A multilevel structural equation modelling approach to understand factors associated with clustering of deprivation in Bolivia

Filippo Temporin, London School of Economics

Keywords: deprivation clustering; Bolivia; multilevel model; structural equation model

Abbreviations

DHS: Demographic and Health Survey

GPS: Global Positioning System

PCA: Principal Component Analysis

PSU: Primary Sample Unit

SEM: Structural Equation Model

Abstract

In the first decade of the XXI century, Bolivia had one of the worst performances in poverty headcount ratio and chronic malnutrition in Latin America, according to World Bank data. The distribution of deprivation within the Bolivian territory is not even, and deprived households are clustered in certain communities. The analysis of the factors related to deprivation clustering is of interest since this phenomenon can be linked to health, social exclusion and to lower access to public services.

This paper aims to answer the following research question: "What are the contextual factors associated with the clustering of household deprivation in Bolivian communities?". By analysing 2008 Demographic and Health Survey data, this analysis aims to quantify the extent to which household deprivation is clustered within Bolivian communities, and to explore contextual factors associated with deprivation clustering.

Ethnicity, education, administrative region, distance to urban centres, and drought-induced migration significantly predict differences in the mean level of deprivation across Bolivian villages. By identifying the factors associated with the uneven sorting of deprivation, the analysis of deprivation segregation will inform policy makers on implementing policies related to urban planning and schooling policies at the local level.

In comparison to descriptive measures of poverty clustering, a multilevel structural equation modelling approach allows us to make statistical inference on segregation, and to model the variance as a function of contextual predictors. This analysis involves a continuous latent variable as an outcome, and therefore represents an extension of the multilevel models used in previous work.

1) Introduction

Segregation can be defined as a form of physical separation where population groups are isolated into different neighbourhoods (in case of residential segregation) or schools (in case of educational segregation), “shaping the living environment at the neighbourhoods [or school] level” (Kawachi and Berkman, 2003).

Clustering is the grouping of identifiable minorities – in this case, deprived people – in a way that people within the same cluster (group) have more similar characteristics than those belonging to other clusters (Schoen and Nau, 2008). Together with other aspects such as concentration and centralization, the phenomenon of clustering is one of the dimensions of segregation (Massey and Denton, 1988; Singh et al., 2009). Geographical clustering of deprived people is commonly associated with economic, ethnic, or physical segregation, being the consequence of variation in characteristics under study across areas. Clustering of deprivation may be related to social exclusion[1], with important consequences for social and health policies. Among the effects of social exclusion, we can highlight a diminished access to public services and decreased opportunities for human capital development. In Bolivia, for instance, social exclusion has been identified as a possible mechanism through which individuals belonging to certain ethnic groups reside in areas that tend also to have lower education and income (Gray-Molina et al., 2002). There is some evidence that the opportunities and even the conduct of people residing in certain neighbourhoods is shaped, among other factors, by the characteristics of their neighbourhood (Jencks and Mayer, 1990). Geographic and social isolation could therefore be among the factors underlying certain social pathologies among the poor (Greene, 1991).

[1] Social exclusion is the mechanism through which members of a certain group are denied full access to resources and opportunities that are available to others, associated for instance with housing, employment, or healthcare, and linked to social integration (Silver, 1994)

The analysis of deprivation and poverty segregation can help to identify the most deprived areas, which are economically and socially isolated from the more developed areas. It can provide a tool to determine the economic, social and institutional factors related to spatial unevenness in the distribution of wealth over the area under investigation. Deprivation and poverty segregation might be particularly suitable for policy interventions related to urban planning at a more local level than the national or regional level (Amarasinghe et al., 2005). Moreover, since higher mortality and higher exposure to infectious diseases is likely to be found in contexts of concentrated deprivation (Fiscella and Franks, 1997; Szwarcwald et al., 2002), reducing the differences in deprivation among communities might also be associated with the reduction of mortality.

This paper focuses on the study of clustering of deprivation in Bolivia in 2008. By the end of the first decade of the millennium, Bolivia was one of the poorest countries in South America (Population Reference Bureau, 2013). For instance, compared to other countries in Latin America, Bolivia performed worst in terms of chronic malnutrition (22% of children) (Coa and Ochoa, 2009), and more than half of the population fell below the poverty line, mostly in rural areas (World Bank, 2014). A peculiarity of Bolivia within Latin America is its condition of underdevelopment, while being surrounded by countries that present an overall ongoing state of development: since 1950, Bolivia’s real income level has remained almost constant, while in other South American countries it has doubled (Wiggins et al., 2006). Bolivian economic inequality is still great, with a Gini coefficient of 51.4 in 2008 (against an average of 49.9 of the other South American contries), and large differences exist in the distribution of income (World Bank, 2014). The distribution of wealth within the country is not uniform, with considerable geographic and ethnic dissimilarities (Schroeder, 2007).

This paper aims to answer the following research question: “What are the contextual factors associated with the clustering of household deprivation in Bolivian communities?”. First, the extent of clustering of deprivation across Bolivian communities is quantified, and then area-level variables are used to explain the variation across communities, while allowing for clustering due to unmeasured area characteristics.

Multilevel structural equation modelling (SEM) is applied to data from the 2008 Bolivian Demographic and Health Survey (DHS). SEM allows the simultaneous creation of a latent variable for household deprivation, and its decomposition into between-community and between-household within-community components. By including community-level covariates, SEM allows exploration of the factors associated with deprivation clustering.

Deprivation is conceptualized as a lack of basic needs related to housing conditions and living standards, rather than monetary measures. The use of asset indices has been increasing in the last years, especially in the context of developing countries (Filmer and Scott, 2012). An alternative measure to the DHS wealth index is proposed, taking into account only items related to housing conditions with a sufficient degree of correlation among them, and which can therefore be considered manifestations of the underlying concept of household deprivation. This study builds on the previous use of multilevel modelling that assessed educational segregation in schools and areas (Goldstein and Noden, 2003; Leckie et al., 2012). Here, the main development is the fact that the outcome of interest, household deprivation, is treated as a latent variable.

2) What do we know about measures of poverty and deprivation clustering?

