## Measuring Housing Deprivation: Methodology and an Application to Afghanistan

Carlos Felipe Balcázar and Silvia Redaelli \*

April, 2017

#### Abstract

Housing adequacy is an important dimension of people's wellbeing, yet there is no consensus or international standard to define and measure adequate housing, or the absence of it. We propose a latent utility model to measure housing adequacy that is consistent with the guidelines in the United Nation's Right to Adequate Housing. First, we provide formal proof that if discrete ordinal data on housing indicators meets the ordering consistency conditions defined herein, housing-adequacy rankings can be gleaned from the factor-scores obtained from using the eigenvector corresponding to the largest eigenvalue in Multiple Correspondence Analysis (MCA). Second, we define welfare consistent cut-off points for housing deprivation on the basis of the guidelines aforementioned. The algorithm measures the incidence of housing deprivation as headcounts, providing a multidimensional deprivation rate. Lastly, we provide an example to estimate housing deprivation rates using Afghanistan's Living Conditions Survey.

<sup>\*</sup>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. The comments of 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, are specially acknowledged.

### **1** Introduction

The international human rights law recognizes everyone's right to an adequate standard of living, including 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 (UN, 2014a). Hence households should have access to adequate housing. Nevertheless, there is no consensus or international standard to define and measure adequate housing, or the absence of it. In this paper we aim at addressing this issue by resorting to the guidelines in the United Nation's *Right to Adequate Housing* to define adequacy, and we deal with the problem of deprivation measurement by proposing a latent utility model.

The current practice is to choose a set of structural characteristics of the dwelling and count the number of housing inadequacies according to local standards for housing to measure the incidence of deprivation (e.g., Amore et al., 2013). However, in general, conceptualizing deprivation measurement involves two steps (Tsui, 2002): First, aggregation of data into one index. Second, identification of the destitute. The first step requires defining the weighting scheme to make aggregations across different goods or services or well-being indicators. The second allows identifying who is deprived according to some established norm. In this paper we propose a methodology to endogenize the calculation of weights, which captures the 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 end-product yields housing deprivation rates that are consistent with the UN's Right to Adequate Housing.

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, while other attribute different weights according to various economic criteria or normative considerations (Decancq and Lugo, 2010, Alkire et al., 2015, Aaberge and Brandolini, 2015). 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 and axiomatic bases for such indexes have not yet been established.

One major concern is that, different weighting schemes could result in very different societal indicators and orders (Yalonetzky, 2013, 2014). While some indices are about capabilities (Atkinson, 2003; Bourgignon and Chakravarty, 2003; Sen, 1999, 2009), aggregation across deprivations cannot (in general) yield deprivation measures that are welfare consistent (Ravallion, 2011).

These issues are important. The design of a meaningful measure of housing deprivation calls for clear axiomatic basis, and for a clear understanding of the implications of using different types

of data and statistical methodologies. In this paper we address these concerns to an extent using discrete ordinal data on housing indicators.

To ease exposition, we will label any ordinal variable an indicator and its values, items. An indicator is essentially a characteristic of the dwelling (e.g., floor material) whereas a value is the attribute of that characteristic (e.g., dirt floor). 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 are utility level iii) and the probability of owning a basket of indicators with a high-ordered item is lower than the probability of 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.<sup>1</sup> 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. Within the context of housing, this eigenvector reflects the ordinal information in the level of housing adequacy.

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. Estimates are also robust to blocked cross-validation.

We also 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. In contrast, MCA does not require discrete variables to follow any underlying distribution (Greenacre, 2006).

<sup>&</sup>lt;sup>1</sup>MCA is a factor analysis technique for nominal categorical data, used to detect and represent underlying structures in a low-dimensional space. Thus the procedure appears to be the counterpart of Principal Component Analysis for categorical data, by applying this procedure to the complete disjunctive table (Greenacre, 1993).

Then we define welfare consistent cut-offs for estimating multidimensional deprivation headcounts and define these cut-offs for each indicator included in the analysis based on UN's Right to Adequate Housing. Given that MCA does a good job at predicting wealth rankings for simulated economies with different prices, we argue that headcounts gleaned from this procedure are informative for policy analysis in the context of housing deprivation, where it is difficult to survey or assess the price of housing attributes.

