Measuring Multidimensional Deprivation: A latent utility approach

Carlos Felipe Balcázar and Silvia Redaelli *

April, 2017

Abstract

We propose a latent utility model to measure multidimensional deprivation. We provide formal proof that if discrete ordinal data meets the ordering consistency conditions defined here, wealth rankings can be gleaned from the eigenvector corresponding to the largest eigenvalue in Multiple Correspondence Analysis (MCA). This approach outperforms Principal Component Analysis (PCA), a popular dimensionality reduction technique used to construct wealth indexes. We provide evidence using Monte Carlo experiments that MCA explains 1.7 times more variation than PCA and does a better job at predicting wealth-rankings. We define welfare consistent cut-off points for the deprivation index. Lastly, we provide an example to estimate housing deprivation headcounts using Afghanistan's Living Conditions Survey.

*The World Bank Group, 1818 H St. NW, Washington D.C. United States of America. The findings herein do not necessarily represent the views of The World Bank Group, or its Boards of Directors. We thank Hai-Anh Dang, Dean Joliffe, Hugo Ñopo, Jaime Ramirez, Javier Rojas, Daniel Valderrama, Tara Vishwanath, participants at Western Economic Association International's 13th international conference in Santiago de Chile, and participants at the LAC Information Sharing Lunch at the World Bank, for their useful comments. Corresponding author at cfbalcazars@gmail.com

1 Introduction

Conceptualizing deprivation measurement involves two steps: First, aggregation of data into one index. Second, identification of the destitute. (Tsui, 2002.) The first step requires defining the weighting scheme to aggregate across different goods or services. The second one allows identifying who is deprived according to an established norm. In this paper we propose a methodology to endogenize the calculation of weights, capturing ordinal information on a latent utility function. We also define welfare consistent cut-offs for the deprivation index, so that individuals below it can be considered multidimensionally deprived.

The problem of the aggregation process is involved, and there is no consensus on how to define the weighting scheme. Some authors attribute equal weights to different welfare dimensions in a nested fashion (Alkire et al., 2015), while other attribute different weights according to various economic criteria or normative considerations (Decancq and Lugo, 2010). Another strand of the literature exploits the numerous associations between dimensions along the lines of factorial analysis (Filmer and Pritchett, 2001; Kolenikov and Angeles, 2009; Asselin, 2009), but the normative basis for such indexes have not yet been established.

One major concern in the multidimensional deprivation measurement literature is that, different weighting schemes could result in very different societal indicators and orders (Decancq and Lugo, 2010; Yalonetzky, 2013, 2014). Furthermore, aggregation across deprivations cannot (in general) yield deprivation measures that are *welfare consistent* (Ravallion, 2011).

These issues are important. The design of meaningful measures of deprivation calls for clear axiomatic basis (Ravallion, 2012), and for a clear understanding of the implications of using different types of data and statistical methodologies (Kolenikov and Angeles, 2009). In this paper we focus on discrete ordinal data—fairly common in empirical applications—and address these concerns to an extent.

To ease exposition, we will label any ordinal variable an *indicator* and its values, *items*. We depart from three conditions that are consistent with welfare theory: i) The probability of owning a given item increases with the utility level, ii) The probability of owning a high-ordered item is lower than the probability of owning a lower-ordered item, given a utility level iii) The probability of owning a basket of indicators with a high-ordered item is lower than the probability of owning a basket of indicators with a lower-ordered item, given a utility level.

Under the conditions previously defined, we provide formal proof that the Multiple Correspondence Analysis (MCA) algorithm captures the ordinal information along the latent utility distribution. In particular, we show that wealth-rankings can be obtained from the eigenvector corresponding to the largest eigenvalue in MCA. In other words, factor scores derived from such eigenvector capture the latent ordinal information in the data.

Monte Carlo experiments reveal that MCA does a good job at predicting (latent) utility rankings. Spearman rank correlations, which evaluate the strength of the monotonic relation between MCA factor scores and a simulated vector of endowments, are on average 0.87, indicating a strong monotonic relation between the two.

We compare MCA to the traditional Principal Component Analysis (PCA) approach to construct wealth indexes (Filmer and Pritchett, 2001), showing that the latter does not do a good job at modelling latent utility. Spearman correlations between PCA factor scores and simulated endowments are on average 0.65, indicating a much weaker monotonic relation vis-á-vis MCA. Moreover, MCA does as well a better job at reducing the dimensionality of the data as measured by the amount of variation explained by the eigenvector corresponding to the largest eigenvalue: the former accounts for 72% of total variation in our simulations while the latter accounts for 42%.

These differences owe to the fact that PCA was originally developed for multivariate normal data, and is best used with continuous data; when data are discrete multivariate normality is clearly violated (Kolenikov and Angeles, 2009). In contrast, MCA does not require discrete variables to follow any underlying distribution (Greenacre, 2006).

Our results have other substantial implications as PCA has become a technique commonly used to created asset indexes as a measure of socio-economic status. One prominent example of this are asset-indexes in Demographic Household Surveys. It appears, on the basis of the evidence provided here, that PCA is not well suited for constructing meaningful asset indexes.

We also define welfare consistent cut-offs for estimating multidimensional deprivation headcounts, and argue that cut-offs must be defined on a normative basis for each indicator included in the analysis. Moreover, given that MCA does a good job at predicting wealth rankings for simulated economies with different prices, headcounts gleaned from this procedure are informative for policy analysis.

Our findings contribute to the multidimensional deprivation literature in a number of ways: First, we show that MCA can provide meaningful ordinal information, gleaned from discrete ordinal data. Second, we show that PCA does not do a good job at modelling latent utility. Third, We provide an alternative to compute deprivation measures that can satisfy welfare consistency.

As an empirical exercise, we use the methodology presented here to assess housing deprivation in Afghanistan, using the Living Conditions Survey (ALCS) 2013/14. We define housing adequacy using United Nation's Right to Adequate Housing (UN, 2014a). However, since housing markets are usually not well-developed in rural areas and since nomadic communities such as the Kuchis do not have a clear preference for adequate housing, we restrict our analysis only to urban areas.

