the review of income and wealth

Review of Income and Wealth 2014
DOI: 10.1111/roiw.12127

DEPRIVATION AND THE DIMENSIONALITY OF WELFARE: A VARIABLE-SELECTION CLUSTER-ANALYSIS APPROACH

by Germán Caruso

University of Illinois at Urbana-Champaign

Walter Sosa-Escudero* and Marcela Svarc

Universidad de San Andrés and CONICET

We approach the problems of measuring the dimensionality of welfare and that of identifying the multidimensionally poor, by first finding the poor using the original space of attributes, and then reducing the welfare space. The starting point is the notion that the "poor" constitutes a group of individuals that are essentially different from the "non-poor" in a truly multidimensional framework. Once this group has been identified through a clustering procedure, we propose reducing the dimension of the original welfare space using recent blinding methods for variable selection. We implement our approach to the case of Latin America based on the Gallup World Poll, which contains ample information on many dimensions of welfare.

JEL Codes: C49, D31, I32

Keywords: clusters, factor analysis, multidimensional welfare, poverty

1. Introduction

Well-being and its related notions, like deprivation or inequality, are elusive concepts, and the efforts leading to define them precisely cannot be disentangled from the practical need to quantify them; to make valid comparisons, or to assess their importance. To complicate matters, a massive body of recent literature points toward the multidimensional nature of welfare (Sen, 1985). The mere notion of a concept being "multidimensional" is elusive as well, but it clearly suggests the inability to measure it based on a single scalar dimension, like income or consumption in the case of welfare. Moreover, even when there is agreement on the multidimensionality of well-being, there remains the problem of deciding how many dimensions are relevant, and which attributes or variables should be considered for a more accurate assessment.

The multidimensionality of welfare translates almost directly into that of poverty or deprivation. A recent line of research has focused on first solving the

Note: We thank Ricardo Fraiman for very useful insights and for kindly providing his computational routines. We also thank the Associate Editor and two anonymous referees for valuable comments which helped improve this paper considerably. All remaining errors are our responsibility. This paper is a follow-up of a study commissioned by the IDBs Latin American Research Network on Quality of Life in Latin America and the Caribbean. Gallup has generously provided the microdata of the Gallup World Poll 2006. Financial support from the Fondo para la Investigación Científica y Tecnológica (PICT 2008-1630) is acknowledged.

*Correspondence to: Walter Sosa-Escudero, Department of Economics, Universidad de San Andres, Vito Dumas 284, Victoria, Buenos Aires, Argentina (wsosa@udesa.edu.ar).

problem of dimensionality of welfare, that is, to identify how many relevant dimensions must be considered to measure welfare, and then proceeding to identify the "poor," based on this reduced set of variables. For example, Gasparini *et al*. (2011) and Ferro Luzzi *et al*. (2008) start with a rather large set of variables that can be seen as alternative measures of an underlying welfare space, and then use factor analytic methods in order to produce a small set of variables (factors) that appropriately capture the variability of welfare. The fact that more than one factor is needed to appropriately reproduce welfare is interpreted as evidence of its multidimensionality. After reducing the dimensionality of the problem, they proceed to find the poor, based on this reduced set of factors. Gasparini *et al*. (2011) identify the poor along each of the relevant dimensions, whereas Ferro Luzzi *et al*. (2008) apply cluster techniques on all relevant dimensions, to find a group of individuals that can be safely labeled as poor, in a multidimensional sense.

In this paper we adopt an alternative route that first identifies the poor and then explores the dimensionality of welfare. The starting point is the notion that the poor constitutes a group of individuals that are essentially different from the "non-poor," in a multidimensional framework. Once this group has been identified, we propose reducing the dimension of the original welfare space by finding the smallest set of attributes that can reproduce as accurately as possible the poor/non-poor classification obtained in the first stage. More concretely, we start by applying cluster methods on a rather large set of attributes, in order to identify a group that can be reasonably labeled as the poor. Once this satisfactory classification has been produced, in order to reduce dimensionality, we use recent methods on variable selection for cluster analysis. We implement the *blinding* approach of Fraiman *et al*. (2008) to find the smallest set of variables that is able to reproduce the poor/non-poor classification of the first stage. In this context, the multidimensionality of welfare (and hence poverty) is related to the fact that this reduced set includes more than one variable.

A first important advantage of this approach is that cluster methods guarantee high similarity within groups and high dissimilarity between groups, and hence, if it exists, the poor is a coherent group, by construction. Reducing the dimensionality first may unnecessarily complicate the goal of finding the poor based on the similarity–dissimilarity requirements of the cluster based approach. For example, the usual "single dimensional" classification based on poverty lines produces a sharp and unambiguous characterization of the poor/non-poor status. But a well known drawback is that individuals close to the poverty line are practically indistinguishable among them, inducing a classification of poor-non/poor that does not satisfy the dissimilarity requirements imposed on the groups. An advantage of our approach is to allow *all* variables in the welfare space to contribute toward the goal of identifying the deprived.

A second advantage refers to the interpretation of the results of factor analytic approaches. Factor methods have well known identification problems due to the fact that factors, even under stringent assumptions, are only identified up to arbitrary linear transformations or "rotations" (see ch. 10 in Hardle and Simar, 2003). That is, the usual output of standard factor analysis is a set of variables (factors) that are linear combinations of the original variables. It is standard

practice to exploit these rotations to favor factors that are formed by combining a few variables (like in the "varimax" rotation), but there is no guarantee that these arbitrary linear combinations will have a meaningful interpretation. On the contrary, our variable-selection approach is free from these ambiguities, since by construction, the reduced set of variables identified in the second stage is a strict subset of the variables originally in the welfare space, hence they are readily interpretable. Moreover, the construction of factors would require us to re-sample all variables whereas the variable-selection approach, if successful, would require us to re-sample only the variables that are selected in the procedure. Clearly, both methods are specific to the initial choice of the welfare space, so their relative merits refer to this choice.

The goals of this paper require the use of a data set that contains a large set of variables that jointly represent all relevant dimensions of welfare. This has usually been a hindrance in applied studies since available data usually focuses on some specific dimensions like those included in standard household surveys (typically income, expenditure, and other socioeconomic variables), but usually excluding aspects that the recent literature on multidimensional welfare emphasizes, in particular those related to subjective notions of welfare. In this paper we implement the proposed strategy using the Gallup World Poll, a comprehensive data set that includes questions on objective and subjective attributes of welfare, that can appropriately provide a starting point for the goals of identifying the poor and studying the multidimensionality of welfare. In spite of being a very rich source of information, its use for research purposes is relatively new; see Gasparini *et al.* (2011) for a detailed review of this data set and a comparison with other more standard sources like national household surveys.

The paper is organized as follows. The next section discusses in more detail the problems of multidimensional welfare and poverty and its empirical consequences. Section 3 describes the proposed methodology, based on recent cluster variable-selection methods. Section 4 describes the Gallup Poll data set and presents the empirical results. Section 5 concludes and discusses further research.