Our methodology also provides a methodological alternative to potentially using the Multidimensional Poverty Index (MPI) by Alkire and Foster (2015), insofar it is not a counting methodology. It is well known that the MPI cannot provide measures that are welfare consistent (Ravallion, 2012). Further, it can weigh differently different attributes without a clear axiomatic basis for doing so (World Bank, 2016; Ravallion, 2012). In contrast, the method herein provides an alternative for a welfare-consistent multidimensional deprivation measure, where the weights are endogenously derived from the underlying variations and co-variations that are reflective of the trade-offs households engage in. Furthermore, we show that we can enhance spatiotemporal comparability through residualization.

Then we define welfare consistent cut-offs for estimating multidimensional deprivation headcounts, and define these cut-offs for each indicator included in the analysis on the basis of UN's Right to Adequate Housing. Moreover, given that MCA does a good job at predicting wealth rankings for simulated economies with different prices, we argue that headcounts gleaned from this procedure are informative for policy analysis in the context of housing deprivation, where it is difficult to survey or asses the price of housing attributes.

We use the methodology that we present herein to assess housing deprivation in Afghanistan, using the Living Conditions Survey (ALCS) 2013/14. 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. 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.

### 2 Housing deprivation

According to UN (2014a): The right to adequate housing contains freedoms, such as the right to choose one's residence 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). On the one hand, 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.

### **3** 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*.

In the context of housing adequacy, an indicator is essentially a characteristic of the dwelling, such as having access to sanitation, electricity, the material of the floor, the material of the walls,

etc. An item instead defines mutually exclusive attributes of the indicator, so for instance in the case of floor material, the attributes can for instance take the value of dirt, wood or tile.

Now, let U denote a one-dimensional latent (utility) variable and let F(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

$$p_i = \int p_i(U) dF(U). \tag{1}$$

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

- i)  $p_i(U)$  is monotonically increasing on U in every indicator. For all i,  $p_i(U_1) \le p_i(U_2)$  for  $U_1 < U_2$ . That is, the probability of owning a given item may not decrease with the utility level.
- ii) The items of an indicator can be ordered such that the levels of U are not intersecting:  $p_i(U) \ge p_j(U)$  for i < j. 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  $p_i(U)p_j(U)$ , and the corresponding conditional probability is  $p_{ij} = \int p_i(U)p_j(U)dF(U)$ . The joint probability of two items *i* and *i'* for  $i, i' \in k$  given a value of U, is zero.

If there is a single latent variable, and (i), (ii) and (iii) hold, then items in *k* can be ordered such that  $p_i > p_j$  for i < j, and  $p_{ij} > p_{i'j}$  for i < i'.<sup>2</sup> Further, if  $p_i(\cdot)$  is totally positive of order 2 then  $p_i(U_1)p_j(U_2) - p_i(U_2)p_j(U_1) \ge 0$  for  $U_1 < U_2$  and k < l. Therefore the baskets of indicators can be ordered such that  $p_{ij}/p_i < p_{i'j}/p_{i'}$  for i < i' (Schriever, 1986). Thus, the probability of owning a basket of indicators with a high-ordered item is lower than the probability of owning the same basket of indicators but with one lower-ordered item, given a utility level. Within the context of adequate housing, this means that having a house with better floor and wall materials, given the values of other housing-adequacy indicators, is associated to exhibiting a higher level of wellbeing vis-à-vis a household with worse floor and wall materials.

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

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

<sup>&</sup>lt;sup>2</sup>This is the double monotonicity property in nonparametric item response theory (Sijtsma and Molenaar, 2002).

#### The Multiple Correspondence Analysis Algorithm 3.1

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.<sup>3</sup> 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.<sup>4</sup>

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

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

**Proposition 1.** Suppose that  $M_k$  of the M vectors, which without loss of generality can be taken as the first  $M_k$ , can be ordered such that (i), (ii) and (iii) hold. Then the elements of v of S corresponding to these vectors satisfy  $v_1 > v_2 > ... > v_{\ell}M_k \ge 0$ .

*Proof.* In the Appendix.

And finally consider the following corollary:

**Collorary 1.**  $v_i + v_i > v_{i'} + v_i$  for i < i'.

*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.

<sup>&</sup>lt;sup>3</sup>An eigendecomposition is the factorization of a matrix into a canonical form, whereby the matrix is represented in terms of its eigenvalues and eigenvectors:  $P = V\Lambda V^T$ . <sup>4</sup>Since  $PV = V\Lambda$ , then  $P(\frac{N}{n}V) = \frac{N}{n}PV = \frac{N}{n}V\Lambda = V(\frac{N}{n}\Lambda)$ .

