–65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*(1), 115–132.

Muthén, B. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika, 29*(1), 81–117.

Muthén, B. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Patrinos, H., & Psacharopoulos, G. (1995). Educational performance and child labor in Paraguay. *International Journal of Educational Development, 15*(1), 47–60.

Patrinos, H., & Psacharopoulos, G. (1997). Family size, schooling and child labor in Peru – An empirical analysis. *Journal of Population Economics, 10*(4), 387–405.

Pattanaik, P., & Xu, Y. (2000). On ranking opportunity sets in economic environments. *Journal of Economic Theory, 93*(1), 48–71.

Psacharopoulos, G. (1997). Child labor versus educational attainment: Some evidence from Latin America. *Journal of Population Economics, 10*(4), 377–386.

Psacharopoulos, G., & Yand, H. (1991). Educational attainment among Venezuelan youth: An analysis of its determinants. *International Journal of Educational Development, 11*(4), 289–294.

Qizilbash, M. (2002). A note on the measurement of poverty and vulnerability in the South African context. *Journal of International Development, 14*(6), 757–772.

Rasmussen, F., Borchsenius, L., Winslow, J., & Ostergaard, E. (1978). Associations between housing conditions, smoking habits and ventilatory lung function in men with clean jobs. *Scandinavian Journal of Respiratory Disease, 59*, 264–276.

Sen, A. (1985). *Commodities and capabilities*. Amsterdam: North Holland.

Sen, A. (1987). *The standard of living*. Cambridge: Cambridge University Press.

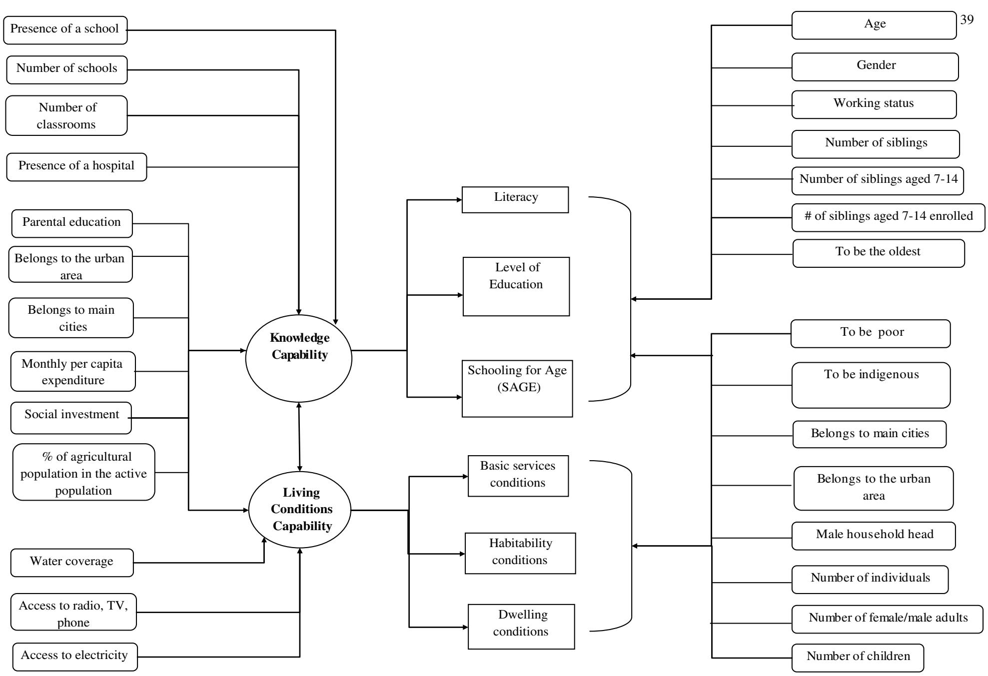
Sen, A. (1992). *Inequality re-examined*. Oxford: Clarendon Press.
Sen, A. (1993). Capability and well-being. In M. Nussbaum, & A. Sen (Eds.), *The quality of life* (pp. 30–53). Oxford: Oxford University Press.
Sen, A. (1999). *Development as freedom*. Oxford: Oxford University Press.
Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.
Tucker, L., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10.
UNDP (1990–2006). *Human development reports*. New York: Oxford University Press.
Vos, R. (1992). *Hacia un sistema de indicadores sociales para América Latina*. Unpublished manuscript, Washington, DC: Inter-American Development Bank.
Xu, Y. (2002). Functioning, capability and the standard of living—An axiomatic approach. *Economic Theory, 20*(2), 387–399.

APPENDIX A.1. ECONOMETRIC MODEL-PATH DIAGRAM[1]

APPENDIX A.2. LIVING CONDITIONS INDICATORS

The indicators of dwelling (y_1), habitability (y_2) and basic services conditions (y_3) of our model are ordered categorical variables with three possible values—low, middle, and high quality—that correspond to a score between minus one and one, with a higher score denoting better quality. This interval is split into three zones. A low quality is reflected by a score lying between minus one to minus one-third, a middle quality by a score ranging from minus one-third to one third, and a high quality by a score of one-third to one.

The dwelling conditions score is obtained by a simple average of the scores assigned to the dwelling materials of the floor, walls, and roof of the house. A score of one (high quality) is gi-



(1) For presentation purposes random errors are not shown in the diagram.

ven to floors built of wood plate, parquet wood, carpet, mosaic, tile, or ceramic; to walls built of plastered mud brick, brick, or concrete blocks; and to roofs built of cement, clay roof tile, or reinforced concrete. A score of zero (middle quality) corresponds to floors built of concrete brick; to walls built of stone; and to roofs built of corrugated iron. A score of minus one (low quality) is assigned to all other materials.

The basic services conditions score is the result of a simple average of the individual scores given to the house's sanitary infrastructure, water supply, and electric connections. The sanitary infrastructure score takes into account the type of excreta disposal—sewage system, septic system, or latrines—and the "ownership" of a toilet—used only by the household, shared with other households, or not having it. A score of one is given to houses that have a sewage system and a toilet used only by the household. A score of zero is assigned to houses with either a septic system or latrines and a toilet used only by the household. A score of minus one is given to houses that share the toilet with other households and with either a septic system or latrines, and to all houses that do not have access to a toilet. The water supply score considers the type of supply and connections. The former distinguishes among water inside the house, outside the house, and outside the lot. The latter differentiates among pipe network/public pipe, cistern, pool, and river, and streams or lake. A water connection inside the house by pipe network or public pipe gets a score of one; while by cistern, pool, or river and streams/lake get a score of zero. Water connections outside the house and outside the lot get a score of zero if by pipe network/public pipe or by cistern and a score of zero and a score of minus one if by pool or river and streams or lake. The electric connections score reflects the availability of electricity (score of one) or the absence of it (score of minus one).

Finally, the habitability conditions score is a weighted average of the scores attached to the degree of overcrowding (measured by the number of people per bedroom and the number of people per room), and to the presence of a kitchen room in the house. A score of one is assigned if there are less than two persons per bedroom, less than four persons per room, and if there is a kitchen room. A score of zero is given to households exhibiting two or three people per bedroom, four to six people per room, and to households that do not have a kitchen room and do not cook. Households with more than three persons per bedroom, more than six persons per room and with no kitchen room (but that do cook) get a score of minus one.