2. Multidimensional Welfare and Poverty

The seminal work by Sen (1985) and its related literature (for a recent collection of results see Kakwani and Silber, 2008) clearly point toward the multidimensionality of welfare, in the sense that it cannot be appropriately represented by a single dimensional notion like income or consumption. Consequently, the status of poor arises as a consequence of assessing all relevant dimensions involved in determining well-being. Were these dimensions conceptually known in advance, the natural way to quantify welfare is to measure each of them empirically, in which case the number of variables coincides exactly with the dimension of the welfare space. The mere fact that welfare is multidimensional simply states that one variable is not enough to properly capture it, without clear signs of which variables to measure and map to each dimension. In this context, a large socioeconomic household survey can be seen as a collection of variables that *together* capture the variability of welfare. Two natural and related questions are: (1) How to find the poor based on the information provided by such a large set of

attributes? (2) Which is the dimensionality of welfare? That is, how many underlying variables are relevant to capture welfare and, eventually, if it is possible to represent the whole welfare space in terms of a few variables or indexes.

The problem of choosing the relevant dimensions to assess welfare, and later on, poverty, is a difficult one, that faces researchers with conceptual as well as operational restrictions. The practice of selecting and measuring coordinates to quantify poverty is usually approached based on a conceptual characterization, usually within the *capability approach* advocated by Sen (1985). But sometimes existing data or previous practices play a dominant role in this choice. Alkire (2007) presents a detailed description of these procedures and highlights the need for a more systematic and better documented practice regarding choosing dimensions to assess multidimensional poverty.

A recent line of research has relied on factor analytic methods to attack the problem of dimensionality. That is, welfare is thought of as being appropriately represented by a few latent, not directly observed factors. Observed variables are then seen as arising from linear combinations of these factors, hence the empirical problem consists in recovering these latent factors based on the observed variables. The fact that welfare is multidimensional is linked to the relevance of more than one factor.

This is the approach adopted by recent papers by Ferro Luzzi *et al*. (2008) and Gasparini *et al*. (2011), with promising results. Gasparini *et al*. (2011) base their analysis on the Gallup World Poll, and their initial data set contains 15 variables, including income, and other monetary and non-monetary measures of welfare, as well as some indicators related to subjective welfare. They conclude that their initial space of 15 variables can be reasonably represented by three factors. The first one is based mostly on income. The second one is interpreted as related to subjective welfare, since it is mostly composed of questions related to this concept, and finally, the third one is related to standard "basic needs" measures, like water access. Ferro Luzzi *et al*. (2008) start with 32 variables from the Swiss Household Panel, and conclude that they can be appropriately represented by four latent factors: financial, health, neighborhood, and social exclusion dimensions.

To summarize, both papers find evidence that the original welfare space, composed of many relevant measures, can be drastically reduced to a few factors, and that more than one variable is needed to adequately represent it, even when income (in the case of Gasparini *et al*., 2011) or variables closely related to it (the financial ones in the case of Ferro Luzzi *et al*., 2008) are included in their data sets. In spite of being strongly associated to a relevant factor, both studies point toward the inadequacy of solely income to capture the multidimensional nature of welfare.

Regarding the problem of finding the poor, both papers attempt to derive the poverty status based on the reduced welfare space, that is, on the factors obtained in the first stage. Gasparini *et al*. (2011) do not attempt to produce a single notion of poverty, instead they compute a poverty status for each of the relevant factors, that is, they set poverty lines for each of the factors separately, and produce poverty rates for each dimension (see Gasparini *et al.*, 2011 for further details). Ferro Luzzi *et al*. (2008), on the other hand, produce a single notion of poverty by using cluster methods based on their reduced welfare space, that is, they use the factors produced in their initial stage as an input for standard clustering

algorithms to identify coherent groups. They find that the scores obtained in the factor analysis stage can be reasonably grouped in three clusters for 1999, two for 2000 and 2001, and four for 2002 and 2003. In all cases, these authors find that one group presents substantially low values for all scores and hence this particular group is labeled as the "multidimensional poor."

There are several methodological concerns related to this approach, which basically consists in a first stage where the dimensionality of the original welfare space is reduced using factor methods, and then the poor are found based on this reduced set. First, though immensely popular in other disciplines (Psychology, for example), covariance methods like factors or principal components are scarce in Economics. This is mostly due to their well known identification issues which harm their direct interpretation. Basically, factors are linear combinations of the original variables, identified up to orthogonal transformations. The standard practice, and the one adopted in both Gasparini *et al*. (2011) and Ferro Luzzi *et al*. (2008), is to rely on "rotations" or other algebraic transformations to produce interpretable results. These rotations, like the popular "varimax" rotation, favor factors that are positive linear combinations of a few variables, so, in the best case scenario, factors are like averages of subsets of variables in the initial set. For example, in Gasparini *et al*. (2011), their second factor (obtained with a varimax rotation) is formed through a positive linear combination of subjective welfare variables, leading them to label this factor as capturing "subjective welfare." But ex-ante, there is no guarantee that the optimal factors will be composed of just a few variables that can be linearly combined to produce interpretable results. Elffers *et al*. (1978) provide a lengthy discussion of the interpretability of factor models.

Second, factors are not directly observable, but constructed as linear combinations of variables in the original space. Consequently, for practical reasons, new information must be constructed by sampling the whole set of initial variables. For example, suppose that the analysis must be repeated for a different period or region, then all the initial variables must be measured in order to construct the factors, even under the assumption that the underlying latent structure remains unchanged.

Finally, reducing the dimensionality first may unnecessarily complicate the identification of a coherent group (the poor) that can be safely distinguished from its complement (the non-poor). This is particularly relevant when most variables in the welfare space consist in categorical (in most cases, binary) variables. The aggregation process implicit in the factor analytic approach may smooth out relevant differences contained in the original welfare space. For example, standard income based poverty lines have serious troubles distinguishing the poor from the non-poor when the distribution of income is densely populated around the poverty line. Other categorical indicators may actually help in separating the poor from the non-poor.

3. The Variable-Selection Cluster Analysis Approach

Based on the concerns of the previous sections, we will explore an approach that (1) preserves the original welfare space in order to identify the poor, and (2)

can reduce its dimensionality by producing unambiguously interpretable variables, that can be resampled or reconstructed easily.

Our strategy starts by applying cluster methods to the original welfare space. Once the poor are satisfactorily identified, the problem of dimensionality is solved by finding the smallest set of variables in the original welfare space, that can reproduce the poor/non-poor classification of the first stage, as accurately as possible. We will use recent results on variable selection for cluster analysis. As in the case of factor methods, "multidimensionality" will be related to finding more than one variable in this reduced set of variables.

Regarding interpretability, and unlike factor approaches, our strategy produces immediately interpretable and reproducible variables, since the reduced set is a strict subset of the variables sampled and contained in the original space. Additionally, and unlike latent-based strategies like factor analysis, further studies would require us to collect information only in the optimal subset. Naturally, both factor and cluster methods are subject to sample variability related to the initial data set, so their relative advantages are to be assessed for a given data set.

