Session Number : 4D

Time: Tuesday, August 26, PM

Paper Prepared for the 30th General Conference of
The International Association for Research in Income and Wealth

Portoroz, Slovenia, August 24-30, 2008

A Model-Based Multidimensional Capability Deprivation Index

Paola Ballon and Jaya Krishnakumar

For additional information please contact:
Paola Ballon or Jaya Krishnakumar
Department of Econometrics
University of Geneva
UNI-MAIL, 40Bd. du Pont d'Arve, CH-1211
Geneva 4, Switzerland.

Email address:
paola.ballon@metri.unige.ch
jaya.krishnakumar@metri.unige.ch

Abstract

This paper proposes a multidimensional capability deprivation index based on a structural economic model (SEM) that explains the capability/deprivation levels in different dimensions, accounting for their multidimensionality and their "unobservable" (latent) nature. Under this framework, the freedom of choice in each capability domain is represented by a latent variable, partially observed through a group of indicators (achievements), and explained by a collection of exogenous variables. The estimators of the different latent variables (scores) provide a measure of the capability levels of the population observed in each dimension. Single and multi-dimensional capability deprivation indices are derived using these scores. The proposed indices are ordinal and fulfill a set of desirable properties in the capability framework. The methodology is applied to analyse the deprivation situation of children in Bolivia in the knowledge and living conditions domains.

Keywords: Capability Approach, Structural Equation Model (SEM), Poverty, Education, Living Conditions, Bolivia.

JEL classification codes: C3, I21, I31, O54.

1 Introduction

In recent decades the assessment of poverty [1] has shifted to a multidimensional perspective to account for the many-sided nature of human deprivation. It is now widely accepted that traditional income-based measures are not informative enough to adequately target the most vulnerable groups, and alternative measures including non income-based measures are necessary for improving the design and effectiveness of anti-poverty policies.

This line of thinking has greatly been influenced by the Capability Approach [2]. By focusing on the real opportunities that people face (capability sets) this freedom-based approach has opened the ground of a novel line of research on the space of capabilities. Though, it is no doubt a more complete framework for poverty assessment compared to approaches based only on achievements or resources or both, its operationalization and practical applicability has been particularly challenging due to its informational and methodological requests.

In addition, the appraisal of poverty in a multidimensional setting brings into consideration several issues, among which the key issues are the selection of the dimensions and their corresponding indicators, and the aggregation of indicators within and across dimensions . Although both are equally important, the latter requires more attention. This leads us to distinguish between the different 'levels' of multi-dimensionality considered in the study of poverty. We would like to stress two of these that are key to our discussion levels the 'indicator' level and the 'dimension' level. While informative, the study of the many-dimensions of poverty without differentiating between the two, might be restrictive. A more suitable level of analysis would be one based on multiple dimensions and multiple indicators. At all levels the choice of the aggregation method (statistical or axiomatic) is important due to its influence on the selection of weights used to obtain the composite index, and the relative importance of the various aspects reflected in the inherent dimensions.

This paper is a contribution along these directions. We propose a multidimensional deprivation index based on an operationalization of the capability approach. This measure, understood as a capability deprivation index, differs from earlier literature on multidimensional poverty assessment in the following: (i) it results from an economic model that explains the different dimensions of poverty instead of simply describing it, (ii) it measures poverty in the functioning-capability space, (iii) it belongs to a class of multiple dimension(capability domains)-multiple indicator(functionings) indices, (iv) it is an ordinal measure fulfilling a set of desirable properties that represent value judgments and ethical principles in the capability framework, (v) it combines statistical and axiomatic methods, at the indicator and dimension levels respectively, for deriving an overall measure of deprivation. We

[1] Although poverty and deprivation are used interchangeably in this section, in the rest of the paper the term deprivation is used to denote poverty in the capability space. This is discussed in detail in Section 3.

[2] Among other approaches towards multi-dimensionality one can cite the Social Exclusion approach (Townsend, 1979; European-Foundation, 1995; Clert, 1999), and the Participatory Approach (Chambers, 1994).

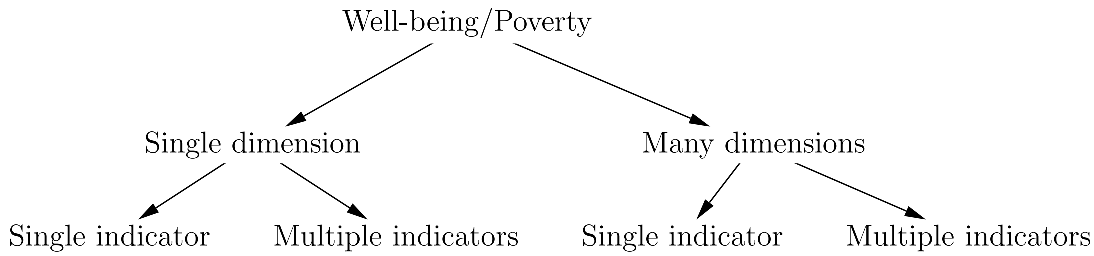
apply it empirically to study the deprivation in the capability domains of knowledge and living conditions for Bolivia's children in 2002. Our results show that deprivation in the single and bi-dimensional cases is more severe in rural areas than in urban and for males compared to females. The bi-dimensional index also shows that deprivation increases as the the degree of substitution between both dimensions decreases.

We begin the paper by discussing the different approaches to multi-dimensionality in poverty assessment in Section 2. Section 3 then describes our operationalization of the Capability Approach through a latent variable model and the estimation of capability indices (factor scores). The study of poverty in the capability space is presented in Section 4. This section first describes the notion of "deprivation" associated with capability poverty and then, proposes two axioms that relate this concept to the economic model of section 3. These axioms are doubly important. They provide the theoretical basis for (i) relying on scores as measures of the well-being and deprivation levels of individuals, (ii) and for defining the individual deprivation function required for the assessment of overall-deprivation. This is treated at the end of the section where emphasis is given to the two levels of multidimensionality addressed in this paper. The resulting measures are ordinal and consider aggregation within a capability domain (single dimension) and between different capability sets (multiple dimensions). Section 5 implements these measures for studying the deprivation in education and living conditions capabilities of Bolivia's children in 2002. Non-parametric bootstrap samples are used to robustify the results of the values of the scores. Section 6 concludes.

2 Different approaches to multi-dimensionality

In the study of the many-dimensions of poverty one can distinguish between two levels of multi-dimensionality: the 'indicator' level and the 'dimension' level. The former, concerns the number of indicators used to summarize the information in a given dimension. The latter, regards the number of dimensions included in the study of poverty. In general their multiplicity leads to four combinations: single dimension-single indicator, single dimension-multiple indicators, multiple dimensions-single indicator, and multiple dimensions-multiple indicators (Figure 1). Single dimension poverty measures have extensively been used in economic poverty studies, where, the emphasis lies on the income/consumption dimension of poverty, and on the use of monetary indicators for its evaluation. The employment of multiple indicators and multiple dimensions in poverty assessment conforms to the awakened interest of practitioners in quantifying the many-sided nature of human deprivation and this is precisely the direction that we have attempted to follow in this paper.

Figure 1: Levels of Multi-dimensionality



The appraisal of deprivation through synthetic measures involves several stages going from the selection of dimensions and the corresponding indicators to data normalization, weighting and aggregation (Nardo et.al (2005)). For the purpose of this paper, we concentrate on the weighting and aggregation procedures. In any empirical study, the selection of dimensions and indicators is generally dictated by the availability of information unless of course one can conduct the survey oneself. Normalization issues are usually limited to a simple empirical minmax method and it is beyond the scope of this paper to compare different procedures in this connection.

The aggregation process in a multidimensional setting involves two steps at the individual (household) level. The first step deals with aggregation of deprivation at the 'indicator' level. This aggregation is necessary because of the presence of multiple indicators, which undoubtedly offer a more complete framework for the assessment of poverty in a given dimension, but become overwhelming unless they are adequately summarized. The resulting measure is some composite index of the indicators in question for each dimension. The second step concerns the derivation of an overall deprivation measure combining different dimensions[3] . In this case, the aggregation procedure will depend on the number of dimensions implicit in the analysis, and on the desirable properties to be respected while carrying out their aggregation. Regarding weighting methods, one can generally distinguish between two types of multidimensional indices - composite indices with exogenous weights and model- based indices with endogenous weights.

Typically composite indices are multidimensional aggregates using exogenous weights like the Human Development Index (HDI). Model-based indices are those with endogenous weights derived from an underlying underlying structure of causes and interactions among variables. Table 1 summarizes the above groups.

Before discussing them in detail, it is important to note that model-based indices are more often employed for summarizing information contained in the different indicators within a dimension. Aggregation across dimensions is usually performed

[3]To deal with this issue researchers often allow for a degree of substitutability or complementarity between dimensions. See for example Bourguignon and Chakravarty (1999, 2002, 2003), Tsui (2002) and Maasoumi and Lugo (2008)

using axiomatic criteria, although statistical procedures (model-based) may also be theoretically applied. The preference for normative exogenous weights when combining different dimensions derives from ethical considerations characterizing the determination of the relative importance of different dimensions. From an ethical point of view, the assignment of weights should rely on some normative principles that the society would like to follow.