Two main approaches to the analysis of poverty clustering can be identified in the literature. The first approach involves the use of descriptive indicators. Conceptually a simple measure of poverty segregation is the proportion of the poor population living in areas with high levels of poverty (Greene, 1991). By making use of this index, the higher the proportion of poor people living in areas with high poverty rates, the higher the level of segregation. A more sophisticated segregation measure is the dissimilarity index (Duncan and Duncan, 1955), which can be interpreted as the percentage of one of the population groups (for instance, the white population in the case of racial segregation) that would have to move to different areas in order to reproduce a distribution matching that of the larger areas. The dissimilarity index has been widely used in the deprivation and poverty segregation literature (Iceland et al., 2014; Sparks et al., 2013). Szwarcwald et al. (2002) propose an index of heterogeneity of poverty intensity per unit area derived from Cramer's contingency coefficient, as a measure of dissimilarity in income concentration across census tracts within neighbourhoods in Rio de Janeiro. Other conventional descriptive measures of segregation are Bell's (1954) and Lieberson's (1981) isolation indices for multiple populations, and Theil's (1972) entropy index, which can be interpreted as the distance of the observed population from the egalitarian state of even distribution of a given characteristic across all individuals. The standardized versions of these indices range from 0 (no segregation, all areas having the same proportion of population groups) to 1 (complete segregation, each area being composed of just one of the population groups) (Hammel et al., 2010).

The second approach to the analysis of poverty clustering makes use of Global Positioning System (GPS) data (Matthews and Parker, 2013). The gradient of spatial clustering can be measured by spatial autocorrelation (Cliff and Ord, 1973), which finds applications, among others, in the study of poverty spatial clustering in Sri Lanka (Amarasinghe et al., 2005) and spatial inequalities in neighborhood walkability (Duncan et al., 2012). Crandall and Weber (2004) measure the proportion of adjacent US census tracts in high poverty, allowing for the identification of poverty clusters. Another index of poverty clustering is defined by Stretesky et al. (2004) as the proportion of adjacent census tracts within cities which have at least 40% poor residents.

The above-mentioned indices are descriptive, meaning they are built on observed proportions that include the effect of random sampling variation. In other words, they fail to take into account the probabilistic component resulting from the sampling process. For instance, Leckie et al. (2012) point out that the dissimilarity index, being based on observed rather than on underlying proportions, has sources of bias depending on the size of the areas and on the underlying proportions.

A multilevel model approach overcomes these limitations, by separating the component of the observed proportion that is due to sampling variation. The first paper in this stream of literature is by Goldstein and Noden (2003) who measured the evenness of the distribution of disadvantaged students across English schools in the period 1994-1999. This paper has been further developed by Leckie et al. (2012), who introduced a third level in the hierarchical structure of the data, with students nested within schools nested within London local authorities. Both these papers have a binary variable as the outcome, namely students' eligibility for free school meals. The present analysis involves a continuous latent variable as an outcome, and therefore represents an extension of the multilevel models used in previous work. Segregation can be measured by estimating the higher-level variance parameter in the multilevel model. This allows assessment of the proportion of variation in the characteristic of interest that is due to the grouping of individuals within areas: the larger it is, the more segregated the neighbourhoods or schools are. By estimating standard errors, statistical inference on segregation can be made. Moreover, multilevel models can explain sources of segregation by including contextual covariates in the models, modelling the variance as a function of such area characteristics (Leckie et al., 2012).

3) Potential explanations for geographical clustering of deprivation in Bolivia

Clustering of deprivation is strictly related to variation across communities. In fact, the higher the between-community variation of the level of deprivation in a country, the higher the level of grouping of deprived people within geographical areas. On the other hand, no between-community variation indicates that no clustering is present in a country. The main aim of this paper is to explain the clustering of deprivation, by looking at the potential factors associated with the between-community variation in deprivation. Among these, ethnic composition, education, distance to urban centres and drought-induced rural-urban migration can have a central role.

The first factor that may affect the clustering of deprivation is ethnicity. The Bolivian population is mainly indigenous, and the ethnic distribution is not uniform, with indigenous populations more concentrated in certain areas – mainly the *Altiplano* (high plateau) and *Valle* (valley) regions. These populations are found in the literature to be more likely to be deprived: the lack of social welfare programmes leads to a high vulnerability to shocks such as droughts, floods and hailstorms (Buzaglo and Calzadilla, 2009). Almost the whole indigenous population (97.5%) of rural areas is found to be chronically poor (Castellanos, 2007). Ethnicity can therefore be a possible source of deprivation clustering: since indigenous households are more disadvantaged, the concentration of these households in certain areas leads to clustering of deprivation. Ethnicity can also be a factor in fostering urban residential segregation, after the increase in migration to Bolivian cities experienced over the last decades (Balderrama, 2011).

Education can play a role in explaining between-community variation in the level of deprivation in the country. The link between parental education and the socioeconomic status of a household is well established (Cornia, 2014; King and Hill, 1993). Education can also be a contextual factor in determining the unevenness of the distribution of deprivation across Bolivian communities. The average degree of education in the community can set the context for a wide set of socioeconomic factors, including economic disadvantage (Wight et al., 2006) which lead to the geographical clustering of deprivation. Education is also strongly related to ethnicity, since a high proportion of indigenous people have no formal education (Castellanos, 2007).

Distance to urban centres might also explain deprivation clustering. Social segregation studied by Gray-Molina et al. (2002) in Bolivian urban environments, can be extended to rural areas. The main activity in rural areas is farming: peasants are vulnerable to shock linked to climate change such as drought (Castellanos, 2007), and lack of roads might affect peasants' access to the market (Buzaglo and Calzadilla, 2009). Rural areas are also associated with a lack of infrastructure (Andersen, 2002) and basic services like sanitation and availability of clean water (Coa and Ochoa, 2008), creating a setting of a higher mean level of deprivation.

Finally, Bolivia has been subject to natural disasters over the last decades. In particular, prolonged droughts have affected the South-West part of the country (Kessler and Stroosnijder, 2006). Agriculture and livestock rely strongly on vegetation resources, the availability of which can be jeopardized by these events: it has been calculated that, in the period 1953-1993, Bolivia lost 30% of its agricultural productivity, and one of the main reasons is related to soil erosion (Benton, 1993). Droughts have fostered migration towards the cities. Bolivia faced a rapid process of urbanization, either temporary or permanent, between the 1980s and the 2000s (World Bank, 2015). Drought-driven rural-urban migration can lead to the uneven residential sorting of rural migrants within cities, which leads to a rise in the level of urban residential segregation. Moreover, there is some evidence of a recent trend towards migration differentiated by age-group. The main mechanism is related to the fact that young men are gradually excluded from access to agricultural soil, due to the increased unavailability of land (Balderrama, 2011). Lands are usually distributed among the children, but there is evidence of the tendency of migrant young men to refuse their share of the inheritance (Michels, 2011).