Before describing in detail our empirical strategy, we must comment on some limitations. First, the cluster approach is not free from identification and interpretation issues. Cluster methods cannot guarantee in advance that the optimal number of groups is necessarily two; moreover, the methods do not guarantee that even if two groups are found, these are economically different. Second, even if two groups are found, this does not necessarily mean that one of them is the poor and the other one the non-poor. For example, one group might consist of the "extremely rich" with the complement group containing all other individuals. The next subsections describe in detail the clustering methods used in this paper, and how they are exploited to deal with the aforementioned problems; in particular, to guarantee that there are actually two separate groups (instead of only one group or more than two) and that one of them can be safely regarded as containing the poor. The second subsection describes the variable selection approach.

3.1. *Clustering Methods and the Poor*

The underlying idea behind our empirical strategy is to understand the poor as a coherent group that can be conceptually and practically distinguished from its complement, the non-poor. Cluster methods seem relevant since, by definition, they solve a within/between similarity trade-off, that is, they try to assign observations to groups so they are close to those in the same group and distant to those in other groups. Even though classical clustering algorithms have long been available, recent advances in data mining and computer intensive methods have driven considerable attention to such techniques; see Cherkassky and Mulier (2007) or Bishop (2006) for a recent overview.

The input for cluster methods is an $N \cdot p$ matrix X, where rows correspond to N observations, and columns to the p variables, together representing (multidimensional) welfare. Each row can be seen as a point of dimension p. A cluster is a collection of these points. If there are K clusters, a clustering mechanism can be characterized by a function or "encoder" $C(i)$: $(1, \ldots, N) \rightarrow (1, \ldots, K)$ that assigns each point to only one group. The main goal of cluster analysis is to

produce assignments so that points within a group are similar and simultaneously different from points in all other groups. Hence, the notion of "similarity" is crucial. Let $d(x_j, x_i)$ be a distance function for any two p dimensional points x_j and x_i, each corresponding to rows of the matrix X. Consider the following loss function

$$W(C) = \frac{1}{2}\sum_{k=1}^{K}\left[\sum_{i,j/C(i)=C(j)=k} d\left(x_i, x_j\right)\right],$$

defined over the space of all possible encoding functions. Cluster algorithms find the encoding that minimizes this penalty function, which measures aggregate discrepancies within each cluster. A natural approach consists in checking all possible clusterings, which is usually computationally inviable. A useful alternative is to specify a distance function. For example, the k-means algorithm (MacQueen, 1967) takes $d(\cdot, \cdot)$ to the be the square of the Euclidean distance $d(x_j, x_i) = \|x_i - x_j\|^2$. It can be shown that in such case

$$W(C) = \sum_{k=1}^{K} N_k\left[\sum_{i/C(i)=k} \|x_i - \bar{x}_k\|^2\right],$$

where N_k is the number of points in cluster k, and $\bar{x}_k$ is the vector of means of cluster k. Standard algorithms start with an initial classification, compute the K vectors of means, reassign observations to clusters so observations are closest to the previously computed means, and iterate this process. Alternative choices for distance functions and centers lead to different solutions, like the *k—medians* algorithm that replaces the mean by the median. Standard results on the asymptotic behavior of the *k—means* procedure are given by MacQueen (1967) and Hartigan (1978). Pollard (1979) established conditions that ensure the almost sure convergence of the cluster centers as the sample size increases, in any general metric space with compact balls. In addition, it has a good performance on many real data examples, as recently pointed out by Coates *et al*. (2011).

This characterization of the poor as a cluster leads to a natural comparison with available methods to find the poor in a multidimensional framework, like the recent proposal by Alkire and Foster (2011). Standard methods to identify the poor multidimensionally usually start by defining deprivation along each dimension, that is by defining a threshold along each welfare dimension, below which a person is considered poor. "Union" methods define as multidimensionally poor a person who is deprived in *at least* one dimension, while "intersection" methods require that the person is deprived in all dimensions. Alkire and Foster (2011) propose a "counting," intermediate method that lies between one and all dimensions.

Cluster-based poverty can be seen as an alternative, intermediate strategy. First, the methods discussed above depend on exogenous cutoffs. In the cluster approach, the cutoffs are determined endogenously, as a solution to the dissimilarity optimization problem that simultaneously defines clustering. Second, the clustering algorithm discussed above leads to a partition of the welfare space that separates the poor from the non-poor. More concretely, suppose there are two groups. The optimal solution

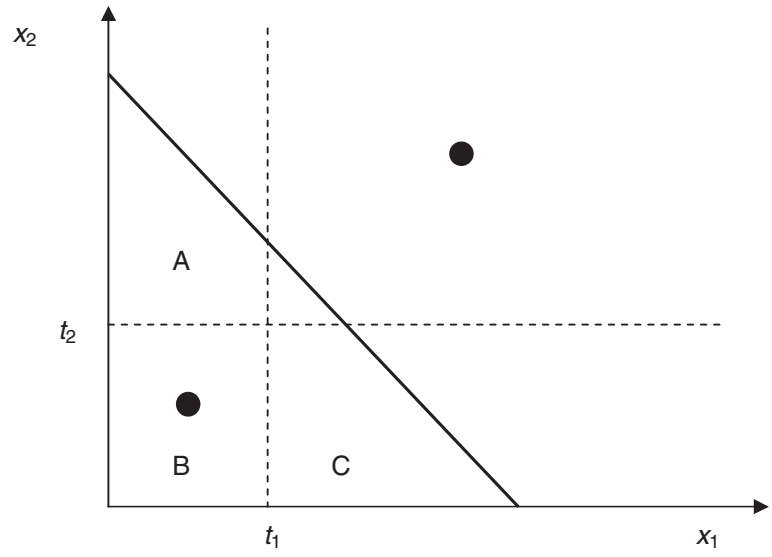

Figure 1. The Poor as a Cluster

defines two centers, labeled as c_1 and c_2, so in the final step of the clustering algorithm all points belong to the cluster that leaves them the closest to either center. Consequently, the separation in two groups implicitly defines a partition, the so-called *Voronoi tesselation*, whose frontier is a hyperplane that separates, in our case, the poor from the non-poor, defined implicitly by the inequality

$$d(z, c_1)^2 < d(z, c_2)^2,$$

where z is a point in $\mathbb{R}^p$. Union and intersection methods imply particular partitions of this space. In the case of the union method, the partition is defined by the inequalities $z_j < t_j$ for at least one j in $(1, \ldots, p)$, where t_j is the poverty threshold for dimension j. For the intersection method, the separating hyperplane is defined by the inequalities $z_j < t_j$ for all j.

Figure 1 illustrates this point graphically. For the case of two dimensions, t_1 and t_2 represent the poverty thresholds in each dimension. The intersection method defines the rectangle labeled as "B" as containing the poor, while the union approach adds regions A and C. The solid dots represent the "centers" for the poor and the non-poor, and the solid line renders as poor all points below it. The line is implicitly defined as separating points closest to each of the centers. Duclos *et al.* (2011) present recent advances on the subject.