First of all, we show that MCA outperforms PCA substantially, with the capacity of explaining more than three times the amount of variation explained by the latter. Then, after establishing an appropriate cut-off based on the literature on housing development and human rights, we find that 6 out of 10 households can be considered housing-deprived. We also find striking regional differences and (as expected) higher incidence of deprivation for monetary-poor households.

The rest of the paper proceeds as follows. In the next section we present the latent utility model, define its properties, describe the methodological approach and present the results from our simulations. In Section 3 we discuss how to establish cut-offs and compute multidimensional deprivation headcounts. In section 4 we introduce the problem of measuring housing deprivation in Afghanistan. In Section 5 we present the results of the empirical exercise. The last section concludes.

2 Methodological framework

Consider the multidimensional space $N \times K$ where N denotes the populations size (or number of rows) and K denotes the number of deprivation indicators (or number of columns). Let k, with $k = \{1, 2, ..., K\}$, denote any indicator of M_k ordered categories or items. Items are ordered from worst to best. Then let $X(N, M)$ be the indicator matrix of N rows and M binary items, with $M = \sum_k M_k$. Thus X consists of M binary $(0, 1)$ vectors of length N. The M vectors will be indexed by i and j.

Now, let U denote a one-dimensional latent (utility) variable and let $L(U)$ be its probability density function. Let also $p_i(U)$ denote the response function for item i. Thus the unconditional probability of having attribute i is given by

$$p_i = \int_{-\infty}^{\infty} p_i(U) dL(U). \tag{1}$$

Next, we define three requirements for the latent utility model:

i) $p_i(U)$ is monotonically increasing on U in every indicator. That is, $p_i(U_1) \leq p_i(U_2)$ for $U_1 < U_2$ for every $i \in k$. In other words, the probability of owning a given item increases with the utility level.

ii) The items of an indicator can be ordered such that the levels of U are not intersecting: $p_i(U) \geq p_j(U)$ for $i < j$, with $i, j \in k$. So, the probability of owning a high-ordered item is lower than the probability of owning a lower-ordered item, given a utility level.

iii) The joint probability of two items $i \in k$ and $j \in l$ with $k \neq l$ for a value of U is given by $p_i(U) p_{ij}(U)$, and the corresponding conditional probability is $p_{ij} = \int_{-\infty}^{\infty} p_i(U) p_{ij}(U) dL(U)$.

If there is a single latent variable, and (i) and (ii) hold, then items in k can be ordered such that $p_i > p_j$ for $i < j$, with $i, j \in k$. Note also that the joint probability of two items $i, j \in k$ for a value of U is zero, whereas $p_{ij} \in (0, 1]$ if $i \in k$ and $j \in l$ with $k \neq l$. If (iii) also holds, then

$p_i(U)p_{ij}(U) - p_{i'}(U)p_{i'j}(U) > 0$, $i < i'$, $i, i' \in k$ and $j \in l$ with $k \neq l$. Thus, the probability of owning a basket of indicators with a high-ordered item is lower than the probability of owning a basket of indicators with a lower-ordered item, given a utility level.

Having the previous properties in mind, let us define the squared $M \times M$ matrix

$$P = [p_{ij}]. \tag{2}$$

2.1 The Multiple Correspondence Analysis Algorithm

Matrix P is inconvenient inasmuch not all M_k are equal; it is well known that indicators with a higher M_k will be given more weight by the eigendecomposition algorithm.[1] To address this issue let us consider that

$$P = \frac{X^T X}{N} = \frac{B}{N},$$

where B denotes the associated Burt Matrix. Therefore, we can use the correspondence matrix $C = \frac{N}{n}(P) = \frac{B}{n}$ where $n = \sum_{i,j} b_{ij}$, instead of P, as it results in the same eigenvector.[2]

We can standardize C by means of calculating the standardized residuals matrix

$$S = [s_{ij}] = [(c_{ij} - r_i r_j)/\sqrt{r_i r_j}], \tag{3}$$

where $r_i = \sum_i c_{i*}$ and $r_j = \sum_j c_{*j}$ are the row and column totals (which are the same since S is a symmetric squared matrix). Now, consider the following direct adaptation of Warrens and Raadt (2013) Theorem 3:

Proposition 1. *Suppose that M_k of the M vector, which without loss of generality can be taken as the first M_k, can be ordered such that (i) and (ii) hold. Then the elements of ν of S corresponding to these vectors satisfy $\nu_1 > \nu_2 > ... > \nu_{M_k} \geq 0$.*

[1] An eigendecomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors: $P = V \Lambda V^T$.

[2] Since $PV = V\Lambda$, then $P(\frac{N}{n}V) = \frac{N}{n}PV = \frac{N}{n}V\Lambda = V(\frac{N}{n}\Lambda)$.

Proof. In the Appendix. □

And finally consider the following corollary:

Collorary 1. *If the elements of M can be ordered such that (i), (ii) and (iii) hold, and* $\nu_i > \nu_{i'}$, $i, i' \in k$, *then the elements of* ν *of S satisfy* $\nu_i + \nu_j > \nu_{i'} + \nu_j$, $i < i'$, *with* $j \in l$ *and* $k \neq l$.

Proof. In the Appendix. □

Proposition 1 and Collorary 1 indicate that ordinal information on latent variable models can be obtained from the eigenvector corresponding to the largest eigenvalue. More specifically, this proposition indicates that if items are ordered for every indicator, then the magnitude of weights across items is consistent with the former ordering; we refer to this as *first ordering consistency*. Corollary 1 indicates that if items are ordered within indicators and we know valuations across bundles, then weights are consistent with the global ordering of individuals; we refer to this as *global ordering consistency*.[3]

Note that if first ordering consistency and global ordering consistency are satisfied, then it follows that if any individual improves his/her situation in relation to one of the indicators, latent utility improves; we refer to this as *composite deprivation ordering consistency*.[4] It is thus possible to uncover meaningful orderings from categorical data using eigenvectors.