Table 1: Summary of Multidimensional Indices

			For model based indices	
Index		**Type of weight**	**Exogenous Causes**	**Simultaneous relations among latent variables**
Composite or crude measure		Exogenous	no	no
Principal Components		Endogenous but no underlying model	no	no
Model-based				
	FA	Endogenous	no	no
	MIMIC	Endogenous	yes	no
	SEM	Endogenous	yes	yes

2.1 Composite indices with exogenous weights

Such indices are characterized by the subjective nature of weighting and aggregating structures as it is the analyst or the user who decides on the weights of the different components and the aggregation procedure. In other words, the aggregation scheme and the weights are selected exogenously. The criteria used for the choice of indicators are often based on the relevance and importance of the indicator to the concept under study. The Human Development Index (HDI) developed since 1990 by the United Nations Development Programme is a clear example of this type of index. It follows a broad definition of human development and gives equal weight (1/3) to three dimensions of human life -longevity, knowledge and decent living standards; further these dimensions, expressed as normalized (0-1) indices, are aggregated by a simple mean (UNDP, 1990). Other examples could be found in Bandura(2006) who surveys 130 indices and Nardo et.al (2005) who examine the different stages in the construction of composite indices. It is important to notice that composite indices lack an explanatory model, their theoretical framework relies exclusively on the researcher's assessment of the real-world phenomenon.

In addition to equal weighting schemes which produce simple averages, we also find in the literature unequal exogenous weighting structures such as the multidimensional human poverty index (Anand and Sen, 1997) where the weights are unequal but still decided unilaterally by the analyst. Typically they involve some parameter whose value is often dictated by normative judgments. Notice that here we refer to a multi-attribute (dimension) unequal weighting structure for a single individual.

2.2 Model-based indices and principal components

Unlike composite indices, model-based indices are derived from an underlying theoretical model that offers an explanation for the inclusion of the variables composing the index as well as a theoretical justification for the choice and values of the weights in the construction of the index. The idea behind these indices is that the theoretical concept that we are trying to measure is not directly observed (is latent) and that each (observed) indicator used to quantify it only partially reflects the theoretical phenomenon. These indicators are indeed taken to be manifestations of various aspects of the unobserved phenomenon, and hence provide a good starting point for its measurement, but none of them is sufficient in itself to get a complete picture of it.

Factor analysis, MIMIC (multiple indicators and multiple causes) and structural equation models (SEM) all fall into this line of reasoning Table1. Though principal components (PC) is not a latent variable model, we have also included it in this section for two reasons. First it is widely used in empirical applications as an 'aggregating' technique and secondly the PC's can be shown to be equivalent to the factor scores under certain conditions (Krishnakumar, 2008).

The use of principal components (PC) or a combination of principal components is a popular technique in the construction of multidimensional indices. One of the earliest studies in this direction is Ram (1982) who first applies PC on three dimensions, namely life expectancy at age one, infant mortality and adult literacy, and combines it with per capita GDP, again using PC, to form a composite index. Slottje (1991) follows the same approach by selecting 20 attributes for 126 countries across the world, calculating a PC-based index and comparing it with indices obtained using hedonic weighting procedures. This method, which is essentially a data reduction technique, dates back to Hotelling (1933) in the statistical literature with a wide range of applications in numerous fields such as psychology, biology, anthropology and more recently in economics and finance. The basic idea behind this method is to determine orthogonal linear combinations of a set of observed indicators chosen in such a way as to reproduce the original variance as closely as possible. But this method lacks an underlying explanatory model which the factor analysis offers.

The FA model assumes that the observed variables (indicators) are different manifestations of one or more underlying unobservable variables called factors. The MIMIC model (cf. Joreskog and Goldberger, 1975) represents a step further in the theoretical explanation of the phenomenon under investigation as it is not only believed that the observed variables are manifestations of a latent concept (or a few latent concepts) but also that there are other exogenous variables that "cause" and influence the latent factor(s). SEM extends this structure by introducing simultaneity among the latent variables in the structural explanation[4] . This structure is highly relevant in our context as it provides us with a framework for operationalizing the capability approach, acknowledging its indirect measurement, and assessing poverty in the capability space. It also offers an explanatory framework of the causes of capability poverty and interactions among its dimensions, which is fundamental for understanding the phenomenon and for making policy decisions.

The use of non-statistical methods comprises scaling techniques and fuzzy sets theory. The scaling of functionings consists in a projection of each variable onto a 01 range, which are further aggregated into a composite measure. The Human Development Index and the Human Poverty Index (UNDP, 1997-2008) are the two major examples of the employment of such techniques, and are widely accepted as the first major operationalizations of the CA in the space of observed functionings. Following the classification provided in Table1, we could classify them as two-level indices belonging to the class of multiple dimensions-multiple indicators, with exogenous weights (at both levels). The dimensions, in this case, correspond to observed functionings and the indicators their corresponding measures. Fuzzy sets methodology is a mathematical tool used to provide a summary measure of the "degree" of poverty or well-being associated to the distribution of functionings under analysis. In this case, poverty is not a zero or one concept but rather "a broad and opaque" one (Sen, 1992). Cerioli and Zani (1990) and Cheli and Lemmi (1995) have applied it to the study of well-being measurement in general, while Chiappero-Martinetti (2000) and Lelli (2000) to functionings in particular.

3 Operationalization of the Capability Approach through Latent Variable Modelling

In this section we describe the main features characterizing the Capability approach and its operationalization through a structural equation model (SEM). Our main concern is the estimation of the latent variables or capability indices, in our case, which provide a measure of the well-being status of individuals in the different capability domains. These are presented at the end of the section.

According to Sen (1985), (1992), (1993), (1999), the basic purpose of development is to enlarge people's choices so that they can lead the life they want to. This notion of freedom is the essence of the Capability Approach (CA). Under this framework, the choices that one has are termed "capabilities" and the levels of achievement in these capability domains are called "functionings". Resources or entitlements (commodities and their characteristics) lack of an intrinsic value and are rather instrumental. In other words, functionings are the individual's "beings" and "doings" resulting from a given choice, capabilities are all the possible functionings that the individual can achieve, and resources are the means to achieve. The conjunction of these three notions (capabilities, functionings and resources) leads to a conversion process of resources to possible functionings which is individual-specific, influenced

[4]See e.g. Bollen (1989), Muthen (2002), and Skrondal and Rabe-Hesketh (2004) for a survey.

by personal, social and cultural characteristics. Thus, in this approach, human development is understood as the enhancement of the set of choices or capabilities of individuals, and poverty corresponds to a notion of deprivation defined as the individual's inability to achieve "minimal" functionings.

Operationalization of the CA, assessing either well-being or the lack of it, has been particularly challenging due to its informational and methodological requirements. Further, there is no common agreement about which dimensions ought to be included, nor how they should be summarized. To make it operational, composite indices have been constructed on the means of statistical and non-statistical methods. Yet, the operationalization has mainly concerned the aggregation of functionings and/or resources .

The employment of statistical methods in the area of human development and well-being measurement have essentially been confined to the use of principal components (Ram, 1982; Slottje, 1991; Klasen, 2000; Rahman, Mittelhammer and Wandschneider, 2005; Noorbaksh, 2003; McGillivray, 2005), and exploratory factor analysis (Schokkaert and Van ootehgem, 2000; Balestrino and Sciclone, 2000; Lelli, 2000; and Maasoumi and Nickelsburg, 1988). SEM have been used by very few authors (Kuklys, 2005; Krishnakumar and Ballon, 2008) and have not been applied so far, to our knowledge, in the field of capability deprivation measurement.

In this paper, we will concentrate on Structural Equation Models as they provide a complete framework to take into account the interactions among the different capability dimensions, and the influence of the surrounding environment [5]. In a SEM framework, the (not directly observable) degree of freedom in the different dimensions relevant for wellbeing and poverty assessment are represented by latent variables, partially observed through a set of indicators, and explained by a collection of exogenous variables. The estimators of the latent variables provide a measure of the individual's capability status in the different dimensions. On the basis of these latent variable scores we derive, in section 4, multidimensional capability deprivation indices.
The SEM is formalized as follows [6]:

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\mathbf{x}_i + \boldsymbol{\zeta}_i \tag{1}$$

where $\boldsymbol{\eta}_i$ is a $(m \times 1)$ vector representing the unobserved capability of individual i in each of the m domains; $\mathbf{x}_i$ is a $(k \times 1)$ vector of k exogenous factors representing the social, cultural and political environment; $\boldsymbol{\zeta}_i$ is a $(m \times 1)$ vector representing the unknown omitted factors in the explanation of $\boldsymbol{\eta}$ that are not explicitly modelled in the equation (random errors); $\boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\Gamma}$ are the corresponding coefficient vector and matrices.

The observed indicators in the different capability domains can either be continuous or qualitative variables. In order to be able to treat these two types in a uniform

[5] See Krishnakumar (2007) for further explanations regarding the adequacy of SEM in the operationalization of the CA.