There are several difficulties that must be sorted out for the case of finding the poor. First, our data is of a mixed nature, that is, it contains categorical (mostly binary) as well as continuous variables (income, for example). This impacts on the choice of an appropriate clustering technique, since these methods are sensitive to the choice of distances, standardizations, and initial conditions. Second, as previously discussed, the final goal is to guarantee that the process finds two essentially different groups, one of them containing the poor.

Regarding the choice of a clustering method for our mixed data, we started by standardizing all variables. This is common practice in this literature, to avoid scale effects. Each variable is divided by its range, that is, for the observation x_{ij} we consider y_{ij}, the standardized observation,

$$y_{ij} = \frac{x_{ij}}{\max_j(x_{ij}) - \min_j(x_{ij})}.$$

This procedure is applied to all variables except to monthly household income, a continuous and highly positive skewed variable. For this case we standardize based on its natural logarithm. Consequently, all standardized variables have the range [0, 1], except for the monthly household income that has the range [–1, 1].

The *k*-means algorithm is sensitive to the choice of an appropriate distance. We have chosen an additive measure that can handle mixed as well as continuous variables. The L_1-norm is a natural choice for our type of data. The distance between two observations y_i and y_j is given by

$$d_{ij} = \sum_{l=1}^{p} |y_{il} - y_{jl}|.$$

So, it can be seen as being the standard L_1-norm for continuous or ordinal variables, and in the case of binary variables, as the number of points where the observations take different values, that is, the same information as in the standard Jaccard index (see Hand *et al.*, 2001), one of the most well known measures of similarity for binary variables.

K-means procedures are often very sensitive to the choice of initial conditions, that is, to the position of the initial centroids used to start the algorithm. Several proposals have been made to handle this effect (see Steinly and Brusco, 2007). We have followed the recommendations in this last reference and considered ten random initializations, keeping the one with minimum within-cluster sum of squares.

Regarding the number of clusters, most methods produce forced partitions on any data set; there is either an endogenous structure or not. Hence, in order to find the poor we are interested in two null hypotheses. The first one is the null that no grouping exists versus the alternative that there is more than one group. The second one is the null that only two groups are relevant, against the alternative that more than two groups are needed. We use the standard Calinski and Harabasz (1974) statistic, the most frequently used method, to find the optimal number of clusters. We also use the more modern *gap statistic* introduced by Tibshirani *et al.* (2001). The intuition behind this statistic is that even though within cluster similarity decreases as the number of groups increases, further partitioning a group with already high similarity reduces the within cluster similarity less than partitioning a heterogeneous group. Then, a sharp decrease will be observed at the optimal number of groups.

Finally, even when the previous process leads to two significantly different groups, there is no guarantee that one of them can be safely labeled as containing the poor, that is, the relevant partition might cluster the extremely rich in one group. We will implement some confirmatory tests, based on a multivariate

version of the Komogorov–Smirnov test, to explore the nature of the implied partition and to what extent one of them contains the poor.

3.2. *Dimensionality through Variable-Selection*

After having found an appropriate clusterization that divides the population into poor/non-poor groups, the problem of reducing the dimensionality of the original welfare space will be handled as a variable-selection one. The main advantage of this approach is that, by construction, the resulting variables are directly interpretable since they are originally in the data set used as a starting point to represent welfare space.

In recent years, and driven by the increased popularity of data mining methods, several proposals have been made to overcome this problem. Most of them involve a clustering technique, a rule to determine the number of clusters, and a procedure to select the variables. We adopt the recent strategy in Fraiman *et al*. (2008) based on a "blinding" process that eliminates unnecessary variables. These authors show that the process has good empirical performance, especially as compared to alternatives like those by Tadesse *et al*. (2005) and Raftery and Dean (2006).

Fraiman *et al*.'s (2008) procedure selects relevant variables after a satisfactory clustering procedure has been implemented. Their approach is based on the idea of blinding unnecessary non-informative or redundant variables. We will discuss the main intuitive ideas behind the procedure; details are provided in the Appendix. For simplicity, suppose there are only two variables in the original space, X and Y. Given an appropriate clusterization based solely on X, Y is redundant if (a) it is strongly related to X, so given X it adds little information to the clusterization, and (b) it is independent of X and non-informative about any clusterization (it only adds "noise"). In these cases, the clusterization remains relatively unaltered if Y is replaced by its best prediction based on X, its conditional expectation $E(Y|X)$. In the extreme versions of the previous cases, Y will be replaced by X (X strongly related to Y) or by a constant (Y just adding noise). Consequently, the goal is to find the smallest group of original variables that can reproduce the original clusterization as accurately as possible, by replacing redundant variables by their expectations conditional on this reduced subset. The algorithm is detailed in the Appendix and in the original paper by Fraiman *et al*. (2008). The variable selection procedure is shown to be strongly consistent under mild regularity conditions on the partitioning method, and on the (non-parametric) estimation of the conditional expectations in the blinding process. Though intuitively simple, the method can be computationally extremely expensive. Fraiman *et al*. (2008) introduce a forward–backward algorithm in order to find a subset of variables with the desired properties. A Matlab computational code to implement this procedure is available upon request to the authors. Finally, even though there are alternative methods for clustering, we will favor partitioning strategies, like k–means for two reasons. Clustering algorithms can be classified mainly in two categories, hierarchical and partitioning clustering (see Bishop, 2006). Hierarchical clustering strategies are used mainly on small data sets, since they are computationally very expensive, and their convenient interpretation properties of their output are lost in large data sets. Partitioning algorithms decompose the data set into a set of disjoint clusters and

are much less expensive from a computational point of view, and hence more convenient for large data sets. Also, variable selection methods are still not developed for hierarchical methods, hence our preference for partitioning methods.

4. Empirical Results

4.1. *Data*

The main input for our analysis is a set of variables that covers the most relevant dimensions of welfare. To this purpose, the Gallup World Poll, collected by the Gallup Organization, provides a convenient framework. The Poll is based on a consistent and homogeneous questionnaire implemented on national samples of adults from 132 countries, providing an exceptional chance to make cross country comparisons. The Gallup World Poll contains an ample spectrum of questions related to welfare, including self-reported measures of quality of life, opinions, and perceptions. It also incorporates fundamental questions on demographics, education, and family income. Respondents are adults (15 years or older), selected randomly within the household. In spite of its potential, the Gallup Poll is still relatively unexplored for research purposes. Gasparini *et al*. (2011) and Gasparini and Gluzman (2012) provide a detailed account of its adequacy and compare it with standard household surveys. They conclude that in many comparable dimensions, the information contained in the Gallup Poll is a valuable and reliable source for welfare analysis.

The use of the Gallup Poll for the analysis of well-being is relatively recent, since only in 2006 did the Poll extend its coverage to include most countries, in particular those in Latin America, and include questions about life satisfaction. See Graham and Behrman (2009), and, more generally, Graham and Lora (2009) for a collection of welfare studies based on the newer version of the Gallup Poll.