This selective migration (Borjas and Tienda, 1987) can therefore be another explanation for the clustering of deprivation in Bolivia.

4) Data and measures

4.1) The 2008 Bolivian Demographic and Health Survey

The Demographic and Health Surveys (DHS) collect data on a broad range of aspects related to health and living conditions. The target population of DHS is all women of reproductive age (15-49). DHS uses a probability sample, where the units are selected with known and nonzero probability to ensure a geographical coverage of the entire national territory (US Aid, 2012). In the sampling process, clusters of a standard size of 100 households are identified and mapped in the territory of the country under investigation, and a further selection within each of these selected clusters is made: each of these areal units serves as a primary sample unit (PSU). A fixed proportion of those households is selected by systematic sampling, and a face-to-face interview is conducted with all women aged 15-49 who are members of the selected households (US Aid, 2012). In this paper, PSUs are considered to be proxies for the respondents' communities, as in previous studies (Uthman et al., 2011; Robson et al., 2012).

The 2008 Bolivian DHS dataset contains 19,564 households from 999 communities. Among them, 11,361 household have complete records on the ownership of the items related to housing conditions and on the variable included as predictors in the structural model.

4.2) Indicators of deprivation

The full set of items related to housing conditions, living standards and owned assets available in the DHS dataset includes: availability of electricity, availability of clean water, type of sanitation, material of the floor, type of cooking fuels, and ownership of refrigerator, radio, television, motorbike, car, telephone and bicycle. These are the items used in the construction of the DHS wealth index, a composite measure of a household's cumulative living standard (Rutstein and Johnson, 2004).

All the observed variables have been dichotomized, in order to simplify the interpretation of the parameters of the models. The categories for sanitation are gathered into two groups, reflecting improved and unimproved hygiene. Sewage and septic systems are included in the first group, while open pit and surface water (street or stream) are in the second (Günther and Fink, 2010). The indicator of floor material is made binary by creating the categories “adequate floors” (parquet and “machimbre” - tongue and groove joint, carpet, cement, tile, ceramic and bricks) and “mud, dirt and other materials” (Vandemoortele, 2014). Water is considered of adequate quality if it is piped into dwellings, yards, plots, or if its source is a public tap or standpipe, a tube well or borehole, a dug open or protected well, a protected spring, or if it is rainwater or bottled water, and if all these sources are within half an hour walking distance from the interviewee’s residence. Low-quality water comes from unprotected wells and springs, rivers, dams, lakes, ponds, streams, tanker trucks, carts with small tanks, or if it is surface water, or if its source is further than half an hour walking from the home of the person interviewed. This approach reflects the categorization already existing in the literature (WHO and UNICEF 2014).

4.3) Explanatory variables

As noted earlier, there are four main factors that can be linked to the between-community variation in deprivation: ethnicity, education, distance to urban centres and drought-induced migration. These are represented by five explanatory variables listed in Table 1. All of these have been measured at the community level.

Table 1: list of covariates

Variable	**Source**	**Values**
Indigenous village	DHS	Indigenous, Non-indigenous
Male education (years)	DHS	[0.7; 17]
Administrative region	DHS	Beni, Chuquisaca, Cochabamba, La Paz, Oruro, Pando, Potosí, Santa Cruz, Tarija
Distance to the closest municipal capital (km)	GeoBolivia	[0.06; 96.51]
Risk of drought	SINSAAT	Very low, Low, Medium, High

The contextual binary variable *Indigenous*, provided by DHS, indicates whether a household lives in a community which has a majority of indigenous or non-indigenous villages. The mean level of male education within each community has been chosen as a contextual variable. For households with more than one adult male (5.97% of the total), the mean value of years of schooling of the males registered at that household has been calculated. In general, individual-level male education can better explain the level of deprivation than female education: paternal rather than maternal income is a strong determinant of the wealth status of the household (Cornia, 2014; Thomas, 1990), and in Bolivian indigenous groups, men are more likely to assume the position of breadwinners (Paulson et al., 1996). However, it might be argued that female education can have a role in explaining the level of household deprivation, since female headship can be important in single-parent households, including those where husbands have left to work in other cities or overseas. Therefore, a sensitivity analysis using female instead of male education has been carried out (results in Appendix 1).

The distance from the centroid of each DHS cluster to the closest municipal capital has been obtained by linking the DHS GPS dataset and the GeoBolivia dataset (GeoBolivia, 2017a), which provides the location of the 339 Bolivian municipal capitals. The distance has been calculated using the Haverisine formula[2] (Robusto, 1957). This measure aims to be a development of the variable for place of residence provided in the DHS dataset, which has only the two categories “urban” and “rural”. The distance to the closest municipal capital can provide a better measure of the variation between urban and rural environments, approaching the concept of Woods’ (2003) “urban-rural continuum”. The distance to the closest municipal capital ranges from 0.06 to 96.51 kilometres. The mean distance of the communities labelled as urban in the DHS variable is 3.88 kilometres, while it is 16.84 kilometres for the rural communities. The variable related to risk of drought has been created by linking the DHS GPS dataset with the 2002 National System for Early Alert of Food Security (Sistema Nacional de Seguridad Alimentaria Alerta Temprana, SINSAAT) (GeoBolivia, 2017b). This dataset classifies areas into four levels of drought risk, depending on the frequency of drought over the period 1972-2002. Very low risk is defined as one or no drought every fifth year over the 30-year period, low risk as a drought every fourth year, medium risk as a drought every second year and high risk as four or more droughts every five year.