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*.<sup>5</sup> It is thus possible to uncover meaningful orderings from categorical data using eigenvectors.

#### **3.2** The multidimensional deprivation index

Denote the spectral decomposition of S as

$$S = V\Lambda V^T, \tag{4}$$

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

$$F = XV. (5)$$

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

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

where  $F_{i,v}$  denotes the value of the deprivation index for individual *i*, with *v* the eigenvector corresponding to the largest eigenvalue. *D* is then the multidimensional deprivation index.<sup>6</sup> In the context of housing, as per Proposition 1 and Corollary 1, *D* then ranks households in terms of housing adequacy.

#### 3.2.1 Index performance

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

<sup>&</sup>lt;sup>5</sup>The ordering consistency labels used herein are borrowed from Asselin (2009).

<sup>&</sup>lt;sup>6</sup>Note that by Proposition 1, lower values of D are preferred to higher values of it.

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 inertia of the off-diagonal blocks of the Burt matrix

$$\frac{K}{K-1}\left(\sum_{s}\lambda_{s}^{2}-\frac{M-K}{K^{2}}\right).$$

We will use this indicator to proxy the index's performance.

#### 3.2.2 Spatio-temporal comparability

It is important to consider that sampling frames can change substantially over time as the local socioeconomic characteristics change. These type of changes can also occur across geographical areas, as well as along the lifecycle. For instance, certain goods or services may become cheaper or more easily available over time and in certain areas; likewise the likelihood of obtaining durables may be a function of the life-cycle as households obtain more durables during the middle of the life-cycle when their incomes and needs are comparatively higher. To address this we residualize the index obtain from carrying out the MCA, by removing the variation explained by common trends over time (indexed by t), both at the country level and at the regional level (indexed by d), as well as the effect of life-cycle using age fixed effects (indexed by a).

Hence we estimate the following regression

$$D_{idt} = \rho_a + \mu_d + \gamma_t + \theta_{dt} + v_{idt}$$

and compute then its residuals

$$D_{idt}^{R} = D_{idt} - \widehat{\rho}_{a} - \widehat{\mu}_{d} - \widehat{\gamma}_{t} - \widehat{\theta}_{dt}.$$

### 3.3 Cross-validation

One of the limitations of the algorithm above is that it can generate overfitting insofar as the algorithm tries to capture the largest amount of variation in the first dimension, and so on. To address this problem we use cross-validation. This consists of splitting the sample in C-random subsets and then averaging our estimates along these C subsets.<sup>7</sup> However, we believe it is important to perform cross-validation on the basis of the sampling frame to reduce the variability of the errors within strata, hence the random sub-samples should be drawn using blocked randomization, where the blocks are the strata. Thus let us redefine our multidimensional deprivation index derived from using blocked cross-validation as

$$D_i = \frac{\sum_c D_{ic}^R}{C}.$$

#### 3.4 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 consider having dirt floor is the same as being floor-deprived.<sup>8</sup>

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

$$z = D(M_k)$$
 with  $M_k = z_k$  for all  $k$ . (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. Thus when this measure is applied over the indicators for adequate housing we obtain a housing deprivation rate.

The deprivation bundle  $\{z_1, z_2, ..., z_K\}$  is consistent with the choices made by someone living

<sup>&</sup>lt;sup>7</sup>This is referred to as K-fold cross validation, but since we are using the letter K to denote the number of indicators, we use the letter C as an alternative herein.

<sup>&</sup>lt;sup>8</sup>We know from the development literature that having dirt floor is related to poor health (Cattaneo et al., 2009).

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 care 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  $y = p_1k_1 + p_2k_2$ ; 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:  $D = v_1k_1 + v_2k_2$ , 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** Monte Carlo Experiments

Let us consider the following utility maximization problem

$$\max_{k} \left( \sum_{j}^{K} \alpha_{j} k_{j}^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}} \quad so \ that \quad 0 \leq k, \ p \cdot k \leq I,$$

where *I* denotes endowments, *p* is the vector of prices, *k* is the vector of indicators (goods and services),  $\sigma$  is the elasticity of substitution, and  $\alpha$  is the share parameter. The optimal demand function for any *k* is its Marshallian demand and it is denoted by  $k^*$ .