Regarding the purposes of this paper, the levels of poverty rates differ between the Gallup Poll and the corresponding national household surveys for each country. Nevertheless, Gasparini *et al*. (2011) report that the correlation between poverty estimates using these two sources is positive and significant; in particular, poverty rankings by country are remarkably similar using both sources, suggesting that despite providing a rougher approximation to per capita income, the picture of poverty in Latin America that arises from the Gallup data is not very different from the one obtained from the national household surveys.

The choice of an initial set of variables is certainly arbitrary and depends strongly on conceptual as well as pragmatic reasons, as stressed by Alkire (2007). In our case, this choice favors comparison with previous results. Consequently, our initial data set consists of the 15 variables used initially by Gasparini *et al*. (2011) for Latin America and the Caribbean (LAC) region, as their welfare space.[1] The final sample size is 14,108 individuals, for the following countries for which all variables are observed (number of observations per-country in parentheses): Argentina

[1]In the final version of their paper, Gasparini *et al*. (2011) drop two variables (whether in the last year respondents felt they lacked enough money to satisfy their shelter needs, and whether in the last year they felt hungry) to avoid missing observations for some of the countries of the LAC region. For the purposes of our paper, we include three additional variables and lose three countries, favoring a better representation of the initial welfare space.

(1000), Bolivia (1000), Chile (1007), Colombia (1000), Costa Rica (1002), Ecuador (1067), El Salvador (1000), Guatemala (1021), Honduras (1000), Nicaragua (1001), Panama (1005), Paraguay (1001), Peru (1000), and Uruguay (1004).

The initial set of variables in Gasparini *et al*. (2011) is driven by their need to quantify an initial and tentative representation along three concepts of welfare:

1. *Monetary welfare:* income is a widely used measure of welfare, and unlike consumption or expenditures, usually available in standardized household surveys. We use the income measure in the Gallup survey, which consists of monthly household income before taxes. Since the original question in the Gallup data set is posed in terms of brackets of income, we proceed as in Gasparini *et al*. (2011), and take a random value in the corresponding range of the original question in local currency units, assuming that the shape of the income distribution is similar to the one from the national household survey. This value is converted to U.S. dollars using country exchange rates adjusted by purchasing power parity. Household per capita income is constructed by dividing income by an estimate of the number of members in each household, since the Gallup survey does not provide this figure. Number of members is estimated as the number of children under 15 (available in the Gallup Survey) plus the average number of adults older than 15, as reported in the corresponding household surveys for each country. Gasparini and Gluzman (2012) provide a detailed comparison of incomes in the Gallup Poll and in national surveys, and conclude that, in spite of the several limitation of the former, it leads to similar rankings of measures based on income, like mean incomes or poverty.
2. *Non-monetary welfare:* these variables capture alternative access to goods and services that impact directly on welfare, but are not necessarily well captured by income. We include access to running water, electricity, landline telephone, television, computer, internet, or mobile phone.
3. *Subjective welfare:* some recent literature (Ravallion and Lokshin, 2002 is a leading example) has emphasized the importance of complementing standard measures with self perceived notions of well-being, finding significant differences between self-rated and objective measures of welfare concepts like poverty. We include questions on how individuals perceive themselves regarding welfare.

A complete list of variables with a more detailed description, is provided in the Appendix. It is relevant to remark that this ex-ante classification obeys descriptive purposes solely, since the key idea behind our clustering procedure is to find the poor and the dimensionality of welfare without exploiting any grouping of the initial variables.

Table 1 presents the correlation matrix for this initial set of variables.[2] Interestingly, although variables are positively correlated, none of these correlations is

[2]Our data set includes continuous as well as discrete variables, for which the choice of a proper correlation measure is not trivial. For two continuous or ordinal variables we have used the standard Pearson correlation coefficient. For two binary variables, we have computed Cramer's ϕ correlation coefficient. Finally, when one variable is binary and the other one is continuous, the Pearson correlation is equivalent to the *point biserial* correlation coefficient, appropriate for dichotomous variables. See Sheskin (2011) for details on correlations for categorical, continuous, and mixed data.

TABLE 1
Correlation Matrix

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15
V1	1.00	0.42	0.60	0.28	0.22	0.12	0.21	0.14	0.09	0.20	0.19	0.17	0.18	0.15	0.22
V2	0.42	1.00	0.17	0.08	0.11	0.08	0.13	0.09	0.08	0.15	0.10	0.10	0.11	0.10	0.14
V3	0.60	0.17	1.00	0.24	0.16	0.08	0.14	0.13	0.05	0.11	0.20	0.13	0.15	0.14	0.14
V4	0.28	0.08	0.24	1.00	0.24	0.14	0.19	0.08	0.05	0.13	0.10	0.09	0.11	0.10	0.10
V5	0.22	0.11	0.16	0.24	1.00	0.49	0.49	0.11	0.11	0.18	0.14	0.14	0.18	0.15	0.22
V6	0.12	0.08	0.08	0.14	0.49	1.00	0.32	0.08	0.10	0.12	0.08	0.12	0.12	0.08	0.15
V7	0.21	0.13	0.14	0.19	0.49	0.32	1.00	0.13	0.15	0.19	0.15	0.18	0.17	0.11	0.23
V8	0.14	0.09	0.13	0.08	0.11	0.08	0.13	1.00	0.21	0.23	0.09	0.25	0.14	0.09	0.18
V9	0.09	0.08	0.05	0.05	0.11	0.10	0.15	0.21	1.00	0.20	0.15	0.56	0.10	0.06	0.23
V10	0.20	0.15	0.11	0.13	0.18	0.12	0.19	0.23	0.20	1.00	0.16	0.26	0.32	0.25	0.26
V11	0.19	0.10	0.20	0.10	0.14	0.08	0.15	0.09	0.15	0.16	1.00	0.19	0.25	0.21	0.22
V12	0.17	0.10	0.13	0.19	0.14	0.12	0.18	0.25	0.56	0.26	0.19	1.00	0.16	0.09	0.27
V13	0.18	0.11	0.15	0.11	0.18	0.12	0.17	0.14	0.10	0.32	0.25	0.16	1.00	0.57	0.28
V14	0.15	0.10	0.14	0.10	0.15	0.08	0.11	0.09	0.06	0.25	0.21	0.09	0.57	1.00	0.25
V15	0.22	0.14	0.14	0.10	0.22	0.15	0.23	0.18	0.23	0.26	0.22	0.27	0.28	0.25	1.00

Note: The variables in the table follow the order of the variables in the Appendix.