[6] For a detailed formalization see Krishnakumar (2007) and Krishnakumar and Ballon (2008).

way we introduce a response variable y * i which will be taken to be directly observed in the case of a continuous indicator, and which will be latent and linked to the observed variable through a qualitative response model in the case of qualitative data. This gives the following measurement equations:

$$\mathbf{y}_i^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \mathbf{D}\mathbf{x}_i + \boldsymbol{\epsilon}_i\ , \tag{2}$$

where $\mathbf{y}_i^*$ is a $(p \times 1)$ vector representing the response variables of individual i; $\mathbf{x}_i$ is a $(s \times 1)$ vector of s individual characteristics and preferences that have an impact on the choice process transforming capabilities into functionings; $\boldsymbol{\epsilon}_i$ is a $(p \times 1)$ vector of random errors; $\boldsymbol{\nu}, \boldsymbol{\Lambda}, \mathbf{D}$ are the corresponding coefficient vector and matrices. In the case of a continuous observed indicator for individual i in dimension j denoted as y_{ij}, we have:

$$y_{ij} = y_{ij}^*, \tag{3}$$

when the observed indicator is of a qualitative nature, we write:

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > \tau_j \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

for a dichotomous indicator, and

$$y_{ij} = c, \quad \text{if} \quad \tau_{j,c} < y_{ij}^* \leq \tau_{j,c+1} \tag{5}$$

for a categorical indicator.

It is further assumed that:

$$E(\boldsymbol{\zeta}_i) = 0,\ E(\boldsymbol{\epsilon}_i) = 0, \tag{6}$$

$$V(\boldsymbol{\zeta}_i) = E(\boldsymbol{\zeta}_i \boldsymbol{\zeta}_i') = \Psi, \tag{7}$$

$$V(\boldsymbol{\epsilon}_i) = E(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i') = \Theta. \tag{8}$$

Thus, the observations are centered without loss of generality, and the disturbances across individuals are assumed to be homoscedastic and nonautocorrelated. These assumptions do not mean that the individual disturbances from two different equations need to be uncorrelated nor that they have the same variance. Equations (7) and (8) show that these are full matrices allowing for correlations between different capability domains and for heteroscedastic variances.

On the basis of these stochastic assumptions, the above nonlinear model is estimated by minimizing the distance between the sample moments of the observed variables and the corresponding theoretical moments expressed as a function of the unknown parameters, by generalised method of moments (GMM)(see e.g. Browne, 1984). An alternative estimation method is (conditional) maximum likelihood (cf. e.g. Joreskog, 1973; Browne and Arminger, 1995; Muthen, 1984). In this case the parameters are estimated under (conditional) normality of the indicator vector y * given the exogenous variables x and its variance is corrected using the well-known 'sandwich' formula under non-normality (quasi-maximum likelihood, cf. White, 1982; Gouriéroux, Monfort and Trognon, 1984).

Factor scores

Once the parameter estimates are obtained, the final step consists in the estimation of the vector of latent variables for each individual i (factor scores), which is of primary interest to us, because these estimators quantify the degree of freedom in each capability domain. Factor scores could be estimated by the Empirical Bayes estimator or by maximizing the logarithm of their posterior distribution. Both methods lead to similar results. Following the empirical Bayesian approach, which is a standard procedure suggested in the related literature (cf. Skrondal and Rabe-Hesketh, 2004), the latent factors are estimated by their posterior means given the sample, replacing the parameter values by their estimates. In the case of a linear SEM model, with only continuous indicators, the empirical Bayes estimate of the factor scores is[7]:

$$\begin{aligned}\hat{\boldsymbol{\eta}}_i = & \left\{[\mathbf{I} - \Sigma \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \Sigma \boldsymbol{\Lambda}' + \Theta)^{-1} \boldsymbol{\Lambda}]\mathbf{A}^{-1}\boldsymbol{\alpha} - [\Sigma \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \Sigma \boldsymbol{\Lambda}' + \Theta)^{-1} \boldsymbol{\nu}]\right\} \\ & + \left\{\Sigma \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \Sigma \boldsymbol{\Lambda}' + \Theta)^{-1} \mathbf{y}_i^*\right\} \\ & + \left\{[\mathbf{I} - \Sigma \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \Sigma \boldsymbol{\Lambda}' + \Theta)^{-1} \boldsymbol{\Lambda}]\mathbf{A}^{-1}\boldsymbol{\Gamma}\mathbf{x}_i - [\Sigma \boldsymbol{\Lambda}' (\boldsymbol{\Lambda} \Sigma \boldsymbol{\Lambda}' + \Theta)^{-1} \mathbf{D}\mathbf{x}_i]\right\} \end{aligned} \quad (9)$$

where $\mathbf{A} = \mathbf{I} - \mathbf{B}$, and $\boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\ \boldsymbol{\Psi}\ (\mathbf{I} - \mathbf{B})'^{-1}$. Equation (9) shows that the factor score results from a combination of three terms: a '*net* constant' term, summarizing the intercept effect $\boldsymbol{\alpha}$ and $\boldsymbol{\nu}$, an 'indicator' term reflecting the information contained in $\mathbf{y}_i^*$, and a '*net* causal' term resuming the causal effect of the exogenous variables $\mathbf{x}_i$ of the measurement and structural equations. It is interesting to note that if we write it as,

$$\hat{\boldsymbol{\eta}}_i = \mathbf{K} + \mathbf{W}\mathbf{y}_i^* + \tilde{\mathbf{W}}_p \mathbf{x}_i \quad (10)$$

where $\mathbf{K}, \mathbf{W}$, and $\tilde{\mathbf{W}}$ are matrices of appropriate dimensions, we see that pre-multiplying matrices of each term in equation (9) are the associated weights $\mathbf{W}$ and $\tilde{\mathbf{W}}$ of the vector of indicators $\mathbf{y}_i^*$, and of the vector of (net) exogenous causes $\mathbf{x}_i$, respectively, and $\mathbf{K}$ a constant.The above formula shows that, the capability set for a single individual

[7] See Krishnakumar and Nagar (2008) for details regarding the derivation.

is represented by a composite measure that summarizes the information provided by her achievement indicators and her personal, social and cultural environment features, in an endogenous manner.

When the observed indicators are both continuous and qualitative the derivation of factor scores becomes more complex and is carried out by iterative techniques (Muthen, 1998-2004). However, it is interesting to note that only first and second order (conditional) moments are necessary for obtaining factor scores. This is shown by equation (11), which is equivalent to equation (9) except that it is written in moment terms, the expression is then:

$$\hat{\boldsymbol{\eta}}_i = \mu_i + \Sigma \boldsymbol{\Lambda}' \Sigma^{*-1} \left(\mathbf{y}_i^* - \mu_i^*\right) \tag{11}$$

where

$$\boldsymbol{\mu}_i = E(\boldsymbol{\eta}_i|\mathbf{x}_i) = (\mathbf{I}-\mathbf{B})^{-\mathbf{1}}\boldsymbol{\alpha} + (\mathbf{I}-\mathbf{B})^{-\mathbf{1}}\boldsymbol{\Gamma}\mathbf{x}_i\,, \tag{12}$$

$$\boldsymbol{\Sigma} = V(\boldsymbol{\eta}_i|\mathbf{x}_i) = (\mathbf{I}-\mathbf{B})^{-\mathbf{1}}\boldsymbol{\Psi}\,(\mathbf{I}-\mathbf{B})^{\prime-\mathbf{1}}\,, \tag{13}$$

are the mean vector and covariance matrix of the (multivariate normal) prior distribution of $\boldsymbol{\eta}_i$ conditional on $\mathbf{x}_i$, and

$$\boldsymbol{\mu}_i^* = E(\mathbf{y}_i^*|\mathbf{x}_i) = \boldsymbol{\nu} + \boldsymbol{\Lambda}\,\boldsymbol{\mu}_i + \mathbf{D}\,\mathbf{x}_i\,, \tag{14}$$

$$\boldsymbol{\Sigma}^* = V(\mathbf{y}_i^*|\mathbf{x}_i) = \boldsymbol{\Lambda}\,\boldsymbol{\Sigma}\,\boldsymbol{\Lambda}' + \Theta\,, \tag{15}$$

are the conditional expectation and conditional variance of the (latent) response variable $\mathbf{y}_i^*$, that links a qualitative indicator through a latent response model. Thus from (11), we see that to estimate the latent variable vector (capability indices) one only needs information on conditional means $\boldsymbol{\mu}_i$, $\boldsymbol{\mu}_i^*$, and conditional variances $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}^*$.

4 Deprivation in the capability space

Formalization

This section formalizes the assessment of deprivation in the capability space using the capability/freedom measures derived in the previous section. After some notational definitions, we present two axioms that characterize the evaluative space of capabilities as measured by factor scores in our model. On this basis we propose two types of capability deprivation indices (CDI). One that performs aggregation within a capability domain (single dimension-multiple indicators), and a second that, in addition, combines capability sets across dimensions (multiple dimensions-multiple indicators). Both types of indices are ordinal and satisfy a series of axioms representing ethical principals and value judgements.

Consider a population of $i = 1, 2, \ldots, n$ individuals whose degree of freedom (ability to choose) in any capability domain $j = 1, 2, \ldots, m$ is represented by a continuous variable η_{ij} (say the true factor score). We will state the definitions and axioms for true measures of capabilities η_{ij} which have to be replaced in practice by their estimates $\hat{\eta}_{ij}$ from the previous section.