2.2 The multidimensional deprivation index

Denote the spectral decomposition of S as

$$S = V \Lambda V^T, \quad (4)$$

[3]These two labels: *first ordering consistency* and *global ordering consistency*, are borrowed from Asselin (2009).
[4]This label is also borrowed from Asselin (2009).

where V is the eigenvectors matrix, and Λ is the diagonal matrix whose diagonal elements are the eigenvalues. Then, consider the $N \times M$ matrix of factor scores

$$F = XV. \tag{5}$$

Now, let us define the composite index D_i, with $i = \{1, 2, ..., N\}$, simply as

$$D_i = F_{i,\nu}, \tag{6}$$

where $F_{i,\nu}$ denotes the value of the deprivation index for individual i, with ν the eigenvector corresponding to the largest eigenvalue. D is then the multidimensional deprivation index.[5]

2.3 Percentage of explained variation

The percentage of inertia (i.e., variation) explained by the first eigenvalue evaluates how well the MCA algorithm reduces the data's dimensionality. Since MCA codes data by creating several binary columns, it creates artificial additional dimensions, the inertia of the solution space is artificially inflated and therefore the percentage of inertia explained by the first eigenvalue is severely underestimated (Abdi and Valentin, 2007). The "percentage of inertia problem" can be addressed by using adjusted inertias (Greenacre, 1993):

$$\lambda_s^{adj} = \left(\frac{K}{K-1}\right)^2 \left(\lambda_s - \frac{1}{K}\right)^2, \tag{7}$$

with $s = \{1, 2, ..., M\}$ such that $\lambda_1 > \lambda_2 > ... > \lambda_M$. However, the adjusted inertias are calculated only for each eigenvalue that satisfies the inequality $\lambda_s \leq 1/K$ (Greenacre, 1993).

Traditionally, the percentages of inertia are computed by dividing each eigenvalue by the sum of the eigenvalues, however this will give an pessimistic estimation of the percentage of inertia. Greenacre (1993) suggests instead to evaluate the percentage of inertia relative to the average

[5] Note that by Proposition 1, lower values of D are preferred to higher values of it.

inertia of the off-diagonal blocks of the Burt matrix

$$\frac{K}{K-1}\left(\sum_s \lambda_s^2 - \frac{M-K}{K^2}\right).$$

2.4 Monte Carlo Experiments

Let us consider the following utility maximization problem

$$\max \prod_{j=1}^{K} k_j^{\alpha_j} \qquad \textit{so that} \qquad 0 \leq k,\ p \cdot k \leq I,$$

where I denotes endowments, p is the vector of prices and k is the vector of indicators (goods ands services). The marshallian demand function for any k is given by

$$k_j^* = \frac{\alpha_j}{\sum_{j=1}^{K} \alpha_j}\left(\frac{I}{p_j}\right).$$

Without loss of generality, for our simulations, we assume I follows a log normal distribution, such that $\ln(I) \sim \mathcal{N}(\mu_I, \sigma_I^2)$;[6] α are randomly assigned as well using a uniform distribution $\alpha \sim \mathcal{U}(0,1)$ so that $\sum_{j=1}^{K} \alpha_j = 1$; prices are also randomly assigned using a normal distribution $p \sim \mathcal{U}(a,b)$ with $a > 0$ so that $p_j > 0$. By providing these parameters we can compute the marshallian demands for every good, and obtain their distribution, which by construction is also log-normal $\ln(k_j^*) \sim \mathcal{N}(\mu_{k_j^*}, \sigma_{k_j^*}^2)$. Note that we assume all individuals have the same utility function.

We obtain ten thousand simulations with one thousand individuals each, making sure they comply adequately with ordering consistency. After obtaining the distributions for every k_j^*, we discretize them: $k_j^{*\prime} = \lfloor k_j^* \rceil$. We use $k_j^{*\prime}$ to perform MCA and PCA as well (as means of providing comparisons to a commonly used method).[7] Then we obtain the factors scores associated to the eigenvector with the largest eigenvalue.

First of all, we observe that MCA does a much better job than PCA at reducing dimensionality

[6]This obeys the fact that most distributions of wealth are skewed with heavy left tails.

[7]PCA performs the eigendecomposition on the correlations matrix instead of the Burt matrix. See Filmer and Pritchett (2001) for a discussion.

(Figure 1). On average, MCA explains 1.7 times more variation than PCA. Secondly, we check whether MCA or PCA factors scores have a strong monotonic relationship with the vector of endowments using the Spearman rank correlation coefficient and Kendall's τ.[8] We find that while PCA does not do a good job at predicting wealth-rankings, MCA performs well on average on this regard as measured by Spearman and Kendall's τ coefficients (Figure 2).

[FIGURE 1 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

3 Cut-offs and headcounts

Let z_k be the deprivation threshold for indicator k, such that if $M_{ik} < z_k$, i is deprived in the indicator k alone, but is not multidimensionally deprived. z_k should be defined normatively on the basis of the deprivation problem that is being assessed. For example, if indicator k corresponds to the type of floor, which is composed by three categories: dirt, wood or tile, we can use wood (assuming that tile is better) to define the threshold for this good since we can consider dirt floor as being floor-deprived.[9]

We define the deprivation threshold by considering all thresholds z_k so that

$$z = D(M_k) \text{ with } M_k = z_k \text{ for all } k. \tag{8}$$

If $D_i > z$, then individual i is multidimensionally deprived. Thus the multidimensional deprivation rate

$$R = \frac{1}{N} \sum_{i \in Q} 1, \tag{9}$$

where $Q = \{i : D > z\}$ is the set of multidimensionally deprived individuals.

[8] The Spearman a rank correlation coefficient and Kendall's τ measure the ordinal association (or rankings) between two variables, and can be used to assess the significance of the relation between them.

[9] We know from the development literature that having dirt floor is related to poor health (Cattaneo et al., 2009).