TABLE 2
Optimal Number of Groups

Clusters (k)	CH Index	Gap(k)	S_k	Gap($k+1$) − S_{k+1}
1		0.368760	0.004639	0.553242
2	2087.42	0.586392	0.033150	0.381457
3	1441.01	0.391880	0.010423	
4	1570.45			
5	1305.30			
6	1394.86			
7	1283.33			

Notes: CH stands for the Calinski/Harabasz index. The Gap procedure chooses the optimal number of clusters (k) by finding the smallest k such that $Gap(k) \geq Gap(k+1) - S_{k+1}$.

exceedingly high, suggesting that, at least ex-ante, all initial of them seem to contribute to the variability of welfare.

4.2. *The Poor as a Cluster*

As described in the previous section, the first step is to find the optimal number of clusters using the k-means algorithm. Table 2 presents results of the Calinsky/Harabasz index and the relevant information for the Gap statistics. The Calinsky/Harabasz index decreases monotonically, achieving a maximum at two clusters. Also, the procedure based on the Gap statistic suggests that the optimal number of clusters is two.

The previous results and the nature of the clustering algorithm suggest that there are essentially two different groups, at least under the metric used to define similarity in the clustering procedure. Nevertheless, as a robustness check, we have implemented a multivariate variant of the non-parametric Kolmogorov–Smirnov (KS) test, developed by Cuesta-Albertos *et al.* (2006) (see also Opazo *et al.*, 2009 for a recent application), which can be applied to either functional or multivariate

data. Roughly speaking, it is based on performing unidimensional KS tests for the projections of the data in randomly selected directions. We proceeded as suggested by Cuesta-Albertos *et al.* (2006), by selecting 50 random projections, computing the KS statistic for every case, and taking the maximum of these values. The corresponding p-value is less than 0.001, meaning that the distributions of both groups induced by clusterization are significantly different. In addition, the results of a standard Hotelling test for differences in means of the two groups clearly suggest statistically relevant differences, with a p-value smaller than 0.0001. We remark that this can be seen as a confirmatory analysis, since by construction, the clustering algorithm maximizes between group separation in the metric of the *k*-means algorithm.

The previous results point toward the existence of two *statistically* different groups, but it remains to explore whether one of them can be seen as containing the poor. As discussed in Section 3.1, in the cluster approach the notion that one of the groups is poor is endogenous. That is, unlike standard poverty approaches, a natural advantage of perceiving the poor as an internally coherent group different from its complement is that it does not require the ex-ante explicitation of a "poverty line" (multidimensional, in this case). Hence, the sense in which one group is indeed rendered as poor comes from examining that in most variables one of them appears as "deprived" in several relevant dimensions.

Table 3 presents means for the two groups obtained in the clustering process. For completeness, we have also compared the means of the three optimal factors obtained by Gasparini *et al.* (2011), interpreted by these authors as representing monetary, subjective, and non-monetary aspects of welfare. Group one (non-poor)

TABLE 3

Characteristics of the Groups

		Group	
Type of Variable	Total Set of Variables	Non-Poor	Poor
Optimal factors	Monetary welfare	232.5077	96.1866
	Subjective welfare	0.5941	–1.9006
	Non-monetary welfare	0.0147	–0.4568
Monetary welfare	Per capita family income	232.5077	96.1866
Non-monetary welfare	Computer	0.2819	0.0399
	Electricity	0.9806	0.8934
	Internet	0.1190	0.0058
	Landline telephone	0.6435	0.2780
	Mobile	0.5426	0.2519
	Running water	0.9347	0.8110
	Television	0.9545	0.7899
Subjective welfare	Best possible life at the present	6.1280	4.5426
	Best possible life in the future	7.2766	5.8018
	Best possible life in the past	5.7545	4.7527
	Enough resources for housing	0.9456	0.5019
	Enough resources for the family food	0.8786	0.0587
	Hungry at least three times in a year	0.9588	0.3857
	Satisfaction with the standard of living	0.7362	0.3857
Frequency		0.7352	0.2648

contains 73.52 percent, and group two the remaining 26.48 percent (poor) of the individuals in our sample. Group two presents substantially lower values for most variables and all indexes obtained through factor analysis, suggesting that this group contains those individual with low levels of welfare. Per capita family income per month is US$232.50, compared to US$96.18 for the second cluster. Most variables reinforce this result. For example, in group one, 64.35 percent of the observations have access to a telephone line, compared to only 27.8 percent in the second cluster. Also, 73.62 percent of the observations in group one are satisfied with their standard of living, compared to only 38.57 percent in group two. These results suggest that both groups are statistically and economically different, with the second one containing individuals with significantly lower levels of welfare.

4.3. *Dimensionality via Variable-Selection*

After having found an acceptable clusterization, we have proceeded to solve the dimensionality problem by finding a reduced set of variables, initially in the welfare space, that can reproduce the initial grouping. As stressed in the previous section, the Fraiman *et al*. (2008) procedure is computationally very expensive, with required computer time growing exponentially with the sample size and the number of variables. In our case, it is unfeasible to perform the procedure with the complete data set (requiring more than 100 days to compute it with a standard computer). We have implemented a subsampling strategy, by considering ten random subsamples, each of them containing 85 percent of all the observations in the original data set. We have used a standard "probability proportional-to-size" sampling scheme, where sampling probabilities are proportional to the relative sizes of the groups. That is, observations in the non-poor group are chosen to be part of the subsample with probability 73.52 percent, and those in the second group, with probability 26.48 percent. The variable selection procedure was then applied to each of these subsamples. This randomization strategy reduces computational time considerably (approximately six days, with a standard personal computer).

Remarkably, in all subsamples the variables selected are: *monthly household income*; *not having had enough money to buy food over the last year on at least three occasions;* and *having a computer at your home or the place you live*. The correct cluster reallocation rate is always between 90 percent and 92 percent. That means that almost all individuals classified as poor with the initial set of 15 variables are correctly classified as poor/non-poor based on this much smaller set of three variables.

The fact that the reduced space needs more than one variable to adequately reproduce the original welfare space is an indication of its multidimensionality. Nevertheless, income turns out to be one of the variables chosen in the reduced set. This result is consistent with the previous literature which suggests that, though important, income is not sufficient to capture all the dimensions of welfare. As a matter of fact, when the reduced set of variables is forced to keep only income, only 60 percent of the observations are reallocated on the correct cluster.

It is interesting to compare the results of our multidimensional approach, with a standard one based solely on income. Table 4 presents, in Panel A, results of

TABLE 4
CLUSTER CLASSIFICATION AND INCOME BASED MEASURES

Panel A: Percentage of Correct Classification by Poverty Line	
Poverty Line	Correct Classification
0.5	73.91%
1	73.91%
2	73.91%
4	73.90%
Panel B: Income and Cluster Poor	
Decile	Cluster poor
1	54%
2	46%
3	37%
4	30%
5	25%
6	23%
7	21%
8	16%
9	11%
10	3%

Notes: Panel A: Percentage of individuals correctly classified as poor or non-poor by a cluster analysis and by income for each daily poverty line in U.S. dollars.