Without loss of generality, for our simulations, we assume *I* follows a log normal distribution, such that  $\ln(I) \sim \mathcal{N}(\mu_I, \sigma_I^2)$ ;<sup>9</sup>  $\alpha$  are randomly assigned as well using a uniform distribution  $\alpha \sim \mathcal{U}(0,1)$  so that  $\sum_{j=1}^{K} \alpha_j = 1$ ;  $\sigma$  is also randomly assigned using a normal distribution  $\sigma \sim N(0,1)$ ; 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

<sup>&</sup>lt;sup>9</sup>This obeys the fact that most distributions of wealth are skewed with heavy left tails.

that we assume that 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:  $k_j^* = \lfloor k_j^* \rfloor$ . We use  $k_j^*$  to perform MCA and PCA as well (as means of providing comparisons to a standard method).<sup>10</sup> 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 (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  $\tau$ .<sup>11</sup> 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  $\tau$  coefficients (Figure 2).



#### Figure 1: Percentage of explained variation

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

<sup>&</sup>lt;sup>10</sup>PCA performs the eigendecomposition on the correlations matrix instead of the Burt matrix. See Filmer and Pritchett (2001) for a discussion.

<sup>&</sup>lt;sup>11</sup>The Spearman a rank correlation coefficient and Kendall's  $\tau$  measure the ordinal association (or rankings) between two variables, and can be used to assess the significance of the relation between them.





Note: The rankings of the multidimensional deprivation index are defined from its largest value (the lowest rank) to its lowest value (the highest rank).  $\tau_a$  will not make any adjustment for ties whereas the  $\tau_b$  statistic makes adjustments for ties.

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

### 5 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. For security of tenure we explore two dimensions: i) 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.

All in all, Table 1 lists the dimensions, the harmonized ordinal variables and includes a description of each. Table 1 lists the dimensions, the harmonized ordinal variables and includes a description of each.<sup>12</sup>

<sup>&</sup>lt;sup>12</sup>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.

Dimension	Variable	Constructed categories	Description			
Infrastructure	Wall material	<ol> <li>Mud, mud bricks, stone</li> <li>Fired brick/stone</li> <li>Concrete</li> </ol>	Categories were harmonized and were organized from worst to best, according with the structural properties of housing materials.			
	Roof material	<ol> <li>Mud bricks or wood with mud</li> <li>Girder with fired bricks or concrete</li> </ol>	Categories were harmonized and were organized from worst to best, according with the structural properties of housing materials.			
	Floor material	<ol> <li>Dirth/earth</li> <li>Concrete/tile</li> </ol>	Categories were harmonized and were organized from worst to best.			
	Kitchen	<ol> <li>Cooking done in the open</li> <li>Kitchen is part of a room inside the dwelling</li> <li>Kitchen is in a separate room outside the dwelling</li> <li>Kitchen is in a separate room inside the dwelling</li> </ol>	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.			
	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.			
Services	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 electricity at any time in the past month from the electric frid, generator, solar panel, wind power or a battery.			
Security of tenure	Dwelling type	<ol> <li>Temporary shelter/shack</li> <li>Shared house</li> <li>Single family house</li> </ol>	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.			

### Table 1: Dimensions, indicators and items

Source: Authors' compilations.

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 the (monetary) poor fare worse off than their non-destitute counterpart.<sup>13</sup>

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 – although this is neither necessary nor sufficient, just indicative. Table A2 in the Appendix shows that there is a high 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 literature 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):

<sup>&</sup>lt;sup>13</sup>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).

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				

**Table 2: Threshold categories** 

Source: Authors' compilations.

## **6** 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).<sup>14</sup>

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.

<sup>&</sup>lt;sup>14</sup>The percentage of variation explained by PCA is only 25%.

Nonetheless, differences are small in the south and west central regions.

Variable	Constructed categories	MCA		
	1. Mud, mud bricks, stone			
Wall material	2. Fired brick/stone			
	3. Concrete			
	1. Mud bricks or wood with mud			
Roof material	2. Girder with fired bricks or concrete			
	1. Dirth/earth			
Floor material	2. Concrete/tile			
	1. Cooking done in the open			
	2. Kitchen is part of a room inside the dwelling			
Kitchen	3. Kitchen is in a separate room outside the dwelling			
	4. Kitchen is in a separate room inside the dwelling			
	0. No access to improved sanitation			
Sanitation	1. Access to improved sanitation			
	0. No access to improved water			
Water	1. Access to improved water			
	0. No access to electricity			
Electricity	1. Access to electricity			
	0. Overcrowding			
No overcrowding	1. No overcrowding			

 Table 3: Eigenvector corresponding to the largest eigenvalue

Source: Authors' compilations.