The deprivation bundle $\{z_1, z_2, ..., z_K\}$ is consistent with the choices made by someone living at the multidimensional deprivation line, in the sense that if someone living at the deprivation line becomes worse (better) off then measured deprivation rises (falls). Indeed, we showed using Montecarlo experiments that MCA does a good job at predicting wealth rankings. Thus, it does a good job at discriminating the indirect utility level given a vector of outcomes. In this sense, individuals with a lower ranking than another with bundle $\{z_1, z_2, ..., z_K\}$, can be considered poor (non-poor otherwise). Hence, the magnitude of weights is irrelevant inasmuch utility is an ordinal function. We are concerned about the utility rankings.

To illustrate this, assume that there are two indicators k_1 and k_2 without loss of generality, with market prices p_1 and p_2, so that $y = p_1 k_1 + p_2 k_2$;[10] y corresponds to the aggregation index. Let us define z the deprivation line and F_y the cumulative distribution of y, so that $P^a = F_y(z)$ is the deprivation headcount. Then consider a composite index obtained from MCA: $D = v_1 k_1 + v_2 k_2$, so that we define define the deprivation aggregation headcount $P^d = F_D(z)$, where z is defined as $z = D(z_1, z_2)$. P^a and P^d will give the same results when percentile ranks are equivalent for both distributions y and D. In such case, we can think of z as the point on the inverse of the latent utility function corresponding to the deprivation level of utility. Then any exogenous welfare-reducing (increasing) change will be poverty increasing (decreasing), and welfare consistency is assured.

4 Housing deprivation

The international human rights law recognizes everyone's right to adequate housing. Adequate housing was recognized as part of the right to an adequate standard of living in the 1948 Universal Declaration of Human Rights and in the 1966 International Covenant on Economic, Social and Cultural Rights. Since then, other international human rights treaties have since recognized or referred to the right to adequate housing or some elements of it (UN, 2014a).

The right to adequate housing contains freedoms, such as the right to choose one's residence

[10]With equality because we are assuming monotone preferences.

and protection against forced evictions. It contains entitlements, such as security of tenure, land and property restitution and equal and non-discriminatory access to adequate housing. It contains parameters, such as having access to an adequate and enclosed space (i.e., four walls, roof, floor and enough physical space to avoid overcrowding), to safe drinking water, adequate sanitation, energy for cooking, heating, lighting, cooking facilities and to local services, such health, education, childcare and other social facilities.

Not guaranteeing the right to adequate housing has major implications for people's welfare. This is because human rights are interdependent, indivisible and interrelated; the violation of the right to adequate housing may preclude people from enjoying other human rights and vice versa. Access to adequate housing can be a precondition for enjoying access to work, public services, health and social opportunities (Duncan, 2009). As a matter of fact, the link between inadequate housing and poor health is well established (Baggott, 2010): inadequate housing has been linked to increased risks of respiratory infection, cardiovascular conditions, allergies, important medical skin problems like eczema, exposure to hazardous agents and adverse psychological health (Harvey and Blackman, 2001; BMA, 2003; Evans et al., 2003; Parry et al., 2004; Blackman, 2006; Shelter, 2006; Barnes et al., 2008). On another hand, vulnerable and disenfranchised communities, such as those living in slums, often live in areas near or on steep slopes, riverbanks, flood plains and by garbage dumps or other hazardous waste sites, in flimsy structures vulnerable to intrusion or destruction by wind, rains, landslides and floods. Thus it comes as no surprise that those who lack adequate housing are forced to spend more money and time on shelter rather than on other basic needs in comparison to the non-destitute, further entrenching them in poverty (Duncan, 2009).

An adequate house is an important asset in which individuals' lives are often shaped, playing an important role as basis for important social support activities that underpin wellbeing (Bratt, 2002). It stands to reason then, that households should have access to adequate housing.

4.1 Housing deprivation in Afghanistan

For our empirical exercise we choose Afghanistan living conditions survey 2013/14. We choose Afghanistan out of the lack of information on housing adequacy for this country. Analyses on housing in Afghanistan gleaned from household surveys so far explore each component of housing independently, but they do not provide a more comprehensive view of housing adequacy (Central Statistics Organization, 2016).

We restrict ourselves to urban areas, given that in rural areas housing market dynamics are substantially different. In this sense, we consider that deprivation may not be solely a problem of demand if it is unlikely to find dwellings with desirable characteristics in the local housing market. For example, some rural areas may have only mud houses with not access to basic services to offer. We also consider the dynamics of idiosyncratic communities: In the case of the nomadic Kuchis for example, it may not be possible to find mechanisms to guarantee adequate housing for them, and even the need for adequate housing in such communities can be debatable given their itinerant nature. Thus we exclude them altogether.

It is also important to note that there is no information that can be used to measure accessibility, cultural adequacy and affordability; although we can identify variables that allow us to measure security of tenure, access to services, infrastructure of the dwelling, habitability and location (albeit imperfectly due to the lack of comprehensive information). However, there is no survey that encompasses all the dimensions in the Right to Adequate Housing—which is a constraint that development practitioners have to work with.

Regarding the structural aspects of housing, we explore the characteristics of roofs, walls, floor and kitchen. For access to services, we consider access to water, electricity and sanitation. In the case of habitability, we restrict the dimension of habitability to an indicator for overcrowding, defined as a dummy that takes the value of one if each pair of same-sex individuals residing in the dwelling have a bedroom, zero otherwise.[11] For security of tenure we explore two dimensions: i)

[11]For example, if there are 4 individuals in the dwelling: two men and two women, and there are less than two bedrooms in the dwelling, the household is overcrowded; similarly, if there are two men and three women, and there are less than three bedrooms in the dwelling, the household is overcrowded.

the type of dwelling and ii) the occupancy status of the dwelling. The type of dwelling allows us to identify if the household lives in a shelter, shared house or single family house; the occupancy of the dwelling inquires directly about the security of tenure. Table 1 lists the dimensions, the harmonized ordinal variables and includes a description of each.[12]

[TABLE 1 ABOUT HERE]

Table A1 in the appendix shows the distribution of households for each indicator. We observe high prevalence of mud and mud brick houses, of dirt/earth floor, high incidence of overcrowding and high rates of access to electricity and water, and a comparatively lower rate of access to improved sanitation. The descriptive statistics also show that poor people fare worse off than their non-destitute counterpart.[13]

There is no guarantee that using all housing attributes lead to satisfying ordering consistency requirements. In fact, that depends on the structure of the Burt matrix. While for first ordering consistency we can make sure the categories in our indicators are ordered from worst to best, the problem with global ordering consistency is that it is hard to check beforehand. One way to address this problem is by exploring the correlations between indicators in the polychoric sense (Lee et al., 1995). High, positive correlations between the ordered indicators provide a sense that bundles with higher-ordered elements have higher valuations in the composite indicator. Table A2 in the Appendix shows that there is a high correlation among most variables, but many variables are negatively correlated or display very low correlations in the polychoric sense with security of tenure and dwelling type. Hence, in order to satisfy ordering consistency requirements, we do not include security of tenure dimensions when performing Multiple Correspondence Analysis.