Let $\eta_j = (\eta_{1j}, \eta_{2j}, \ldots, \eta_{nj})$ denote the j-th capability vector drawn from the j-th capability space $E_j = \bigcup_{n=1}^{\infty} E_j^n$, where $\eta_{ij} \in E_j$ is some nondegenerate real interval and E_j^n is the set of all n-tuples of elements in E_j. Further, assume that the individual capabilities are arranged in an ascending order such that $\eta_{1j} \leq \eta_{2j} \leq \cdots \leq \eta_{nj}$. Next, let $R(.)$ denote the rank function that maps the η_j vector into the set of positive integers. This is, $R(.) : \eta_j \mapsto \mathbb{Z}_+ = (1, 2, \ldots, n)$. For a given individual $i \in n$, his ordinal position in the distribution of capability scores is given by $R(\eta_{ij}) = r_{ij}$, where $r_{ij} = p, p \in \mathbb{Z}_+ = (1, 2, \ldots, n)$. Let $r_j = (r_{1j}, r_{2j}, \ldots, r_{nj})$ be the resulting vector of increasingly arranged individual ranks in the j-th dimension, with $\texttt{max}(r_j) \leq n$ and $\texttt{min}(r_j) = 1$. For our purposes we set $R(\eta_{1j}) = r_{1j} = 1$. This means that we assign a rank of one to the worst off individual i.e., the one exhibiting the lowest score. Lastly, for $i \neq i'$ both $\in n$, $r_{ij} = r_{i'j}$ whenever $\eta_{ij} = \eta_{i'j}$, or in other words, two different individuals exhibiting the same factor score are ranked equally. Note that the formalization is written down in terms of the true factor scores $\boldsymbol{\eta}$ but in practice these will be replaced by their estimates obtained as explained in the previous section.

We characterize the capability space by the following two axioms:

Axiom 1 (Monotonic freedom of choice) *For a given capability domain j and for any two individuals i and $i' \in \mathbb{Z}_+$, i is better off than $i' \Leftrightarrow \eta_{ij} > \eta_{i'j}$.*

Axiom 2 (Ordinal freedom of choice) *For a given capability domain j and for any two individuals i and $i' \in \mathbb{Z}_+$,*

$$i \textit{ is better off than } i' \Leftrightarrow r_{ij} > r_{i'j}$$

$$i \textit{ is as well as than } i' \Leftrightarrow r_{ij} = r_{i'j}$$

where r_{ij} and $r_{i'j}$ denote the ordinal positions of i and i' in the distribution of capability scores, respectively.

Thus the degree of freedom of an individual is represented by her factor score, and her relative position in the whole population by her rank in the ordering of the (capability) scores. Individual deprivation and interpersonal comparisons can therefore be based on a mixture of cardinal and ordinal information regarding the scores, with greater (score) values and higher ranks denoting greater well-being.

On the basis of Axioms 1 and 2, and using an absolute notion of capability deprivation (Sen, 1983), we propose the following definition for identifying the deprived individuals in a single capability domain.
Let η_{dj} and r_{dj} denote the factor score and ordinal position of a '*fictitious*' deprived individual who only has the capability to achieve a minimal set of functionings. Let us denote this 'minimal' capability as $\eta_{dj} \in E_j$.
Weak definition of the deprived: For a given deprivation threshold $\eta_{dj} \in E_j$ and for all $\eta_j \in E_j$, the j-th capability deprived domain is $S_j(\eta_j) = \{i \in E_j \mid \eta_{ij} < \eta_{dj} \Rightarrow r_{ij} < r_{dj}\}$, and the number of deprived individuals is $q_j(\eta_j) = card\{S_j\}$.
The preceding definition takes the deprivation threshold to be the score (and rank) of a *"fictitious"* individual, denoted by the subscript(d). The term fictitious is used to recall that it is the analyst who decides what the minimal level of functionings that an individual should be able to achieve. According to this definition the set of deprived individuals comprises all individuals whose factor score (and thus rank) is smaller than the factor score (and rank) of the fictitious deprived individual. It is worth noting that using a *weak* definition of the deprived avoids including the fictitious deprived individual in the deprivation set, which by 'construction' is the deprivation threshold.

Deprivation within a capability set

The above definition of the poor and axioms (1) and (2) allow us to characterize individual deprivation, in a given capability domain, as an individual function:

$$d_{ij} = \begin{cases} f(\eta_{ij}, r_{ij}; \eta_{dj}, r_{dj}) & \text{if } i \in S_j \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

where f is a continuous function decreasing with respect to η_{ij} and r_{ij}, and increasing with respect to η_{dj} and r_{dj}. Note that f is not differentiable. Thus, from (16) we see that the individual deprivation function d_{ij} depends on cardinal (score) and ordinal (rank) information regarding the individual and the fictitious deprived status.
The total deprivation within a capability domain for the whole population can be defined as follows:

For a given capability domain j and a deprivation threshold η_{dj} a measure of aggregate capability deprivation, say a capability deprivation index, is a real-valued function:

$$D_j = G\big(\eta_j, r_j, j = 1, ..., n; \eta_{dj}, r_{dj}\big) = G\big(d_{1j}, d_{2j}, \ldots, d_{nj}\big) : E_j \times \mathbb{Z}_+ \to \mathbb{R}_+ \tag{17}$$

where the function G has similar properties as f with respect to its arguments. Clearly there are an infinite number of f and G functions that satisfy the above conditions. In what follows, we propose some additional properties that the individual deprivation function $f(\cdot)$ and aggregate function $G(\cdot)$ should satisfy. These properties represent value judgements and ethical principles in the capability space leading to restrictions on possible functional forms. Under these principles, the resulting measure D_j would be a distribution-sensitive measure allowing for interpersonal comparisons. We shall call it a simple capability deprivation index. The term simple is used to mean the analysis of deprivation in a single capability domain with multiple functioning indicators.

Desirable properties in the capability framework

Inspiring from the literature on economic poverty measurement we propose the following desirable properties for our CDI. These could be regarded as "basic" value judgements that one would like to incorporate in any quantitative measure to be used as a capability deprivation index.

Definitions

Change in non deprived scores: We say that $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a change in non deprived scores if $\tilde{\eta}_{ij} = \eta_{ij} \forall i \in S_j$, and $\tilde{\eta}_{ij} \neq \eta_{ij}$ for some $i \notin S_j$. By Axiom 2, this will lead to a change in the corresponding rank vector[8], where $\tilde{r}_{ij}$ will be different for all $i \notin S_j$ whose ranks were above the rank of the individual whose score has changed in the initial distribution.

Gain (loss) of freedom of choice[9]: We say that $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a gain (loss) of freedom of choice if $\tilde{\eta}_{ij} > \eta_{ij}$ ($\tilde{\eta}_{ij} < \eta_{ij}$ for some $i \in S_j$, and $\tilde{\eta}_{ij} = \eta_{ij}$ for every other $i \in S_j$. By Axiom 2, it will translate into a better (worse) ordinal position for the individual whose score has changed[10] and the ranks will change for all individuals whose ranks were above the rank of the individual whose score has changed in the initial distribution.

Regressive (progressive) transfer: We say that $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a regressive (progressive) transfer if there exists a pair of individuals i and i', such that: (1) $\eta_{ij} < \eta_{i'j}$, (2) $\tilde{\eta}_{i'j} - \eta_{i'j} > 0$; $\eta_{ij} - \tilde{\eta}_{ij} > 0$, (3) $\eta_{kj} = \tilde{\eta}_{kj} \ \forall k \neq i, i'$. Equivalently, by Axiom 2 we could express these three conditions in terms of their ordinal representations, say (1) $r_{ij} < r_{i'j}$ (2) $\tilde{r}_{i'j} - r_{i'j} > 0$; $r_{ij} - \tilde{r}_{ij} > 0$

The first condition says that i is more deprived than i'. The second condition says that there is an increase in the freedom of choice of the less deprived individual (i') and a decrease in the freedom of choice of the more deprived one (i). The third condition says that the freedom of choice of the remaining individuals does not change. Regarding the ranks, it is clear that the rank of i' will increase and that of i will decrease but it is also easy to understand all the other ranks of individuals who were

[8]When estimates are used, the change should be "statistically significant" for the ranks to be different.

[9]This definition is equivalent to the simple increment (decrement) definition used by Zheng (1997).

[10]Once again the gain(loss) of freedom should be "statistically significant" in case of estimates.

above the initial rank of i will be different in the new distribution.
Now the properties:

Single-dimension Focus axiom. If $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a change in non deprived scores then $D_j = \tilde{D}_j$. This means that the the deprivation measure depends only on the scores of the deprived individuals.

Single-dimension Monotonicity axiom. If $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a gain (loss) of freedom of choice, then $D_j \leq \tilde{D}_j (D_j \geq \tilde{D}_j)$. This means that for a given deprivation threshold, the degree of deprivation resulting from an improvement (worsening) of an individual's score cannot be greater (smaller).

Single-dimension Transfer axiom. If $\tilde{\eta}_j \in E_j$ is obtained from $\eta_j \in E_j$ by a regressive transfer among the deprived, then $D_j \geq \tilde{D}_j$. This means that whenever the scores of two deprived individuals change, with the more deprived one ending with even less ability to choose, the degree of deprivation should not decrease.