Panel B: Percentage of individuals in each income decile, classified as cluster poor.

classifying individuals using both our cluster method and a standard poverty line. We considered four alternative poverty lines, ranging from 0.5 to 4 dollars a day, including the widely used one and two dollars a day. The table reports, in Panel A, the percentage of individuals that both methods classify in the same group. That is, for example, using one dollar a day as the income poverty threshold, 73 percent of all individuals are classified the same using both methods. This is a relevant result since it suggests that the multidimensional perspective implicit in the cluster method cannot be replicated with income solely. Second, the misclassification rate remains unaltered to the levels of the poverty line or, equivalently, to the translation of the distribution of income. This is relevant since it suggests that the mismatches between income and multidimensional poverty are not due simply to discrepancies in the levels of income and/or the poverty line used. This result also suggests that the potential effects of income underreport in the Gallup survey, as reported in Gasparini and Gluzman (2012), should be small. Artificially shifting the distribution of incomes to the right (or moving the poverty line to the left) to compensate for misreported income, leaves the classification performance unaltered. Naturally, this effect may alter other functionals of the distribution of income (like inequality) which rely on use of the whole distribution of income, unlike poverty analysis which focuses on its left tail.

Panel B in Table 4 offers another perspective. It shows the proportion of individuals in each income decile that belong to the "poor cluster" group. For example, 54 percent of those in the first income decile are classified as poor. This

TABLE 5
Country Comparison

Country	Cluster Poor	Income Poor
Honduras	47.89%	23.00%
Peru	44.50%	57.80%
Nicaragua	41.88%	59.50%
Paraguay	39.65%	54.9%
Bolivia	38.16%	58.80%
El Salvador	34.29%	60.50%
Ecuador	29.03%	45.80%
Uruguay	28.71%	25.60%
Chile	27.58%	22.00%
Panama	24.64%	32.60%
Colombia	21.77%	35.84%
Argentina	21.10%	22.90%
Costa Rica	20.97%	25.40%
Guatemala	16.89%	50.30%

proportion decreases monotonically with income, to the point that only 3 percent of those in the 10th decile are classified as poor by the cluster method. This result is relevant since it suggests, again, that even though income plays a relevant role in the cluster based multivariate notion of poverty, the relationship is rather weak, especially in low levels of income. In other words, though more income reduces monotonically the chances of falling in the poor cluster, low income is not necessary or sufficient to explain the multivariate version of the poverty status, to the point that, for example, 46 percent of the individuals in the lowest decile are not rendered as poor by the cluster approach. This result, again, is compatible with the large literature that points toward the inadequacy of income as the only factor to identify the poor.

Table 5 explores similarities by country, that is, after implementing the procedure in the original database, we have computed cluster and income poor groups. As expected, the relationship between the two classifications is positive but weak. The cases of Honduras and Guatemala are interesting. Honduras has the higher proportion of cluster based poor, even though in terms of income, it ranks relatively in the bottom. Exactly the opposite occurs in the case of Guatemala. Uruguay and Argentina are cases where the aggregate figures match; for example, in the latter, the cluster poor is 21 percent compared to 22.9 percent based on income.

It is relevant to extend the comparison to other multidimensional methods. The task is not easy, since a particular advantage of our cluster method is that it does not require the ex-ante explicitation of a poverty line. On the other hand, the recent approach of Alkire and Foster (2011) requires, as a starting point, the existence of a threshold in each of the dimensions where welfare is measured, and with respect to which, poverty is computed. This is a complicated task when several variables in the welfare space are of a binary nature, so the only natural threshold is whether a person or household either possesses or not a particular characteristic. Additionally, and to complicate matters, thresholds are particularly difficult to establish for variables measuring subjective welfare, like the sort used in our empirical analysis.

TABLE 6

Comparison with Alkire and Foster (2011)

Number of Deprived Dimensions	Correct Classification
1	53.58%
2	79.58%
3	78.92%

Note: Percentage of individuals correctly classified as poor or non-poor by cluster analysis and Alkire and Foster (2011) method.

Nevertheless, and in order to compare our cluster results with other multidimensional perspectives, we carried out the following exercise. In order to deal with the problem of defining thresholds and with that of dealing with discrete variables, we implemented an analysis along the lines of Alkire and Foster (2011) using the optimal factors that are the output of Gasparini *et al*. (2011). Factors are by construction continuous random variables. Any choice of threshold involves a certain degree of arbitrariness. The problem is magnified in our case, that involves subjective and categorical variables. Consequently, and to emphasize comparison, we have followed the approach in Gasparini *et al*. (2011) strictly, by taking their thresholds for each factor that produces a share of the LAC population below the threshold that coincides with the income poverty headcount ratio based on the 2 dollars a day line (39.9 percent).

Table 6 presents the results of this exercise. Following Alkire and Foster (2011), we classified individuals as poor/non-poor when they are below the previously defined thresholds in one, two, and three dimensions. In each row we report the percentage of individuals in our sample that are classified the same using the counting method and our cluster-based method. For example, when individuals are considered as poor if they are below the poverty threshold of one dimension (no matter which one), the percentage of individuals whose poverty status matches exactly with the one produced by the cluster method is 53.58.

In spite of the simplicity of this exercise, several results are relevant. First, there is a sudden increase in the accordance of both methods when moving from one to two dimensions. Second, the coincidences between the counting and cluster method are maximized when only two dimensions are considered. From this perspective, the cluster method performs as an intermediate case, more stringent than the union method with one dimension, and less restrictive than the intersection method. Third, although similarities are high and remarkable, both methods differ, highlighting, once again, the discrepancies between approaches and the complexity of computing multidimensional poverty. Finally, we stress the fact that the methods are not directly comparable without relying on previous agreements on the welfare space and on the poverty thresholds that are needed to compute the counting approach. A natural characteristic of cluster methods is that they do not require the ex-ante explication of thresholds in order to produce a poor/non-poor classification.

A final discussion refers to use of cluster methods vis-a-vis other strategies. The classical approach (unit- or multi-dimensional) to finding the poor is to rely on

exogenously defined thresholds, as in the initial step of Alkire and Foster (2011) discussed above. For the purposes of this discussion, such strategy will be labeled as defining the poor *exogenously*. As stressed before, a natural advantage of cluster methods is that what they take as exogenous is the notion of the poor as a group, and obtain (multivariate) poverty thresholds endogenously, as a by-product of the problem of optimal classification. Hence, from this perspective cluster methods produce an *endogenous* partition of the welfare space.

In this paper, cluster mechanisms are used for two purposes. The first one it to find the poor in a multidimensional framework, and the second one is to estimate the dimension of the welfare space through a variable-selection procedure that tries to reproduce the initial classification as accurately as possible using a significantly reduced variable space. As expressed before, this strategy requires the validation of the poor as a group, by checking statistically and economically that its members are indeed deprived.