In the most recent DHS surveys, each community is georeferenced during the sample listing process. The GPS readers are in general accurate to less than 15 metres, but the GPS coordinates of each community are randomly displaced due to issues of confidentiality: the error ranges from 0 to 2 kilometres for urban communities and from 0 to 5 kilometres for rural

[2] The Haversine distance does not reflect real distance, especially in a territory like Bolivia, which is highly mountainous in the South-West areas. It is reasonable to think that Bolivians willing to reach the closest municipal capital might have to cover longer distances than the great-circle line connecting their village to the target. A better estimate of such distance would be the walking (or driving) path from each community to the municipal capital. However, no reliable GPS dataset on minor streets and trails has been found. The only available dataset is related to main roads (GeoBolivia, 2013), but this is not specific enough to include all the walking trails that Bolivians might take. Therefore, the Haversine formula has been considered the best available approximation of the real distance to the closest municipal capital.

communities (Perez-Heydrich et al., 2013). While cluster displacement might induce large misclassification errors when calculating the distance between clusters' centroids and health facilities or other specific locations (Skiles et al., 2013), the random displacement of the centroid of the communities is unlikely to affect the results of this study. First, the region of each community is directly calculated from DHS, so no issue of displacement arises even when the random error is introduced. Second, the distance to the closest municipal capital is the variable that mostly could be affected by the random error. Therefore, a sensitivity analysis using the binary variable for urban or rural place of living provided in the DHS dataset instead of the distance to distance to the closest municipal capital has been carried out in Appendix 2. Since no substantial difference in the results is observed, the continuous variable for distance to the municipal capital has been retained in the models, since it is considered a better approximation of the rural-urban continuum (Woods, 2003). Third, the areas for risk of drought are very large and the risk of displacement of a community seems very low. There are 46 communities within 5 kilometres from the borders between areas at different risk of drought (Figure 3 in Appendix 3). A sensitivity analysis has been carried out, changing the categorization of these communities to the area of risk of drought they might have been misplaced from; no substantial difference is found in the results.

5) Statistical methods

5.1) Latent variable model for household deprivation

An index measuring deprivation (or wealth) is an alternative to monetary measures such as income or expenditure, which are often unavailable or unreliable in low- or middle-income countries (Filmer and Scott, 2012). Deprivation can be considered as a concept underlying certain characteristics of living standards and can therefore be derived from a set of observable items.

A key point in the creation of a composite index of deprivation is the choice of weights to be assigned to the observed items. Many approaches exist in the literature, ranging from the simple sum of the owned items to more sophisticated data-driven techniques that take into account the extent to which each item discriminates between households' deprivation (Vandemorteele, 2014). Among these composite indicators, the DHS wealth index, built from principal component analysis (PCA), is probably the most widespread (Rutstein and Johnson, 2004). In the following sections, a critique of the construction of the DHS wealth index is presented, and a latent variable approach is proposed.

5.2) Critique of the DHS wealth index

The DHS wealth index is constructed by means of PCA, a technique that transforms a set of observed correlated items into a set of linearly uncorrelated principal components by means of an orthogonal transformation (Jolliffe, 1986). PCA's major limitation is that it does not take into account the categorical nature of the observed indicators, treating them as continuous, which is analogous to using an OLS regression for the analysis of a categorical outcome (Howe et al., 2008). The wealth index scores are built from the first principal component, which often explains only a low proportion of the total variation in the observed items (Kolenikov and Angeles, 2004). Moreover, since the correlation between the observed indicators has not been investigated before the analysis, the linear dependence between the items could lead to incorrect estimates of the wealth index (ibid.). Finally, using the DHS wealth index as a measure of deprivation in further analyses ignores the measurement error that arises from constructing an index from a set of items.

5.3) Rationale for the construction of a latent variable for household deprivation

SEM is a latent variable approach that incorporates a model for the relationship between a continuous latent variable and a set of observed items, considered as the manifestation of the latent variable (Bartholomew et al., 2011). In this case, for instance, a set of observed items relating to housing conditions and living standards are combined into a latent variable for household deprivation.

A SEM is composed of a measurement model and a structural model, estimated simultaneously. The measurement model describes the relationship between the observed items and the latent variable. The structural model is a regression of the latent variable on a set of covariates (Bartholomew et al., 2011). In contrast to PCA, the items included in the measurement model of SEM can be binary or polytomous (ibid.). Weights are assigned to the items depending on their ability to discriminate between households' scores on the latent variable. By estimating standard errors, SEM also allows testing hypotheses involving parameters of both the measurement and structural models. An important feature of SEM is that it takes into account the measurement error which may bias the estimates of the level of segregation within communities. Latent variables do not have measurement error associated with them, since they are not directly measured, therefore the association between them and other covariates can be estimated without any bias (Muthén and Muthén, 2010).

In comparison to the DHS wealth index, a further development of the proposed approach is the selection of the observed items, which is based on the correlation matrix of all items. Only items relating to the latent concept of deprivation are included in the measurement model, as explained later.

5.4) Measurement model

The measurement model specifies the relationship between the latent variable and the observed items. Denote by y_{rjk} the r^{th} item $(r = 1, \ldots, p)$ of household j $(j = 1, \ldots, n_k)$, nested within community k $(k = 1, \ldots, K)$. Then the logit of the probability that household j in community k owns item r is:

$$\operatorname{logit}\left(P\left(y_{rjk} = 1 \middle| \eta\right)\right) = \operatorname{logit}\left(\pi_{rjk}(\eta)\right) = \alpha_{r1}\eta_{jk} - \alpha_{r0}, \qquad (1)$$

where $\eta_{jk} \sim N(0, \sigma_\eta^2)$ is the latent variable for household deprivation and α_{r0} and α_{r1} are, respectively, the difficulty and the discrimination parameters. The difficulty parameter α_{r0} indicates how "difficult" an item is to be owned, while the discrimination parameter α_{r1} indicates how well the r^{th} item discriminates between households with different scores for deprivation. In order to identify the model, some constraint must be imposed on the item parameters. It is common to constrain one of the α_{r1}s to 1, which sets the scale of the latent variable to be equal to the scale of the chosen item.

5.5) Multilevel structural model

In this paper, the multilevel structural models specify the partitioning of the variance into a between-community component and a within-community between-household component. Of particular interest is the extent to which community variation can be explained by the community-level covariates described earlier. An important characteristic of multilevel SEM is that the creation of the latent outcome variable and the analysis of its between- and within-community components is done simultaneously, while accounting for measurement error (Muthén and Muthén, 2010).

The structural model specifying the decomposition of the latent variable η_{jk} into its within- and between-community components is:

$$\eta_{jk} = \beta_k + u_{jk}^{(hh)}$$

$$\beta_k = \gamma_{00} + u_k^{(PSU)}, \qquad (2)$$

where $u_{jk}^{(hh)} \sim N(0, \sigma_u^{2\,(hh)})$ is the household residual and $u_k^{(PSU)} \sim N(0, \sigma_u^{2\,(PSU)})$ is the community-level random effect. They represent, respectively, the within-community and the between-community components of household deprivation, and their variances $\sigma_u^{2\,(hh)}$ and $\sigma_u^{2\,(PSU)}$ are the within-community and the between-community variances.