Relying on these properties, and recalling the lack of cardinal interpretation of the scores, we propose the following characterizations of the individual deprivation function in terms of the number of positions that an individual is away from the threshold position (rank gap):

$$d_{ij} = (r_{dj} - r_{ij}) \tag{18}$$

The overall capability deprivation index (for a single dimension) can then be defined as:

$$\begin{aligned} D_j &= \frac{1}{n} \sum_{i \in S_j} (d_{ij}) \cdot \omega_{ij} \\ &= \frac{1}{n} \sum_{i \in S_j} (r_{dj} - r_{ij}) \cdot \omega_{ij} \end{aligned} \tag{19}$$

with $\omega_{ij} = \frac{\eta_{dj}}{\eta_{ij}}$ i.e. the rank gap is multiplied a 'weight' given by the inverse of the relative distance of the individual score to the deprivation threshold. In other words, the farther away the score of the individual from the threshold in relative terms (i.e. the smaller the relative score), the bigger the weight, thus giving greater importance to the more deprived.

Alternatively one could also define

$$d_{ij} = (r_{dj} - r_{ij}) \cdot \omega_{ij} \tag{20}$$

incorporating the 'weight' directly into the individual deprivation measure and have

$$D_j = \frac{1}{n}\sum_{i \in S_j}(d_{ij}) \tag{21}$$

which will yield the same aggregate index.

It is easily seen that both ((18)) and ((20)) imply that d_{ij} is decreasing with respect to r_{ij} and increasing with respect to r_{dj} meaning greater deprivation whenever the rank gap increases. If we include the weights, then in addition f becomes decreasing with respect to η_{ij} and increasing with respect to η_{dj} thus satisfying the conditions laid out earlier.

Regarding the aggregate deprivation within a capability domain, the functional form of D_j in (21) shows that the capability deprivation index is a sort of an average of the individual deprivation functions. We say "sort of" because the individual weights used in its computation do not add up to 1, as is required for weighted averages. This however is not necessarily a drawback of the index and is true for other indices like the classical FGT class poverty indices. Our index gives the average number of 'weighted' positions that an individual is away from the deprivation threshold. This per capita interpretation seems quite reasonable as it allows to compare different groups without any additional assumptions on the scale of measurement of the scores. Indeed, any attempt of summarizing the information using the values of the scores as such, and not using them as weights as in our case, needs at least an interval scale of measurement for the scores, under which only means and differences are allowed. If one is interested in performing other operations with scores, then a ratio scale is needed (Stevens, 1946). For the moment, we do not assume that our scores satisfy the above scales of measurement and only consider them as ordinal (continuous) measures.

The above CDI satisfies the properties enumerated in the beginning of this section and is thus sensitive to the distribution of scores among the deprived (see Appendix A for the proofs).

Deprivation across capability domains

Finally, if one is interested in summarizing the deprivation status of individuals in several dimensions then, one need to aggregate deprivation measures across capability domains. One can also envisage a latent variable approach for this purpose considering the capability measures (factor scores) in each dimension as imperfect measurements of the overall well-being, taken to be a (single) latent variable. This will amount to performing a factor analysis using the (estimated) scores in each dimension and will give us an estimate of the overall welfare (latent factor score) along with the (endogenous) weights associated with the individual dimension scores. However, this may not a suitable method for combining information on different dimensions as it does not enable the policy maker or the researcher to include 'normative judgements' on the relative importance of different dimensions based on social and ethical principles. Therefore we turn to non model-based methods for carrying out aggregation across dimensions (to be applied after using model based methods for combining indicators within a dimension).

Studies by Tsui (2002), and Bourguignon and Chakravarty (2003) who have proposed multidimensional poverty measures resulting from axiomatic criteria are particularly relevant in this context[11]. The Bourguignon and Chakravarty measures are quite appealing in our case as their individual deprivation functions in a multidimensional setting directly correspond to our (single dimension) deprivation indicator based on our factor scores. However there is a fundamental difference between their individual deprivation indicators and ours in that our single-dimension deprivation function is already a multi-indicator function (i.e. we have one more level of aggregation within each dimension). They propose many functional forms in their work, and we have chosen some of them for their interesting interpretations.
Thus we define our multi-dimensional deprivation index as follows, restricting our formalization to the bi-dimensional case, as we are only concerned with two dimensions in our application. Generalisation to more than two dimensions is not difficult.

$$\rho(\eta_i,\eta_d)=\begin{cases} 1 \quad \text{if} \quad \exists \quad j \,\epsilon\, \{1,2,\ldots,m\} : \eta_{ij} < \eta_{dj} \\ 0 \end{cases} \tag{22}$$

According to (22) an individual is considered as deprived if her score is below the deprivation threshold in at least one dimension. In terms of the poverty literature language, this definition corresponds to the 'union' approach. From an ethical point of view, this seems more appropriate as no deprivation according to this index is only possible if there are no deprived in any dimension. One could also consider an 'intersection' approach where an individual is identified as deprived if she is unable to achieve a minimal level of functionings in *all* the dimensions considered. On the basis of the above definition of the deprived, we define the individual bidimensional deprivation function :

$$\begin{aligned} \delta(\eta_1,\eta_2;\eta_d) &= \delta(\eta_{i1},\eta_{i2}, i=1,...,n;\eta_{d1},\eta_{d2}) \\ &= I\Big\{\mathtt{Max}[(r_{d1}-r_{i1})\omega_{i1}\,;0]\,;\mathtt{Max}[(r_{d2}-r_{i2})\omega_{i2}\,;0]\Big\} \end{aligned} \tag{23}$$

where $\omega_{ij}=\frac{\eta_{dj}}{\eta_{ij}}$ is the relative individual weight as before, and $I(u_1,u_2)$ is an increasing, continuous (but not differentiable), and quasi-concave function with $I(0,0)=0$. Note that arguments in (23) are the individual deprivation functions in the single dimensional case. Bourguignon and Chakravarty propose a CES functional form:

$$I=\Big\{a\,\mathtt{Max}[(r_{d1}-r_{i1})\omega_{i1}\,;0]^{\theta}+(1-a)\mathtt{Max}[(r_{d2}-r_{i2})\omega_{i2}\,;0]^{\theta}\Big\}^{\alpha/\theta} \tag{24}$$

where $a \,\epsilon\, [0,1]$ is the dimensional weight, $\alpha=0,1,2$ denotes the aversion to capability deprivation, and $\theta \geq 1$ is a parameter that accounts for the degree of substitution

[11] Information theory based multidimensional measures have also been proposed by Maasoumi and Lugo (2008); however it will take us outside the scope of the paper to discuss them here.

between dimensions. If $\theta = 1$ we face perfect substitution, whereas if $\theta \to \infty$ we have no substitutability (Leontief). The CES functional form has an additional parameter α capturing the aversion to capability deprivation. Aggregating over the whole population we get the bi-dimensional CDI:

$$BD = \frac{1}{n} \sum_{i \epsilon \text{deprived}} \left\{ a \, \texttt{Max}[(r_{d1} - r_{i1})\omega_{i1} \, ; 0]^{\theta} + (1-a) \, \texttt{Max}[(r_{d2} - r_{i2})\omega_{i2} \, ; 0]^{\theta} \right\}^{\alpha/\theta} \quad (25)$$

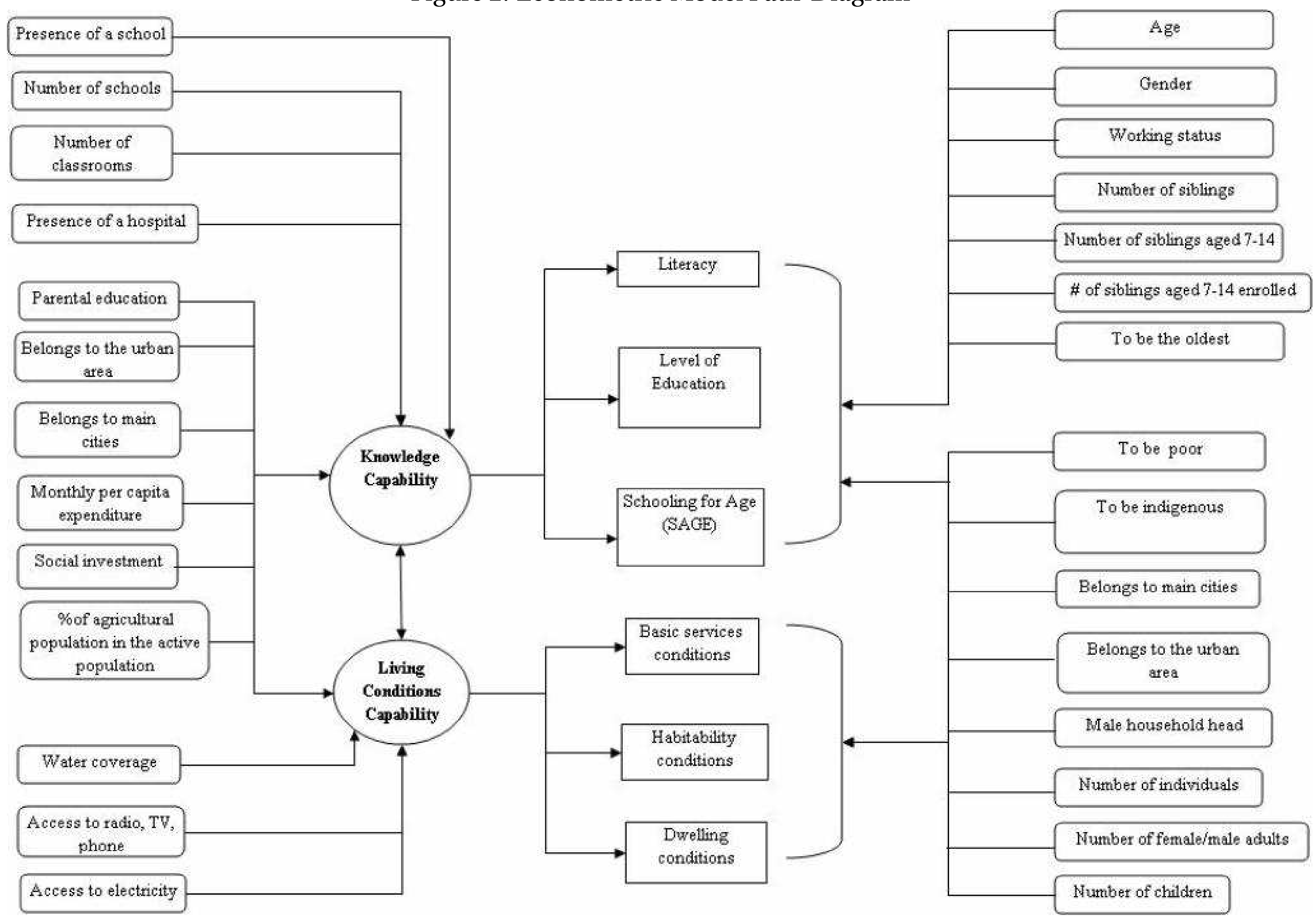
Bourguignon and Chakravarty show that this index satisfies strong focus, multidimensional transfer principle (MTP), and non-decreasing correlation increasing switch (NDCIS). The strong focus axiom requires the bi-dimensional deprivation measure to be independent from non-deprived scores. The MTP is a generalization of the single-dimensional transfer principle previously mentioned to the case of several dimensions. The NDCIS deals with the degree of substitution or complementarity of capability deprivation across dimensions.