Note: Percentage of inertia explained by the first component: 78.5%.

Variable	A 11	Poverty status			
variable	All	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		

# Table 4: Multidimensional deprivation headcounts

Source: Authors' calculations.

### 7 Conclusion

Measuring housing 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 that both housing adequacy and deprivation will be defined. We address the first problem by fitting a latent utility model, obtaining 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 that are consistent with UN's Right to Adequate Housing. We provide an example for our methodology using Afghanistan's Living Conditions survey, and show that our results are informative about (housing) deprivation profiles.

### 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.

Bourguignon, F., Chakravarty, S.R. (2003) The Measurement of Multidimensional Poverty. The Journal of Economic Inequality 1, 25–49. https://doi.org/10.1023/A:1023913831342

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.

Gantmacher, F.R. (1977) The Theory of Matrices. London: Chelsea Publishing.

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.

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

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 University 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.

McKay, S. (2004). Poverty or preference: what do 'consensual deprivation indicators' really mean?. Fiscal studies, 25(2), 201-223.

Rao, C.R. (1973) Linear Statistical Inference and Its Applications. 2nd Edition, Wiley, New York. https://doi.org/10.1002/9780470316436

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.

Ravallion, M. (2015). The economics of poverty: History, measurement, and policy. Oxford University Press.

Sen, A. K. (1999). Development as Freedom. OUP.

Sijtsma, K., & Molenaar, I. W. (2002). Introduction to nonparametric item response theory (Vol. 5).

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]

UNDP. (1993). Human Development Report. Oxford: Oxford University Press.

UNDP. (2015). Work for Human Development. Human Development Report 2015. New York, NY. Walker, Gordon, Simcock, Neil, & Day.

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

World Bank. (2016). Monitoring global poverty: Report of the commission on global poverty. The World Bank.

### A Mathematical appendix

*Proof of Proposition 1.* Following Warrens and de Raadt (2015):

First, consider a weaker version of the Perron-Frobenius theorem whereby if a square matrix D has strictly positive elements, then the eigenvector corresponding to the largest eigenvalue of D has strictly positive elements (Rao, 1973; Gantmacher, 1997).

Second, let A denote the upper triangular matrix of size  $M_k \times M_k$  ( $2 \le M_k \le M$ ) with unit elements on and above the diagonal, and all other elements zero. So for a  $3 \times 3$  matrix,

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$
 (A.1)

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

$$A^{-1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$
 (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

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

and

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

Third, since *T* is non-singular, *v* is an eigenvector of *S* corresponding to  $\lambda$  if and only if  $w = T^{-1}v$  is an eigenvector of  $D = T^{-1}ST$  corresponding to  $\lambda$ . Application of the Perron-Frobenius theorem referenced above yields that the eigenvector *w* of *D* has strictly positive elements.

Fourth, the matrix  $G = T^{-1}S$  has elements

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

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

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

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  $s_{ij} \ge s_{i+1,j}$  and the matrix U has non-negative elements except for  $s_{i,i+1}$  for  $1 \le i \le M_k - 1$ . But since  $s_i > s_{i+1}$  it follows that

$$g_{ii} + g_{i,i+1} = s_{ii} - s_{i,j+1} + s_{i,j+1} - s_{i+1,j+1} = s_i - s_{i+1} > 0$$
(A.7)

for  $1 \le i \le M_k - 1$ . Thus the matrix D = GT has non-negative elements. Moreover, because the elements in the last row and last column of D are strictly positive, the elements of  $D^2$  are strictly positive.

*Proof of Corollary 1.* If  $s_i$  is both monotonically increasing and satisfy total positivity of order 2, Schriever (1986) proved that

$$\frac{s_{ij}}{s_i} > \frac{s_{i+1,j}}{s_{i+1}},$$

which follows from Proposition 1.

## **B** Empirical appendix

	Percentage of people						
Variable	All	Pove Poor	rty status 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				

### Table B. 1: Dimensions, indicators and items

Source: Authors' calculations.

 Table B. 2: Polychoric correlations

Wall material	1.00									
<b>Roof material</b>	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.