Naturally, a *hybrid* approach may be exploited in situations where there is a well established notion of multivariate poverty or in cases where more than two groups exist. That is, in a first stage, *any* classification mechanism is used to define the poor and the non-poor, like for example the output of Alkire and Foster (2011), or a hierarchical partition method. Then, the blinding process is applied to this initial classification, for the purpose of finding a reduced set of variables that is able to reproduce the initial (now not necessarily endogenous) classification. In the language of recent machine learning methods, a fully cluster-based approach in the two stages (finding the poor and reducing dimensionality) is an *unsupervised* learning strategy, where the poor/non-poor arise as a result of a within similarity/between dissimilarity trade-off. The hybrid proposal implies a first stage where the poor are found exogenously, and this resulting classification is used in *supervised* stage to reduce the dimension of the welfare space. See Murphy (2013, ch. 25) for a recent discussion.[3]

5. Conclusion

The fact that welfare is progressively accepted as an essentially multidimensional notion implies many conceptual and practical challenges, which usually suggest a trade-off related to the desired degree of aggregation. On the one hand, and for pragmatic and conceptual reasons, it seems reasonable to attempt to summarize welfare in a few readily available indexes that can help monitor social performance as well as implement valid comparisons. On the other hand, the complex nature of well-being points toward retaining as many factors as possible in order to fully characterize it. In this context, this paper suggests a simple procedure that (1) treats the poor as a coherent, clearly identifiable group that can be economically and statistically distinguished from its complement, (2) fully exploits available information to detect it, and (3) summarizes the initial welfare space into a few unambiguously interpretable variables.

The empirical implementation based on the Gallup Poll suggests that three variables can reproduce quite accurately the role of the original 15 variables in the

[3]We thank an anonymous referee for bringing out this important point.

goal of identifying the poor. From a practical perspective, once this "cluster poor" group of individuals is successfully identified using a large data set, further classification or evaluations can be implemented by assessing just the variables in the reduced set.

From a methodological perspective, the use of multivariate methods in Economics is scarce, which is surprising in light of the massive acceptance these techniques have in closely related areas. For this reason we have tried to stay as close as possible to standard grouping techniques, relegating more modern and sophisticated approaches (like CART methods as in Keely and Tan, 2008) to further research.

References

Alkire, S., "Choosing Dimensions: The Capability Approach and Multidimensional Poverty," in Kakwani, N. and J. Silber (eds), *The Many Dimensions of Poverty*, Palgrave Macmillan, New York, 89–119, 2007.

Alkire, S. and J. Foster, "Counting and Multidimensional Poverty Measurement," *Journal of Public Economics*, 95, 476–87, 2011.

Bishop, C. M., *Pattern Recognition and Machine Learning*, Springer, New York, 2006.

Calinski, R. B. and J. Harabasz, "A Dendrine Method for Cluster Analysis," *Communications in Statistics*, 3, 1–27, 1974.

Cherkassky, V. and F. M. Mulier, *Learning from Data: Concepts, Theory and Methods*, 2nd ed, Wiley, New York, 2007.

Coates, A., H. Lee, and A. Y. Ng, "An Analysis of Single-Layer Networks in Unsupervised Feature Learning," *Proceedings of 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

Cuesta-Albertos, J. A., R. Fraiman, and T. Ransford, "Random Projections and Goodness-of-Fit Tests in Infinite-Dimensional Spaces," *Bulletin of Brazilian Mathematical Society*, New Series 37, 477–501, 2006.

Duclos, J. Y., D. E. Sahn, and S. D. Younger, "Partial Multidimensional Inquality Orderings," *Journal of Public Economics*, 95, 225–38, 2011.

Elffers, H., J. Bethlehem, and R. Gill, "Indeterminacy Problems and the Interpretation of Factor Analysis Results," *Statistica Neerlandica*, 32, 181–99, 1978.

Ferro Luzzi, G., Y. Fluckiger, and S. Weber, "A Cluster Analysis of Multidimensional Poverty in Switzerland," in Kakwani, N. and J. Silber (eds), *Quantitative Aproaches to Multidimensional Poverty Measurement*, Palgrave Macmillan, New York, 2008.

Fraiman, R., A. Justel, and M. Svarc, "Selection of Variables for Cluster Analysis and Classification Rules," *Journal of the American Statistical Association*, 103, 1294–303, 2008.

Gasparini, L. and P. Gluzman, "Estimating Income Poverty and Inequality from the Gallup World Poll: The Case of Latin America and the Caribbean," *Journal of Income Distribution*, 21, 3–27, 2012.

Gasparini, L., W. Sosa Escudero, M. Marchionni, and S. Olivieri, "Multidimensional Poverty in Latin America and the Caribbean: New Evidence from the Gallup World Poll," *Journal of Economic Inequality*, 1–20, 2011.

Graham, C. and J. Behrman, "How Latin Americans Assess Their Quality of Life: Insights and Puzzles from Novel Metrics of Well-Being," in Graham, C. and E. Lora (eds), *Paradox and Perception: Measuring Quality of Life in Latin America*, Brookings Institution Press, Washington, 1–19, 2009.

Graham, C. and E. Lora, *Paradox and Perception: Measuring Quality of Life in Latin America*, Brookings Institution Press, Washington, 2009.

Hand, D., H. Mannila, and P. Smyth, *Principles of Data Mining*, MIT Press, Cambridge, 2001.

Hardle, W. and L. Simar, *Applied Multivariate Statistical Analysis*, Springer, New York, 2003.

Hartigan, J. A., "Asymptotic Distributions for Clustering Criteria," *Annals of Statistics*, 6, 117–31, 1978.

Kakwani, N. and J. Silber, *Quantitative Approaches to Multidimensional Poverty Measurement*, Palgrave Macmillan, New York, 2008.

Keely, L. and C. Tan, "Understanding Preferences for Income Redistribution," *Journal of Public Economics*, 92, 944–61, 2008.

MacQueen, J. B., "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, 281–97, 1967.
Murphy, K., *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, MA, 2013.
Opazo, L., C. Raddatz, and S. Smuckler, "The Long and the Short of Emerging Market Debt," World Bank Working Paper, 2009.
Pollard, D., "Strong Consistency of K-Means Clustering," *Annals of Statistics*, 9, 135–40, 1979.
Raftery, A. E. and N. Dean, "Variable Selection for Model-Based Clustering," *Journal of American Statistical Association*, 101, 168–78, 2006.
Ravallion, M. and M. Lokshin, "Self-Rated Economic Welfare in Russia," *European Economic Review*, 46, 1453–73, 2002.
Sen, A., *Commodities and Capabilities*, Oxford University Press, Oxford, 1985.
Sheskin, D., *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th ed, Champan-Hall, New York, 2011.
Steinly, D. and M. J. Brusco, "Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques," *Journal of Classification*, 24, 99–121, 2007.
Tadesse, M. G., N. Sha, and M. Vannucci, "Bayesian Variable Selection in Clustering High-Dimensional Data," *Journal of the American Statistical Association*, 100, 602–17, 2005.
Tibshirani, R., G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *Journal of the Royal Statistical Society, Serie B (Statistical Metodology)*, 63, 411–23, 2001.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix 1: Variable description
Appendix 2: The Fraiman, Justel and Svarc (2008) algorithm