5 Empirical Results

In this section we illustrate the use of the preceding measures to analyze the deprivation situation of Bolivia's children in the capability domains of knowledge and living conditions. The deprivation measures are applied to the estimated factor scores of the latent variable model proposed by Krishnakumar and Ballon (2008). The data used in this model corresponds to the 2002 MECOVI program, a National Household Survey conducted by the National Statistical Institute of Bolivia, with the support of the World Bank. The 2002 survey covers 5,952 households and 24,933 individuals and contains information at the national and regional levels, on education, health, migration, labor, income, household characteristics, and living conditions. The information of the MECOVI 2002 survey is complemented with information from the National Institute of Statistics (INE) on social investment and school conditions at the municipal level for data on some exogenous variables of our model. Our sample comprises 5313 enrolled primary school children aged 7-14. Figure 2 presents path-diagram of the econometric model used to operationalize the capability theory to the Bolivian case.

As shown by the path-diagram, knowledge and living conditions capabilities, represented by circles, are measured by three functioning indicators in each case. The indicators for educational achievements include literacy, level of education, and schooling for age (SAGE). The SAGE variable reflects the lag in a child's schooling with reference to a 'normal' achievement rate (see, Psacharopoulos and Yand, 1991). A score under 1 is considered as being below normal progress in the school system because of late entry or dropping out and/or re-enrollment and the further away it is below 1 the lower the performance of the child. Living conditions outcomes are measured by the quality of basic services, and the quality of dwelling and habitability conditions enjoyed by the household. These three indicators are measured by an ordered categorical variable with three categories indicating low, middle, and high quality. The exogenous variables included in the structural and measurement equations (1) and (2), respectively are split into supply and demand factors. These are classified according to their influential role in the enhancement of capabilities, and in the choice process. Thus, supply variables (number of schools, number of classrooms, parental education, etc) are included in the structural equations, and demand variables (age, gender, indigenous status, etc.) in the measurement equations. In the path-diagram these are represented on the left and right hand side, respectively.

Figure 2: Econometric Model Path-Diagram



Tables 9 and 10 of the appendix present the estimated parameters of the econometric model described by the path-diagram. We will not discuss them in detail here, as our main concern is the deprivation situation. The reader is referred to Krishnakumar and Ballon (2008) for a detailed interpretation of the results. Briefly speaking, we see that there is a positive simultaneous relationship between knowledge and living conditions capabilities implying that they mutually enhance each other. The structural equation results also show that parental education, and supply variables of health and education access, along with, urban location have a positive influence in the enhancement of both capability sets. Regarding the measurement equations, we observe a positive loading for each of the capability domains on its corresponding indicators. Among the exogenous variables included in these equations we see that being indigenous or poor has a negative effect on the achieved functionings in both dimensions given the same capability level. The sibling structure and working status have a mixed and negative effects on educational achievements, respectively. Urban environment seems to favour better housing conditions.

Computing the CDI's

On the basis of the estimated parameters and the estimated factor scores (obtained as explained in section 3), we compute the capability deprivation indices presented in section 4. We only look at 14-year old children as it is only or this age group that one can talk of not being able to achieve a minimal functioning level. Our analysis is carried out for 14-year-old children belonging to one of the following groups: female/rural, female/urban, male/rural and male/urban. We only consider 14 year-old children because it is only at this age that one can define an absolute criteria for identifying the capability deprived in the knowledge domain as they are expected to have finished primary education at this age. For these children, we fix the minimal level of (threshold) educational functionings corresponding to the 'fictitious' child at the threshold deprivation level as being literate (if we fix it as illiterate, we end up with too few deprived)[12], a SAGE (lag in schooling progress) value of 0.5 and with incomplete primary education. The living conditions threshold is represented by the same fictitious child living in a house with a 'low' quality of basic services, habitability, and dwelling conditions. Along with these characteristics of the indicators, we also need to fix the exogenous variables in our model for calculating the threshold deprivation level. As these exogenous factors influence the choice process and the capability set, we propose to fix them at their minimum or maximum observed value depending on whether the associated coefficient is positive or negative (Tables 7 and 8 of the appendix). Thus the fictitious deprived child (representing the threshold of deprivation) will be surrounded by the most unfavorable environment. Table 2 presents the conditions characterizing the deprivation threshold.

Once the model is estimated, we compute the scores (representing the capabilities) for all individuals in each dimension. In order to calculate the ranks of the in-

[12]The issue of where to fix the minimum level of achievement for the threshold is a crucial one as the results are obviously sensitive to the value chosen; we will say more about this problem while examining the results.

Table 2: Deprivation thresholds

Exogenous variable	Sign of the estimated coefficient	Selected value	
Father's level of education	+	Min	No education
Mother's level of education	+	Min	No education
Monthly per capita expenditure	+	Min	25 (Bs)
Number of siblings	-	Max	6
Number of siblings aged 7 to 14	-	Max	6
Number of siblings aged 7 to 14 enrolled	+	Min	0
Male household head	-	Max	Yes
Number of children	-	Max	6
Use of medical services	+	Min	No
Number of schools	+	Min	1
Number of classrooms	+	Min	1
Working status	-	Max	Yes
Indicators			
Literate	+		Yes
Sage	+		0.5
Level of education	+		Incomplete primary
Quality of basic services	+		Low
Quality of habitability conditions	+		Low
Quality of dwelling conditions	+		Low

dividuals in each of the domains, one needs to perform an additional test of the difference between any two scores being significant or not, to decide whether the two individuals should be given the same rank or two different ranks. In formal terms, the null hypothesis that we are interested in testing between any two individuals i and k is:

$$H_0 : \eta_{ij} - \eta_{kj} = 0 \tag{26}$$

To make our rank derivations 'robust' to the estimated values of the scores, we performed a nonparametric bootstrap simulation and obtained the score vectors for 100 bootstrap samples drawn from the full sample of children aged 7 to 14. With each of these 100 samples, we re-estimated the above econometric model and computed the scores. Thus we have 100 score values per individual i in each dimension j, $\hat{\eta}_{ij,s}, s = 1, 2, \ldots, 100$ whose average was used as the capability measure of the individual in that particular domain. We have, by the Central Limit Theorem,

$$\bar{\hat{\eta}}_{ij} \overset{a}{\sim} N\left[E(\hat{\eta}_{ij}); \frac{V(\hat{\eta}_{ij})}{100}\right] \tag{27}$$

Under H_0 the corresponding statistic, given by the estimated value of the expression in (26) is asymptotically normal, with mean zero and variance equal to twice the variance given in (27) as we assume that individuals are independent and the moments of the distribution of latent factors are invariant over individuals. If H_0 is rejected, i and k will be assigned two different ranks.

Given that the factor scores are continuous latent variables, and because of their implications on the well-being status of the individuals, our ranking algorithm based on the above hypothesis testing procedure must be reluctant to accept the equality of two factor scores leading to identical rankings. That means that in our case, the Type II error is more important than the Type I error, as assigning equal ranks between two individuals implies that their well-being status is the same. Thus we decided to minimize the Type II error by reducing the null hypothesis region of acceptance to a 5% interval.