When including a contextual variable calculated as the mean of a household-level variable $\bar{X}_k$, it is common to include the group mean centred household-level variable, in this case $X_{jk} - \bar{X}_k$. Therefore, the coefficient associated with the group mean centred household-level variable is the within-group effect, while the coefficient associated with $\bar{X}_k$ is the between-group effect, measuring the relationship between the covariate and the outcome at the community level (Snijders and Bosker, 2012). Therefore a model with a community mean of a household variable can be specified as:

$$\eta_{jk} = \beta_{0k} + \beta_1 (X_{jk} - \bar{X}_k) + \varsigma_{jk}$$

$$\beta_{0k} = \gamma_{00} + \gamma_{01} \bar{X}_k + u_{0k}, \qquad (3)$$

where $\eta_{jk} \sim N(0, \sigma_\eta^2)$, $\varsigma_{jk} \sim N(0, \sigma_\varsigma^2)$ and $u_{0k} \sim N(0, \tau_{00}^2)$.

The models are fitted by maximum likelihood, and likelihood ratio tests can be used to compare the fit of nested models. The analyses have been carried out using the *gsem* function in the Stata software (StataCorp, 2013).

6) Construction of the latent variable for household deprivation

6.1) Inspection of the correlation matrix of deprivation indicators

The full set of 12 items available in the DHS dataset includes *Electricity, Water, Sanitation, Floor, Cooking fuels, Radio, Television, Refrigerator, Motorbike, Bicycle, Car* and *Telephone*. These are the same items used for the construction of the DHS wealth index. These items can be divided into two sets: the first five items are related to the living environment, while the last seven are assets or possessions.

The aim of the investigation of the correlation matrix is to select the observed items used to construct the latent variable, in order to avoid multicollinearity and to have a coherent set of indicators measuring household deprivation. Tetrachoric correlations estimate the correlation between two theorized normally distributed latent variables from two observed binary variables (Divgi, 1979). With the aim of analysing a unique latent variable for household deprivation, the aforementioned observed variables are selected according to their tetrachoric correlations. Only pairs of items having a correlation between 0.5 and 0.9 (those in white in Table 2) are considered, in order to retain items measuring the same latent concept but which are not too highly correlated.

The items *Bicycle, Motorbike, Car*, and *Radio* show a weak tetrachoric correlation with the rest of the items, and have therefore been excluded from the measurement model.

Although the correlations between *Television*, *Telephone* and the retained items are sufficiently strong, they have been excluded from the measurement model on a theoretical basis. These items cannot be considered as basic needs in the context of a low-income country such as

Bolivia. Similar to relative poverty, deprivation is a relative concept, and as such it depends on the society (Runciman, 1966). For example, while a household that does not own a telephone in a high-income country can be considered deprived, it is not the case in Bolivia. On the other hand, the asset *Refrigerator* is the only one that has been retained in the measurement model. Not owning a refrigerator has been commonly considered as a lack of basic needs in the context of high-income countries (Townsend, 1979). Its importance as a manifestation of deprivation can be extended to low- or middle-income countries, due to its strong association with health outcomes. By allowing us to keep food fresh, a refrigerator can indeed be related to hygiene and diseases (Lagendijk et al., 2008).

Therefore, the six selected items for the measurement model of household deprivation are *Electricity*, *Water*, *Sanitation*, *Floor*, *Cooking fuel* and *Refrigerator*. These items have a tetrachoric correlation higher than 0.5 (Table 3), suggesting that they are manifestations of the same underlying concept.

Table 2: tetrachoric correlation matrix, all items

	Electricity	Water	Sanitation	Floor	Cooking fuel	Radio	Television	Refrigerator	Motorbike	Car	Bike	Telephone
Electricity	1											
Water	0.66	1										
Sanitation	0.65	0.5	1									
Floor	0.78	0.64	0.58	1								
Cooking fuel	0.86	0.64	0.67	0.84	1							
Radio	0.26	0.17	0.1	0.24	0.2	1						
Television	0.94	0.63	0.64	0.78	0.85	0.31	1					
Refrigerator	0.81	0.52	0.68	0.7	0.72	0.28	0.81	1				
Motorbike	0.24	-0.03	0.34	0.09	0.2	0.08	0.28	0.36	1			
Car	0.47	0.34	0.34	0.51	0.48	0.29	0.53	0.55	0.18	1		
Bike	0.02	-0.03	0.03	-0.02	-0.08	0.22	0.07	0.11	0.1	0.19	1	
Telephone	0.76	0.53	0.65	0.73	0.75	0.39	0.73	0.76	0.18	0.53	0.11	1

Table 3: tetrachoric correlation matrix, retained items only

	Electricity	Water	Sanitation	Floor	Cooking fuel	Refrigerator
Electricity	1					
Water	0.66	1				
Sanitation	0.66	0.51	1			
Floor	0.78	0.64	0.59	1		
Cooking fuel	0.86	0.64	0.67	0.84	1	
Refrigerator	0.81	0.52	0.68	0.71	0.72	1

6.2) Measurement model for household deprivation

The measurement model of equation (1) can be interpreted as a single-level model. The total variance of the latent variable σ_η^2 is estimated as 19.15. The Spearman rank correlation with the DHS wealth index is high in the single-level latent variable, with a value of 0.92. This result is consistent with previous attempts to construct a latent variable for wealth (Vandemoortele, 2014).

Note that the discrimination parameter related to the item *Electricity* has been constrained to 1 for identification. As can be seen in Table 4, *Cooking fuel* and *Electricity* are the items that best discriminate between households with different deprivation scores, while *Water* and *Sanitation* have the least discriminatory power. Therefore, having electricity discerns household deprivation better than, for instance, having clean water. Moreover, *Water* and *Sanitation* are items that are more likely to be owned (those with lower values in the difficulty parameters), while *Cooking fuel* is the least likely.

Table 4: discrimination and difficulty parameters, single-level model

Item	Discr. (α_{r1})	SE (α_{r1})	Diff. (α_{r0})	SE (α_{r0})
Electricity	1.00	(constrained)	-4.05	0.12
Water	0.41	0.02	-6.39	0.06
Sanitation	0.43	0.02	-3.91	0.04
Floor	0.72	0.04	-3.00	0.07
Cooking fuel	1.02	0.06	-2.32	0.36
Refrigerator	0.57	0.03	1.86	0.04

7) Results

7.1) Empty multilevel model

The aim of the multilevel structural models of equations (2) and (3) is to analyse the distribution of the latent variable for household deprivation between and within Bolivian communities.