In order to define the deprivation threshold, we use a combination of characteristics for which households can be considered non-deprived. For this endeavor we guide ourselves from the lit-

[12]To establish the order amongst categories for each variable, we guided ourselves from the literature on housing development and human rights (Tully, 2006; Amore et al., 2013; JMP, 2015). Additionally, we consulted experts on urbanization, infrastructure and architecture at the World Bank to validate our orderings.

[13]The ALCS 2013/14 did not survey for food consumption due to the rotating module methodology. To estimate poverty rates at the national level, a survey-to-survey imputation technique was applied using the National Risk and Vulnerability Assessment (NRVA) survey 2011/12, which has consumption data (Central Statistics Organization, 2016).

erature on housing development and human rights (Tully, 2006; Amore et al., 2013; JMP, 2015), in which having access to electricity, improved water, improved sanitation and cooking facilities is a need, and having a dwelling providing safe enclosure to its inhabitants: floors other than dirt, sturdy ceiling and wall materials, and no-overcrowding, provides a strong criterion for adequate housing. Thus, we define the threshold as the value of D obtained for the following combination of dwelling attributes (Table 2):

[TABLE 2 ABOUT HERE]

5 Results

Table 3 shows the results of the algorithm; it shows the eigenvector corresponding to the largest eigenvalue. Note that weights within indicators satisfy ordering consistency requirements. Therefore the index obtained from this exercise reflects the ordinal information in the latent utility model. The first factor explains a considerable share of variation. The percentage of variation explained by the largest adjusted eigenvalue is 78.5% (46.7% unadjusted). In contrast, the percentage of variation explained by PCA is 25%.

[TABLE 3 ABOUT HERE]

Table 4 shows the results obtain from calculating the multidimensional deprivation headcounts (R) for different population groups on the basis of the defined threshold. Our results show that almost six out of ten households can be considered housing-deprived, with marked differences across the socioeconomic spectrum. Overall, housing deprivation incidence is particularly high for households with young household heads, household heads with low or no educational attainment, and for those households at the bottom of the wealth distribution. We also find regional differences, for example the north east and west central regions have strikingly high housing-deprivation rates; more populated regions, such as central and south, have rates below the mean. As expected, poor in general present much higher incidence of housing deprivation than their non-poor counterparts. Nonetheless, differences are small in the south and west central regions.

[TABLE 4 ABOUT HERE]

6 Conclusion

Measuring multidimensional deprivation is not a straightforward task. It involves making many decisions regarding the way data will be aggregated in a single meaningful measure and the way deprivation will be defined. In this document we opt for fitting a latent utility model, and seek to obtain ordinal information using MCA. We show that under ordering consistency, ordinal information on discrete ordered data can be gleaned from the eigenvector corresponding to the largest eigenvalue in MCA. Then we discuss how a cut-off, that is consistent with the choices made by someone living at the multidimensional deprivation line, can be established to estimate multidimensional deprivation headcounts. We also provide an example for our procedure using Afghanistan's Living Conditions survey, and show that results are informative about (housing) deprivation profiles.

Our results have substantial implications as well for constructing indexes of socio-economic status. They show that PCA may not be well suited for the purpose of capturing wealth, and thus provide a meaningful measure of it. This is relevant since PCA has become a technique commonly used to created asset indexes as a measure of socio-economic status.

References

Abdi, H., & Valentin, D. (2007). Multiple correspondence analysis. Encyclopedia of measurement and statistics, 651-657.

Angel, S. (2000). Housing policy matters: A global analysis. Oxford University Press.

Amore K, Viggers H, Baker, MG, & Howden-Chapman, P (2013). Severe housing deprivation: The problem and its measurement, Official Statistics Research Series, 6. Available from www.statisphere.govt.nz.

Asselin, L.M. (2009). Analysis of Multidimensional Poverty. New York, NY: Springer New York. http://doi.org/10.1007/978-1-4419-0843-8.

Alkire, S., & Foster, J. (2011). Counting and multidimensional poverty measurement. Journal of public economics, 95(7), 476-487.

Alkire, S., Foster, J., Seth, S., Santos, M. E., Roche, J. M., & Ballon, P. (2015). Multidimensional poverty measurement and analysis. Oxford University Press, USA.

Baggott, R. (2010). Public health: policy and politics. Palgrave Macmillan.

Balcazar, C. F., Olivieri, S., Cerani, L., and Ranzani, M. (2016). Rent imputation for welfare measurement: A review of methodologies and empirical findings. https://sites.google.com/site/cfbalcazars/Rentimputationforwelfaremeasurement.pdf?attredirects=0&d=1.

BMA (British Medical Association Board of Science and Education). Housing and health: building for the future. Accessed on accessed July 26 2016http://www.bma.org.uk/ap.nsf/AttachmentsByTitle/PDFhousinghealth/$FILE/Housinghealth.pdf.

Bratt, R. G. (2002). Housing and family well-being. Housing studies, 17(1), 13-26.

Cattaneo, M.D., Galiani, S., Gertler, P.J., Martinez, S. and Titiunik, R. (2009). Housing, health, and happiness. American Economic Journal: Economic Policy, 1(1), pp.75-105.

Central Statistics Organization. (2016). Afghanistan Living Conditions Survey 2013-14. National Risk and Vulnerability Assessment. Kabul, CSO.