The following tables report our capability deprivation indices for the four groups. Panel A of Table 3 describes the within deprivation indices in the knowledge domain. Comparing urban and rural groups, we see that the head count ratio (i.e. the proportion of deprived children) is higher among the rural population, with a bigger difference among males compared to females. This means that 67% of rural females, 73% of rural males, 47% of urban males, and 49% of urban females do not enjoy the ability to achieve minimal functionings in knowledge capability. Regarding the the intensity of deprivation given by our per capita deprivation (rank) gap (CDI), once again, we see that it is higher in rural areas for both males and females, with a greater gap difference among males (16 per capita positions compared to 9 in the female case). We can therefore say that living in urban areas provides a bigger choice range and more so for male children than the female ones.

In what follows we will be looking at some descriptive statistics regarding the achievement indicators and exogenous variables of the deprived children in each of these groups (see Tables 4, 5 and 6). Perhaps it is necessary to explain why it is useful to look at these statistics. As mentioned earlier, the capability deprivation threshold was calculated by fixing the achievements at at a 'minimal' level and the exogenous factors at their most unfavourable state. The resulting deprivation threshold is an endogenous combination of all this information coming out of the model. Then we determine the 'deprived' children as those being below the threshold thus derived. But this does not necessarily mean that they will all be below the minimal/least favourable level in all the functionings and exogenous variables. That is why we compute these descriptives to see which indicators/exogenous variables seem to influence deprivation the most.

Looking at Table 4 presenting the distribution of 'knowledge' deprived children among the three achievement indicators, we see that level of education is the most constraining component with 98% of children exhibiting incomplete primary education, in rural regions, and 96% in urban areas. [13] It is interesting to note that the average value of SAGE for 'knowledge' deprived children is between 0.6 and 0.7 units

[13]The Literacy column is not very informative as they are all literate, which is logical according to the minimum we fixed for this indicator.

i.e. above the level corresponding to the fictitious deprived child (0.5) though not too far from it.

From Table 5 we find that parental education is low among ‘knowledge’ deprived children, with high proportion of parents exhibiting incomplete primary (even though the threshold taken was ’no education’ there are more in ’incomplete primary than in ’no education’ except for the female urban group). This situation worsens in the case of female children. Regarding other exogenous variables (see Table 6) we find that the use of medical services is also low in all the four groups with values ranging between 10 to 16% among rural and urban females, respectively, and 11% and 12% among their peer males. Finally, we see that working status is higher among the rural groups. Though, the deprived children indeed face a disadvantageous milieu, it is less restrictive compared to that of our fictitious child. The above analysis is also useful to identify the aspects that need the most attention for improving the capability status of children in the knowledge domain.

Turning to living conditions capability, Panel B of Table 3, we note that the percentage of deprived individuals and the intensity of their deprivation are higher among rural groups in this case too. However, the differences between females and males by region are less important for both the head count and the CDI. The head count ratio is between 2 to 6% in the female group and between 3 to 6% in the male one. The per capita deprivation gap is around 1 position in each of the four groups. Thus it seems that the situation is much better for living conditions than for knowledge capability. However, the small values of these measures need to be interpreted with caution as the threshold positions in both dimensions are not strictly comparable. Recall that the fictitious deprived child in this case is one exhibiting low quality in all the indicators relating to basic services, habitability and dwelling conditions, which really depict the bare minimum and rather desperate conditions of living.

Our explanation for the low index values here is that the threshold itself being so low the percentage of population below this level and the intensity of deprivation are very small. We come back to our earlier remark about the sensitivity of the index to the choice of the threshold level which is clear in this case. The values would be really different (and much bigger) if we had chosen ’medium’ quality for the minimum achievement level. It is our intention to explore this issue further in future work.

Our explanation is supported by the comparison between the two sets of deprived (’knowledge’ deprived and ’living conditions’ deprived) in terms of descriptive statistics relating to exogenous variables (see Tabes 5 and 6). All the variables such as parental education level, average monthly per capita expenditure, use of medical services are much lower for the ’living conditions’ deprived.

Deprivation in the bi-dimensional space is reported in Table 7 for the male groups, and in Table 8 for the female groups. The first column of both tables, labelled ’a’, is the weight assigned to the living conditions dimension. The substitution parameter (θ) ranges from perfect substitution (one) to no substitution (Leontief). The alpha values correspond to the head count ratio (union approach, $\alpha = 0$), the bi-dimensional deprivation intensity per capita gap ($\alpha = 1$), and the bi-dimensional deprivation severity per capita gap ($\alpha = 2$). Comparing urban and rural male children (Table 7) we observe that for the three deprivation measures, the deprivation status of rural males is higher than that of urban males. The head count ratio is higher by 21%, and the per capita intensity measure is between 12 to 16 positions higher. Looking at the weights, we see that as the weight attached to living conditions increases (from 0.3 to 0.8) the intensity and severity measures also increase. In addition, we observe that as the degree of substitution decreases the intensity and the severity measures increase for all weights. This is in accordance with the notions of substitution and complementarity underlying the CES functional form. As θ increases it becomes more difficult to compensate one dimensional deprivation with another. This is related to the 'nature' of the dimensions: whenever they reflect very different aspects of life, it seems reasonable to assume low degree of substitution. In our analysis we have considered a range of possibilities (one to Leontief) although, we think that no substitution (Leontief) is more reasonable to assume in the case of living conditions and education. The results for female groups (Table 8) show a similar pattern regarding the proportion of deprived, and the bi-dimensional intensity and severity measures, for all weights. However, the differences between rural and urban female groups are much smaller compared to males. This is also true for the no substitution case. We see that these measures increase as the degree of substitution decreases. As in the single dimensional case we evidence that living in urban areas has an enhancing effect on the choice set of male children.

Table 3: Education CDI (panel A) and Living Conditions CDI (panel B)

Panel A	Female		Male	
	Urban	Rural	Urban	Rural
Group size (Rmax)	178	129	195	125
Deprivation Threshold (rank)	88	83	92	89
Headcount Ratio	49%	67%	47%	73%
Deprivation Gap*	4871	4623	5307	5412
Per capita Deprivation Gap** (number of positions)	27	36	27	43

Panel B	Female		Male	
	Urban	Rural	Urban	Rural
Group size (Rmax)	178	129	195	125
Deprivation Threshold (rank)	4	9	6	7
Headcount Ratio	2%	6%	3%	6%
Deprivation Gap*	7.1	35.12	19.2	19
Per capita Deprivation Gap** (number of positions)	0.04	0.27	0.10	0.15

* Value of CDI before diving by n
** CDI value

Table 4: Distribution of functioning indicators among the deprived children

		Level of education		
Literate	Sage (average)	None	Incomplete primary	Complete primary
Female Rural				
96%	0.61	2%	98%	0%
Female Urban				
99%	0.72	4%	96%	0%
Male Rural				
100%	0.68	2%	98%	0%
Male Urban				
99%	0.74	4%	96%	0%

Table 5: Distribution of parental level of education among the deprived children in Education (Panel A) and in Living Conditions (PanelB)

Panel A

	None	Incomplete primary	Complete primary	Incomplete secondary	Complete secondary	Higher
Female rural						
Father's	26%	62%	6%	5%	1%	0%
Mother's	44%	52%	1%	2%	0%	0%
Female urban						
Father's	37%	34%	8%	10%	2%	9%
Mother's	35%	41%	7%	10%	4%	3%
Male rural						
Father's	27%	56%	6%	8%	2%	1%
Mother's	45%	48%	6%	1%	0%	0%
Male Urban						
Father's	27%	43%	5%	12%	5%	7%
Mother's	27%	55%	3%	8%	2%	4%

Panel B

	None	Incomplete primary	Complete primary	Incomplete secondary	Complete secondary	Higher
Female rural						
Father's	67%	33%	0%	0%	0%	0%
Mother's	100%	0%	0%	0%	0%	0%
Female urban						
Father's	50%	50%	0%	0%	0%	0%
Mother's	50%	25%	25%	0%	0%	0%
Male rural						
Father's	60%	40%	0%	0%	0%	0%
Mother's	80%	20%	0%	0%	0%	0%
Male Urban						
Father's	33%	50%	0%	17%	0%	0%
Mother's	17%	83%	0%	0%	0%	0%

Table 6: Distribution of exogenous variables among the deprived children children in Education (Panel A) and in Living Conditions (PanelB)

	Use of medical services	Average % social investment	Average monthly p.c. expenditure (Bs)	% of Agricultural population (in child's district)	Working status (yes)
Panel A					
Female rural	16%	65%	16.45	55%	34%
Female urban	10%	71%	30.53	22%	28%
Male rural	11%	58%	16.20	56%	53%
Male Urban	12%	67%	28.77	21%	21%
Panel B					
Female rural	16%	63%	14.89	47%	33%
Female urban	18%	79%	11.79	31%	75%
Male rural	0%	62%	5.15	54%	60%
Male Urban	40%	59%	15.57	11%	50%

Table 7: Bi-dimensional CDI. Urban male (panel A) Rural male (panel B)

Panel A

$\alpha = 0$

	θ				
a	1	2	3	5	Leontief
0.3	64.6%	64.6%	64.6%	64.6%	64.6%
0.5	64.6%	64.6%	64.6%	64.6%	64.6%
0.8	64.6%	64.6%	64.6%	64.6%	64.6%

$\alpha = 1$

	θ				
a	1	2	3	5	Leontief
0.3	20	25	27	30	35
0.5	22	27	29	31	35
0.8	25	28	30	31	35

$\alpha = 2$

	θ				
a	1	2	3	5	Leontief
0.3	1027	1493	1830	2235	3220
0.5	1291	1846	2165	2503	3220
0.8	2019	2375	2556	2749	3220

Panel B

$\alpha = 0$

	θ				
a	1	2	3	5	Leontief
0.3	85.6%	85.6%	85.6%	85.6%	85.6%
0.5	85.6%	85.6%	85.6%	85.6%	85.6%
0.8	85.6%	85.6%	85.6%	85.6%	85.6%

$\alpha = 1$

	θ				
a	1	2	3	5	Leontief
0.3	32	37	40	44	52
0.5	35	40	43	46	52
0.8	40	43	45	47	52

$\alpha = 2$

	θ				
a	1	2	3	5	Leontief
0.3	1906	2353	2727	3241	4676
0.5	2321	2853	3218	3651	4676
0.8	3262	3603	3812	4053	4676

Results for and $\alpha = 1, 2$ are in per capita number of positions.