When taking into account the hierarchical structure of the data, the between-community variance component does not vary substantially from the total variance of the single-level models (19.51, versus 19.15 in the single-level model), while the within-community variance components is 1.77 (Table 5). The intra-community correlation, that is the proportion of variation in the latent variable explained by the grouping of households within communities, allows an assessment of the level of segregation: a high level of community-level variance reflects substantial differences in household deprivation across communities. For this model, a high proportion of variation in the latent variable (around 92%) is due to the grouping of households within communities. Thus, households within the same community have extremely similar scores on the latent variable of deprivation. This finding is consistent with previous studies: Castellanos (2007) pointed out the relatively low level of inequality among indigenous households in rural Bolivian communities. In such aggregates, income is similar across households, since it mainly relies on agriculture. The value of the Gini coefficient for income is lower when calculated for communities than for the Bolivian nation as a whole (ibid.).

A comparison between the selected SEM model and a model including the continuous DHS wealth index has been made, focusing on the differences in the partitioning of the variance. Using the DHS wealth index, the between-community variance is 0.82 and the within-community variance is 0.19, giving an intra-community correlation of 0.81. Therefore, the DHS wealth index, built with PCA, leads to an underestimation of the proportion of variation in wealth explained by the grouping of households within communities. This difference is due to the fact that the DHS wealth index does not take into account measurement error, as well as to the different selection of observed items in the construction of the two indices.

Table 5: variance decomposition, all models

	Between-community variance	**SE**	**Within-community variance**	**SE**	**Total variance**	**Intra-community correlation**
Empty multilevel	19.51	1.69	1.77	0.16	21.28	0.92
Indigenous	17.8	1.54	1.77	1.16	19.57	0.91
Male education	6.59	0.55	1.24	0.12	7.83	0.84
Regions	18.64	1.61	1.77	0.16	20.41	0.91
Distance to the closest municipal capital	14.09	1.23	1.78	0.16	15.87	0.89
Risk of drought	18.76	1.63	1.78	0.16	23.64	0.95
All	5.12	0.43	1.25	0.12	7.61	0.89

7.2) Models including contextual factors of deprivation clustering

Table 6: results for the structural models, all models

Model	Variable	Univariate models		Multivariate model	
		Coeff.	95% CI	Coeff.	95% CI
Indigenous [ref. Non indigenous]	Indigenous	-2.67	[-3.25; -2.09]	-1.47	[-1.84; -1.10]
	(LR test - vs empty model)	X^2 =86.23	d.f.=1		
Male education	Community-level mean years of male education	1.08	[0.99; 1.17]	0.92	[0.84; 1.01]
	Group mean centred years of male education	0.19	[0.17; 0.20]	0.19	[0.17; 0.20]
	(LR test - vs empty model)	X^2 =1720.81	d.f.=2		
Regions [ref. La Paz]	Chuquisaca	-0.66	[-1.76; -0.43]		
	Cochabamba	0.48	[-0.49; 1.45]		
	Oruro	-0.70	[-1.77; 0.37]		
	Potosí	-1.44	[-2.49; -0.40]		
	Tarija	0.53	[-0.56; 1.62]		
	Santa Cruz	0.76	[-0.15; 1.68]		
	Beni	-2.40	[-3.77; -1.13]		
	Pando	0.94	[-0.72; 2.60]		
	(LR test - vs empty model)	X^2 =40.31	d.f.=8		
Distance to the closest municipal capital	Distance	-0.19	[-0.21; -0.16]	-0.07	[-0.08; -0.05]
	(LR test - vs empty model)	X^2 =269.15	d.f.=1		
Risk of drought [ref. High]	Very low	2.14	[-0.46; 4.73]	0.30	[-1.16; 1.70]
	Low	4.92	[2.33; 7.52]	2.03	[0.60; 3.47]
	Medium	3.89	[1.36; 6.43]	1.53	[0.14; 2.92]
	(LR test - vs empty model)	X^2 =43.50	d.f.=3		

Table 6 shows the results of the univariate and multivariate models of equations (2) and (3). First, the coefficient of *Indigenous* is significantly negative: communities with a majority indigenous population are more likely to have higher mean deprivation. Indigenous origins are found to be associated with poverty in rural Bolivian communities (Albo, 1994; Grootaert and Narayan, 2004); the Bolivian indigenous population is mainly clustered in the *Altiplano* and *Valle* regions in isolated rural communities, with a subsequent lack of roads, access to markets, and social infrastructure (Buzaglo and Calzadilla, 2009). The lack of social welfare programmes leads to a high vulnerability to shocks such as droughts and floods (ibid.). Moreover, a high proportion of indigenous people have no formal education (Castellanos, 2007). Therefore, due to their disadvantaged position, the concentration of indigenous households in certain areas leads to the clustering of deprivation. Moreover, Zoomers (2006) highlights the fact that heterogeneity exists not only between indigenous and non-indigenous communities, but also between neighbouring indigenous villages depending on agricultural activity, arising from differing access to irrigation water and roads.

Second, while including the variable for community mean male education, as explained earlier in equation (3), the group mean centred household-level education is also included in the model, in order to separate the between- and within-community effects. Both coefficients related to male education are significant and positive. The between effect indicates that the higher the mean level of male education within a community, the lower the mean level of deprivation of that community. While the role of parental education in explaining the socioeconomic status of a household is well established (Cornia, 2014; King and Hill, 1993), the mean level of education within a community can also be interpreted as a contextual factor explaining the variation in the level of deprivation across communities. Education underlies a broad range of socioeconomic factors, including lower economic conditions (Wight et al., 2006), leading to deprivation clustering. However, this relationship can have variations within the Bolivian territory. For instance, Punch (2004) pointed out that education might have a minor role in predicting socioeconomic status for young people in Bolivian rural areas close to the Argentinian border, where migration can have a bigger impact on occupational status, and consequently on the level of deprivation within the community. The multivariate model in this paper indicates that education is associated with clustering of deprivation while also taking into account ethnicity. While including the variable for community mean male education, a drop in the total variance is observed, due to a large decrease in the between-community variance component: the inclusion of the contextual variable for mean community level of education explains 64.2% of the between-community variance (Table 5). A sensitivity analysis including female instead of male education has been carried out. Results of this model, shown in Appendix 1, indicate no substantial differences in comparison to the model including male education.