Decancq, K., & Lugo, M. A. (2013). Weights in multidimensional indices of wellbeing: An overview. Econometric Reviews, 32(1), 7-34.

Duncan, J. (2009). A right to a decent home: mapping poverty housing in the Asia-Pacific region. Accessed on 27 July 2016 http://www.habitat.org/asiapacific/know_more/poverty_housing_in_asia_pacific_report.aspx.

Evans, G. W. (2003). The built environment and mental health. Journal of Urban Health, 80(4), 536-555.

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. Demography, 38(1), 115-132.

Greenacre, M.J. (1993). Correspondence analysis in practice. London: Academic Press.
Greenacre, M., and Blasius, J. (eds.) 2006. Multiple Correspondence Analysis and Related Methods. Boca Raton, FL: Chapman & Hall.

Harvey, T., and Blackman, J. (2001). Housing renewal and mental health: a case study. Journal of Mental Health, 10(5), 571-583.

JMP (2015). Progress on Sanitation and Driking Water. Geneva and UNICEF, New York.

Karlin, S. (1959). Positive operators. Journal of Mathematics and Mechanics, 8, 907–937.
Kolenikov, S., & Angeles, G. (2009). Socioeconomic status measurement with discrete proxy variables: Is principal component analysis a reliable answer?. Review of Income and Wealth, 55(1), 128-16

Karlin, S. (1968). Total Positivity. Stanford Univeristy Press, Stanford, Calif, USA.

Lee, S.-Y., Poon, W. Y., & Bentler, P. M. (1995). "A two-stage estimation of structural equation models with continuous and polytomous variables". British Journal of Mathematical and Statistical Psychology, 48, 339–358.

MacCluer, C. R. (2000). The many proofs and applications of Perron's theorem. Siam Review, 42(3), 487-498.

Ravallion, M. (2011). On multidimensional indices of poverty. The Journal of Economic Inequality, 9(2), 235-248.

Ravallion, M. (2012). Mashup indices of development. The World Bank Research Observer, 27(1), 1-32.

Thomson, H., Thomas, S., Sellstrom, E., & Petticrew, M. (2013). Housing improvements for health and associated socio-economic outcomes. The Cochrane Library.

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

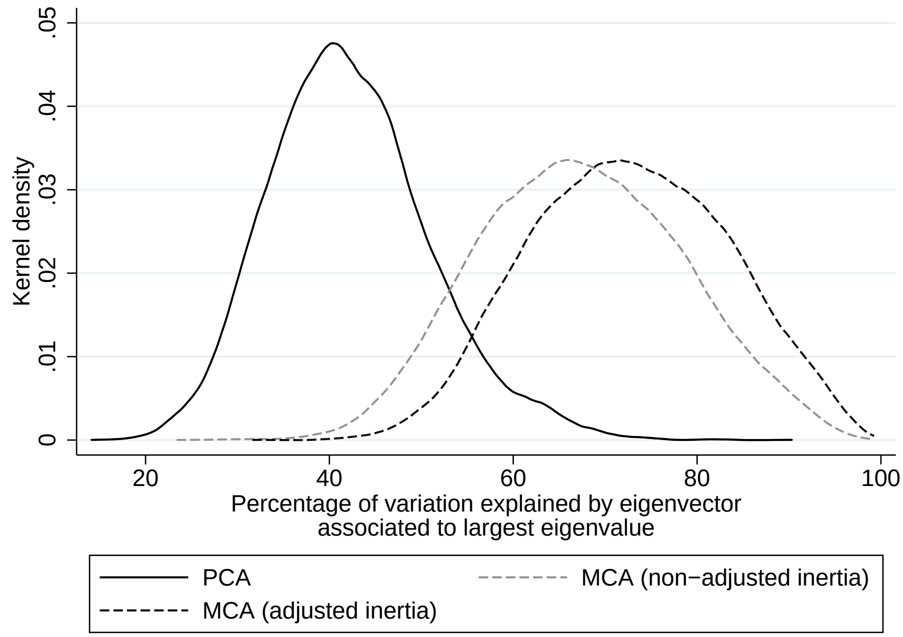
Tully, S. (2006). The human right to access electricity. The electricity journal,19(3), 30-39.

U.N. (2014a). The Right to Adequate Housing, Fact Sheet 21. Available at: `http://www.ohchr.org/Documents/Publications/FS21_rev_1_Housing_en.pdf` [accessed 10 May 2016]

U.N. (2014b). World Urbanization Prospects. Available at: `https://esa.un.org/unpd/wup/Publications/Files/WUP2014-Highlights.pdf` [accessed 10 May 2016]

Warrens, M., and Raadt, A. (2015). Ordering properties of the first eigenvector of certain similarity matrices. Journal of Mathematics, V. 2015.

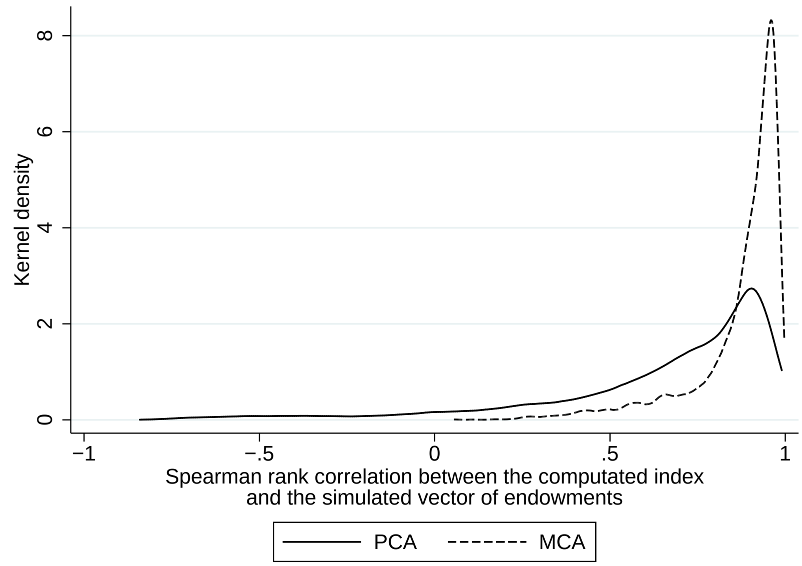
Figure 1: Percentage of explained variation