Table 8: Bi-dimensional CDI. Urban female (panel A) Rural female (panel B)

Panel A

$\alpha = 0$

	θ				
a	1	2	3	5	Leontief
0.3	69.7%	69.7%	69.7%	69.7%	69.7%
0.5	69.7%	69.7%	69.7%	69.7%	69.7%
0.8	69.7%	69.7%	69.7%	69.7%	69.7%

$\alpha = 1$

	θ				
a	1	2	3	5	Leontief
0.3	19	24	26	28	34
0.5	22	26	28	30	34
0.8	25	28	29	31	34

$\alpha = 2$

	θ				
a	1	2	3	5	Leontief
0.3	993	1363	1647	2020	3011
0.5	1298	1738	2011	2322	3011
0.8	2019	2300	2456	2627	3011

Panel B

$\alpha = 0$

	θ				
a	1	2	3	5	Leontief
0.3	78.3%	78.3%	78.3%	78.3%	78.3%
0.5	78.3%	78.3%	78.3%	78.3%	78.3%
0.8	78.3%	78.3%	78.3%	78.3%	78.3%

$\alpha = 1$

	θ				
a	1	2	3	5	Leontief
0.3	27	31	33	36	43
0.5	29	33	36	38	43
0.8	33	36	37	39	43

$\alpha = 2$

	θ				
a	1	2	3	5	Leontief
0.3	1458	1948	2324	2799	4050
0.5	1831	2415	2770	3165	4050
0.8	2740	3114	3310	3525	4050

Results for and $\alpha = 1, 2$ are in per capita number of positions.

6 Conclusions

In this paper we propose a capability deprivation index that is based on scores derived from a structural economic model *explaining* the different capabilities using exogenous (social, economic, institutional) *causes*. We specify the freedom of choice given by any capability set as a latent variable partially observed through a vector of functioning indicators, and influenced by a collection of exogenous variables. Thus, our model measures deprivation in the functioning-capability space. The estimator of the latent variable vector, or factor scores, provides us with a measure of the well-being status of the individuals in the different dimensions. We propose an index that only uses the ordinal information (ranks) for defining the deprivation status including the quantitative information for defining weights. We distinguish two types of indices, one regarding deprivation within a capability set, and one considering aggregation across capability sets. That is, our index belongs to a class of multiple dimension(capabilities)- multiple indicator(functionings) indices. Both types of indices fulfill a set of desirable properties in the capability framework.

We applied our indices to analyze capability deprivation in the knowledge and living conditions of Bolivia's children in 2002. We differentiated four groups according to their gender and urban/rural location. Looking at individual dimensions, we find urban environment seems to enhance the choice set and more so for males in the knowledge dimension. The situation is similar for the living conditions domain though the difference between males and females is much less pronounced.

The bi-dimensional measures show the importance of living conditions in the overall deprivation as the deprivation measures tend to increase when their associated weight increases. They are also sensitive to the degree of substitution between dimensions, showing greater values, as the degree of substitution decreases. Finally, from the policy angle, it is interesting to note the role played by the exogenous variables in the CDI. The supply and demand factors, especially parental education, and the use of health services, largely account for how individuals turn out to be deprived. This highlights their importance as policy instruments for targeting the most deprived.

An important issue which needs to be explored and which we have not done in this paper is the choice of the minimum levels of achievement for determining the deprivation threshold, as we have used absolute criteria for identifying the deprived. It is clear that the results are sensitive to the criteria, and in order to make 'robust' conclusions, further investigations are needed.

7 Appendices

Table 9: Knowledge measurement equation results (panel A) and living conditions measurement equation results (panel B)

Panel A

Variable	Literacy		Level of Education		Sage	
	Standardized Coefficient	Significance	Standardized Coefficient	Significance	Standardized Coefficient	Significance
Knowledge capability						
Number of siblings	-0.106	***	-0.029	**	-0.068	*
Number of siblings aged 7-14	-0.091	***	-0.045	***	-0.027	***
Number of siblings aged 7-14 enrolled	0.453	***	0.166	***	0.189	***
Age	0.534	***	0.447	***	-0.156	***
Being indigenous	-0.041	*	-0.029	***	-0.082	***
Male			-0.033	***	-0.028	***
Male household head	-0.228	***	-0.097	***	-0.111	***
Working status			-0.039	***	-0.017	*
R^2	0.759		0.567		0.507	

Panel B

Variable	**Dwelling**		**Habitability**		**Basic Services**	
	Standardized Coefficient	Significance	Standardized Coefficient	Significance	Standardized Coefficient	Significance
Living conditions capability	0.532	-	0.546	***	0.532	***
Being indigenous					-0.031	***
Being poor	-0.052	***	-0.112	***	-0.065	***
Male household head	-0.048	***	-0.050	***	-0.031	***
Number of female adults	0.036	***	0.028	**	0.076	***
Number of children	-0.080	***	-0.233	***	-0.078	***
Urban	0.278	***	-0.076	***	0.436	***
R^2	0.500		0.446		0.695	

Table 10: Structural Model Results

Variable	**Knowledge capability equation**			**Living conditions capability equation**		
	Coefficient	Standardized. Coefficient	Significance	Coefficient	Standardized. Coefficient	Significance
Knowledge capability	-	-	-	0.078	0.129	***
Living conditions capability	0.124	0.075	***	-	-	-
Father's level of education				0.074	0.160	***
Mother's level of education	0.141	0.171	***	0.067	0.134	***
Belongs to main cities	0.109	0.044	***	0.128	0.085	***
Use of medical services	0.170	0.023	*			***
Number of schools	0.001	0.200	***			***
Number of classrooms	0.000	-0.273	***			***
% Agricultural population/PEA				-0.045	-0.016	
Monthly per capita expenditure				1.275	0.457	***
R^2	0.065			0.441		

***,* denote significance at 1% and 10

References

Bollen, K. (1989). *Structural Equations with Latent Variables.* Jhon Wiley and Sons, New York.

Bourguignon, F. and S. Chakravarty (1999). A family of multidimensional poverty measures. In: *Advances in Econometrics, Income Distribution and Methodology of Science* (D.,Slottje, Eds.). Physica-Verlag, London.

Bourguignon, F. and S. Chakravarty (2002). Multi-dimensional poverty orderings. Working Paper No. 2002-22, Delta, Paris.

Bourguignon, F. and S. Chakravarty (2003). The measurement of multidimensional poverty. *Journal of Economic Inequality* **1**(1), 25–49.

Chambers, R. (1994). The origins and practice of participatory rural appraisal. *World Development* **22**(7), 953–969.

Clert, C. (1999). Evaluating the concept of social exclusion in development discourse. *European Journal of Development Research* **11**(2), 176–199.

European-Foundation (1995). *Public Welfare Services and Social Exclusion: The Development of Consumer Oriented Initiatives in the European Union.* The European Foundation, Dublin.

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. and A. Nagar (2008). On exact statistical properties of multidimensional indices based on principal components, factor analysis, mimic and structural equation models. *Social Indicators Research* **86**(3), 481–496.

Krishnakumar, J. and P. Ballon (2008). Estimating basic capabilities: A structural equation model applied to bolivia. *World Development* **36**(6), 992–1009.

Maasoumi, E. and M. Lugo (2008). The information basis of multivariate poverty assessments. In: *Quantitative Approaches to Multidimensional Poverty Measurement* (Kakwani, N. and Silber, J., Ed.). Palgrave-MacMillan.

Muthen, B.O. (2002). Beyond sem: General latent variable modelling. *Behaviormetrika* **29**(1), 81–117.

Sen, A. K. (1983). Poor, relatively speaking. *Oxford Economic Papers* **35**(2), 153–69.

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models.* Chapman Hall, Boca Raton, USA.

Townsend, P. (1979). *Poverty in the United Kingdom: A Survey of Household Ressources and Standard of Living*. University of California Press, Berkley.

Tsui, K. (2002). Multidimensional poverty indices. *Social Choice and Welfare* **19**(1), 69–93.

Zheng, B. (1997). Aggregate poverty measures. *Journal Economic Surveys* **11**(2), 123–162.