Third, two regions, Potosí and Beni, have a significantly higher level of deprivation than La Paz. The territory of Potosí, located in the South-West of the country (Figure 1), is mainly mountainous, posing issues of accessibility, as well as difficulties in promoting extensive agricultural exploitation. This region presents the highest presence of indigenous population (Castellanos, 2007), and has been affected several times by severe drought (Gray-Molina et al., 2002). Beni's case is different: this region is rich in raw materials and represents one of the biggest agricultural centres in Bolivia (Vadez et al., 2004). Despite its richness in natural resources, the level of poverty is still high, being a mainly rural territory, lacking big urban centres and being in a logistically marginal area when compared to the leading Bolivian economic poles (Weisbrot and Sandoval, 2008).

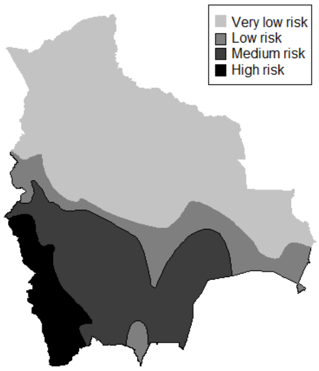
Figure 1: Bolivian regions with a significantly higher level of deprivation than La Paz

As a fourth result, the coefficient of *Distance to municipal capital* is significantly positive: every additional kilometre of distance from the closest municipal capital is associated with an average decrease of 0.18 in the community-level score of the latent variable for household deprivation. On average, rural villages have higher mean deprivation. The time series of the Bolivian poverty headcount ratio highlights the persistency of deprivation in rural areas; while it decreased by 23.4% between 2000 and 2014 in urban areas, the extent of the decrease has been only 15.4% in rural areas (my own elaboration on World Bank (2014) data). Rural populations are strongly dependent on farming productivity, which leads to a high vulnerability to shocks such as drought or flooding (Castellanos, 2007). Access to the market for agricultural goods might also be limited by geographical isolation and lack of roads, leading to a strong dependency on intermediaries that, in the presence of asymmetrical information, can affect peasants' income (Buzaglo and Calzadilla, 2009). Due to the distribution of Bolivian's rural population over extended mountainous and forested areas, issues of geographical accessibility contribute to the high cost of the extension of basic services to the totality of the population (Andersen, 2002). 26.7% of rural households retrieve water from a source considered unsafe, and 56.7% lack basic sanitation services (against, respectively, 5.4 and 9.3% in urban areas) (Coa and Ochoa, 2008). This exposes rural populations to endemic diseases that can affect labour productivity and consequently levels of deprivation (Buzaglo and Calzadilla, 2009).

Little impact on the within-community variance component is observed after the inclusion of the variable *Distance to municipal capital*, while a substantial drop of almost one-third is observed for the between-community component in comparison to the empty multilevel model (Table 5). Given the random error introduced in the DHS GPS datasets, a sensitivity analysis using the binary variable for urban or rural place of living provided in the DHS dataset instead of the distance to distance to the closest municipal capital has been carried out (results in Appendix 2). No substantial differences in comparison to the model including the continuous variable are observed.

Moreover, the coefficients indicate that the communities located in the medium- and low-risk areas of drought have a lower mean level of deprivation than the communities in areas of high risk. Prolonged droughts have affected the South-West part of Bolivia, causing soil erosion and reducing the presence of vegetation (Kessler and Stroosnijder, 2006). These phenomena have a great impact on rural populations, which strongly rely on farming and livestock, and is the main cause of the drop in agricultural productivity observed over the last decades in Bolivia (Benton, 1993). Climate change has triggered rural-urban migrations; a rapid process of urbanization has been observed in Bolivia between the 1980s and the 2000s (World Bank, 2015). Punch (2004) observed that in a rural Bolivian village in Tarija (located in the area at medium risk of drought) migration rather than education is considered the best way to improve living standards, since migrant work offers more security and immediate benefits. Rural-urban migration is associated with the uneven residential sorting of the migrants within the urban environment, increasing the level of urban residential segregation. In my models, the only non-significant difference in mean levels of deprivation has been observed between the low-risk and the high-risk areas. This can be partially explained by analysing the rural-urban migration flows. Balderrama's (2011) study on migration from the rural Northern Potosí area identified four main destination areas: Cochabamba (for construction works), Llallagua (mining, construction work, trade and education), Huanuni (mining), Santa Cruz (construction work). None of these destinations are located in the areas at very low risk of drought; the first three are in the medium risk territory, while Santa Cruz is at low risk. Therefore, selective migration might be the reason for the significant difference in community means of deprivation between low- and medium-risk communities and communities belonging to high-risk areas. Due to the random error introduced in the DHS GPS datasets, some communities might have been misplaced into a different area; Appendix 3 presents a sensitivity analysis in which those communities are allocated to the closes area of risk of drought; no substantial difference is found in the results.

Figure 2: risk of drought in Bolivia



Little difference is found in the multivariate models simultaneously including these variables: rural, indigenous communities with a lower mean level of male education and at higher risk of drought are significantly more likely to have higher mean deprivation. *Region* has not been included in the model, since it is highly correlated with *Risk of drought*: the areas of risk overlap with many of the Bolivian regions. For instance, the whole area at high risk of drought is included in the regions of Oruro and Potosí. Therefore, *Risk of drought* is preferred because of its higher theoretical value as a potential explanation for clustering of deprivation within communities, being a cause of selective rural-urban migration (Balderrama, 2011).

8) Discussion

Bolivia in 2008 presented among the highest indicators of poverty and deprivation in Latin America (Coa and Ochoa, 2009). This paper explores the distribution of deprivation within the country, with a focus on the contextual factors affecting the geographical clustering of deprivation. Deprivation clustering manifests itself when more deprived households are isolated and physically separated into certain areas, and is associated with social exclusion (Gray-Molina et al., 2002).