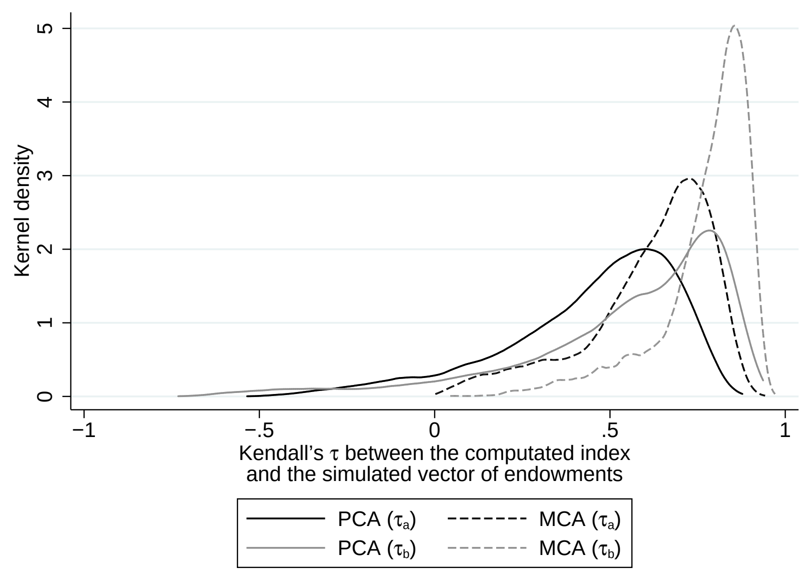
Source: Author's calculations based on ten thousand Monte Carlo simulations, with one thousand individuals each.

Figure 2: Strength of the monotonic relation between indexes and simulated wealth

(a)



(b)



Note: The rankings of the multidimensional deprivation index are defined from its largest value (the lowest rank) to its lowest value (the highest rank). τ_a will not make any adjustment for ties whereas the τ_b statistic makes adjustments for ties.
Source: Author's calculations based on ten thousand Monte Carlo simulations, with one thousand individuals each.

Table 1: Dimensions, indicators and items

Dimension	Variable	Constructed categories	Description
Infrastructure	Wall material	1. Mud, mud bricks, stone 2. Fired brick/stone 3. Concrete	Categories were harmonized and were organized from worst to best, according with the structural properties of housing materials.
	Roof material	1. Mud bricks or wood with mud 2. Girder with fired bricks or concrete	Categories were harmonized and were organized from worst to best, according with the structural properties of housing materials.
	Floor material	1. Dirth/earth 2. Concrete/tile	Categories were harmonized and were organized from worst to best.
	Kitchen	1. Cooking done in the open 2. Kitchen is part of a room inside the dwelling 3. Kitchen is in a separate room outside the dwelling 4. Kitchen is in a separate room inside the dwelling	Categories were harmonized and were organized from worst to best, according with the development literature.
Habitability	Number of rooms	1 if there is no overcrowding; 0 otherwise	The dummy the value of 1 if each pair of same-sex individuals residing in the dwelling have a bedroom.
Services	Sanitation	1 if there is a pit latrine with slab, or a pit latrine covered, or an improved pit latrine, or a flush toilet; 0 otherwise.	Categories were harmonized across surveys, and improved access to sanitation was defined on the basis of UN standards.
	Water	1 if there is piped water, hand pumped water, or a protected spring, well or karitz.	Categories wer harmonized across surveys, and improved access to water was defined on the basis of UN standards.
	Electricity	1 if there is access to electricity in the households, from any source; 0 otherwise.	A household has access to electricity if it reports having electrictiy at any time in the past month from the electric frid, generator, solar panel, wind power or a battery.
Security of tenure	Dwelling type	1. Temporary shelter/shack 2. Shared house 3. Single family house	Categories were harmonized and were organized from worst to best.
	Security of tenure	1. Charity 2. Caretaker 3. Tenant 4. Owner	Categories were harmonized and were organized from worst to best in terms of long-run security of tenure.

Source: Authors' compilations.

Table 2: Threshold categories

Variable	Threshold category
Wall material	Mud, mud bricks, stone
Roof material	Mud bricks or wood with mud
Floor material	Concrete/tile
Kitchen	Kitchen is part of a room inside the dwelling
Sanitation	Access to improved sanitation
Water	Access to improved water
Electricity	Access to electricity
No overcrowding	No overcrowding

Source: Authors' compilations.

Table 3: Eigenvector corresponding to the largest eigenvalue

Variable	Constructed categories	MCA
Wall material	1. Mud, mud bricks, stone	1.04
	2. Fired brick/stone	-1.53
	3. Concrete	-2.56
Roof material	1. Mud bricks or wood with mud	1.04
	2. Girder with fired bricks or concrete	-1.90
Floor material	1. Dirth/earth	1.29
	2. Concrete/tile	-1.54
Kitchen	1. Cooking done in the open	1.86
	2. Kitchen is part of a room inside the dwelling	0.65
	3. Kitchen is in a separate room outside the dwelling	0.31
	4. Kitchen is in a separate room inside the dwelling	-0.99
Sanitation	0. No access to improved sanitation	1.64
	1. Access to improved sanitation	-0.49
Water	0. No access to improved water	1.05
	1. Access to improved water	-0.10
Electricity	0. No access to electricity	2.28
	1. Access to electricity	-0.03
No overcrowding	0. Overcrowding	0.42
	1. No overcrowding	-0.40

Source: Authors' compilations.
Note: Percentage of inertia explained by the first component: 78.5%.

Table 4: Multidimensional deprivation headcounts

Variable (at the level of the household head)	All	Poverty status Poor	Non-poor
All	**55.03**	**74.24**	**42.58**
Age			
24 years of age or less	61.60	75.62	52.14
25-34 years of age	55.55	77.05	43.12
35-44 years of age	58.63	73.23	46.36
45-54 years of age	53.12	72.59	39.40
55-64 years of age	52.20	73.95	41.04
65 years of age or more	52.42	77.14	40.35
Gender			
Male	54.99	77.14	40.35
Female	59.92	74.21	42.52
Marital status			
Married	54.84	79.35	50.01
Formerly married	53.41	74.02	42.36
Single	63.78	80.68	41.39
Education			
No formal education	64.86	81.45	51.84
Primary	57.98	78.45	52.42
Secondary	57.38	75.99	46.00
High School	44.84	73.36	48.24
Higher education	30.97	65.19	35.49
Employment status			
Employed	53.34	71.07	38.91
Underemployed	74.53	84.71	59.58
Unemployed	61.82	81.54	47.94
Inactive	46.33	69.69	35.47
Region			
Central	47.03	68.99	32.49
South	46.44	39.33	39.27
East	73.03	85.50	63.20
Northeast	85.44	92.70	77.87
North	78.00	92.97	71.33
West	59.13	60.02	25.65
Southwest	50.18	65.14	31.24
WestCentral	97.23	99.58	95.72

Source: Authors' calculations.

Appendix

Proof of Proposition 1. Let J denote the upper triangular matrix of size $M_k \times M_k$ ($2 \leq M_k \leq M$) with unit elements on and above the diagonal, and all other elements zero. So for a 3×3 matrix,

$$J = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{A. 1}$$

Its inverse J^{-1} is the matrix with unit elements on the diagonal and with elements -1 adjacent and above the diagonal, thus for our example

$$J^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}. \tag{A. 2}$$

Furthermore, let I be the identity matrix of size $(M-M_k)\times(M-M_k)$, and let T denote the diagonal block matrix of size $M \times M$ with diagonal elements J and I. Thus examples for T and its inverse T^{-1} are

$$T = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A. 3}$$

and

$$T^{-1} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{A. 4}$$

Since T is non-singular, v is an eigenvector of S corresponding to λ if and only if $z = T^{-1}v$ is an eigenvector of $W = T^{-1}ST$ corresponding to λ. Application of Lemma 1—in Warrens and Raadt (2015)— yields that the eigenvector z of W has non-negative elements.

Now, the matrix $U = T^{-1}S$ has elements

$$u_{ij} = s_{ij} - s_{i+1,j} \tag{A. 5}$$

for $1 \leq i < M_k$ and $1 \leq j \leq M$ and

$$u_{ij} = s_{ij} \tag{A. 6}$$

for $M_k \leq i < M$ and $1 \leq j \leq M$. Under requirements (i) and (ii) for the latent utility model, we have that $p_{ij} \geq p_{i+1,j}$ and $r_i > r_{i+1}$ as long as $i, i+1 \in k$. Thus $s_{ij} > s_{i+1,j}$ as long as $i, i+1 \in k$, and the matrix U has non-negative elements except for $u_{i,i+1}$ for $1 \leq i \leq M_k - 1$. But since $s_i > s_{i+1}$ it follows that

$$u_{ii} + u_{i,i+1} = s_{ii} - s_{i,j+1} + s_{i,j+1} - s_{i+1,j+1} = s_i - s_{i+1} > 0 \tag{A. 7}$$

for $1 \leq i \leq M_k - 1$. Thus the matrix $W = UT$ has non-negative elements. □

Proof of Corollary 1. Using Proposition 1, $v_i > v_{i+1}$ for $i, i+1 \in k$. Under requirements (i), (ii) and (iii) for the latent utility model, subtracting v_j on both sides: $v_i + v_j > v_{i+1} + v_j,\ j \in l$ and $k \neq l$, yields $v_i > v_{i+1}$. □

Table A. 1: Dimensions, indicators and items

Variable	All	Poor	Non-poor
	Percentage of people		
		Poverty status	
Variable	**All**	**Poor**	**Non-poor**
Infrastructure			
Walls			
Mud, mud bricks, stone	64.47	80.88	53.33
Fired brick/stone	23.74	14.57	29.62
Concrete	11.79	4.55	17.05
Roof			
Mud bricks or wood with mud	64.74	79.79	55.64
Girder with fired bricks or concrete	35.26	20.21	44.36
Floor			
Dirth/earth	54.52	71.04	43.22
Concrete/tile	45.48	28.96	56.78
Kitchen			
Cooking done in the open	9.13	11.73	7.39
Kitchen is part of a room inside the dwelling	21.80	26.14	19.66
Kitchen is an a separate room outside the dwelling	27.42	30.09	23.32
Kitchen is in a separate room inside the dwelling	41.65	32.04	49.63
Habitability			
No overcrowding	51.45	34.47	62.29
Access to services			
Sanitation	76.82	70.39	82.19
Water	91.38	90.21	92.01
Electricity	98.66	97.80	99.17
Security of tenure			
Dwelling type			
Temporary shelter/shack	2.66	4.20	1.53
Shared house	40.15	41.67	37.76
Single family house	57.20	53.59	58.67
Security of tenure			
Charity	1.78	2.41	1.34
Caretaker	2.26	2.77	1.96
Tenant	20.38	24.78	17.36
Owner	75.58	70.04	79.34

Source: Authors' calculations.

Table A. 2: Polychoric correlations

Wall material	1.00									
Roof material	0.89	1.00								
Floor material	0.79	0.79	1.00							
Kitchen	0.31	0.26	0.37	1.00						
Sanitation	0.39	0.34	0.61	0.50	1.00					
Water	0.11	0.09	0.22	0.18	0.33	1.00				
Electricity	0.27	0.31	0.34	0.17	0.31	0.38	1.00			
No overcrowding	0.19	0.23	0.22	0.08	0.05	0.02	0.13	1.00		
Dwelling type	0.12	0.13	0.06	-0.01	-0.18	-0.09	-0.01	0.18	1.00	
Security of tenure	0.00	0.04	0.05	0.01	0.00	0.03	0.03	0.14	0.24	1.00

Source: Authors' calculations.