By analysing 2008 DHS data, a latent variable for deprivation is created from a set of six observed items (electricity, water, sanitation, floor, cooking fuel and refrigerator), and simultaneously included in the SEM models, overcoming issues related to measurement error (Muthén and Muthén, 2010). The multilevel structure of the data allows us to investigate the extent of the clustering of deprivation within Bolivian communities, and the inclusion of contextual variables to predict the variation in the level of deprivation across communities. This paper contributes to the study of measures of poverty clustering, since a multilevel structural equation modelling approach allows us to overcome issues related to the measurement error and to make statistical inference on segregation. This analysis involves a continuous latent variable as an outcome, and therefore represents an extension of the multilevel models used in previous work (Goldstein and Noden, 2003; Leckie et al., 2012). Moreover, this analysis involves the investigation of the correlation matrix of the items used in the construction of the latent variable for household deprivation, ensuring that the selected items measure the underlying concept of household deprivation. This analysis highlights the differences in the use of the latent variable in comparison to the DHS wealth index; the inclusion of this latter measure leads to an underestimation of the magnitude of the clustering of deprivation in Bolivia, since the DHS wealth index does not take into account measurement error and the items used in the construction of the two indices are slightly different.

Bolivia is found to have a high level of clustering of deprivation, since the main source of variation in deprivation arises from differences across communities, rather than within communities. Ethnicity, education, administrative region, distance to urban centres and drought-induced migration are found to explain differences in the mean level of deprivation across Bolivian villages. This paper has implications for social and health policies. By identifying the contextual factors associated with the clustering of deprivation, this paper provides evidence on the mechanisms leading to economic and social segregation. This analysis helps in identifying clusters of deprivation within Bolivia, and highlights crucial sectors to be developed in order to fight spatial unevenness in the distribution of wealth, linked to social exclusion, diminished opportunities for human capital development and lower access to public services. Finally, reducing inequality across Bolivian communities could also positively affect health indicators, since contexts of concentrated deprivation are associated with higher mortality and higher exposure to infectious diseases (Fiscella and Franks, 1997; Szwarcwald et al., 2002).

9) Acknowledgements

The author would like to thank Professor Arjan Gjonca for his extensive and helpful comments on earlier drafts and Professor Fiona Steele for his advice on statistical methods.

The study was funded by the ESRC under Grant no. 201557818/1-23-A000. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the paper.

References

Albo, X. (1994). Ethnic violence: the case of Bolivia; *in The culture of violence, 119-143, The United Nations University, Shibuya-ku, Tokio*

Amarasinghe, U., Samad, M., & Anputhas, M. (2005). Spatial clustering of rural poverty and food insecurity in Sri Lanka. *Food Policy, 30*(5), 493-509.

Andersen, L. E. (2002). Rural-urban migration in Bolivia: Advantages and disadvantages. *Documento de Trabajo, Instituto de Investigaciones Socio-Económicas, Universidad Católica Boliviana. (No. 05/02).*

Balderrama, C. (2011). Rural migration in Bolivia: the impact of climate change, economic crisis and state policy. *IIED; No. 31*.

Bartholomew, D. J., Knott, M., & Moustaki, I. (2011). Latent variable models and factor analysis: A unified approach. *John Wiley & Sons*; *Vol. 904.*

Bell, W. (1954). A probability model for the measurement of ecological segregation. *Social Forces*, 357-364.

Benton, J. (1993). The role of womens organisations and groups in community development. A case study of Bolivia. *USAid Working Papers; Document No. 127039; retrieved from: https://www.popline.org/node/339234*

Borjas, G. J., & Tienda, M. (1987). The economic consequences of immigration. *Science, 235*, 645-652.

Buzaglo, J., & Calzadilla, A. (2009). Towards a new consensus: poverty reduction strategies for Bolivia. *International Journal of Development Issues, 8*(1), 18-39.

Castellanos, I. O. V. (2007). Extreme Poverty: Vulnerability and Coping Strategies Among Indigenous People in Rural Areas of Bolivia. *Ecology and Development Series No 51*

Cliff, A., & Ord, J. K. (1973). Spatial autocorrelation. *Taylor & Francis; No. 04; QA278. 2, C5*.

Coa, R., & Ochoa, L. H. (2009). Bolivia 2008: Results from the Demographic and Health Survey. *Studies in Family Planning, 41*(4), 301-306.

Cornia, G. A. (2014). Falling inequality in Latin America: policy changes and lessons. *OUP Oxford*.

Crandall, M. S., & Weber, B. A. (2004). Local social and economic conditions, spatial concentrations of poverty, and poverty dynamics. *American Journal of Agricultural Economics, 86*(5), 1276-1281.

Divgi, D. (1979). Calculation of the tetrachoric correlation coefficient. *Psychometrika, 44*(2), 169-172.

Duncan, O. D., & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review, 20*(2), 210-217.

Duncan, D. T., Aldstadt, J., Whalen, J., White, K., Castro, M. C., & Williams, D. R. (2012). Space, race, and poverty: Spatial inequalities in walkable neighborhood amenities?. *Demographic Research*, *26*(17), 409.

Filmer, Deon & Kinnon, Scott (2012). Assessing asset indices. *Demography* 49.1: 359-392.

Fiscella, Kevin, and Peter Franks. "Poverty or income inequality as predictor of mortality: longitudinal cohort study." *Bmj* 314.7096 (1997): 1724.

GeoBolivia. (2013). Red vial fundamental de Bolivia 2012. *Retrieved from*: *https://geo.gob.bo/geonetwork/srv/esp/catalog.search#/metadata/9c141353-e99b-44c9-b121-6f8fe75ac240*

GeoBolivia. (2017a). Capitales municipales de Bolivia (339 municipios), 2013. *Retrieved from*: *https://geo.gob.bo/download/?w=fondos&l=CapitalesMunicipales*

GeoBolivia. (2017b). Mapa de amenaza de sequía meteorológica. *Retrieved from*: *https://geo.gob.bo/geonetwork/srv/esp/catalog.search#/metadata/e304b259-8dcf-4886-b0f8-d0bcf99c02d0*

Goldstein, H., & Noden, P. (2003). Modelling social segregation. *Oxford Review of Education, 29*(2), 225-237.

Gray-Molina, G., Jimenez, W. L., & de Rada, E. P. (2002). Social exclusion: residential segregation in Bolivian cities. *Inter-American Development Bank; Research Network working papers; R-440.*

Greene, R. (1991). Poverty concentration measures and the urban underclass. *Economic Geography*, *67*(3), 240-252.

Grootaert, C., & Narayan, D. (2004). Local institutions, poverty and household welfare in Bolivia. *World Development, 32*(7), 1179-1198.

Günther, I., & Fink, G. Water, sanitation and children's health: evidence from 172 DHS surveys. *World Bank Policy Research Working Paper Series, Vol (2010).*

Hammel, E. A., Mason, C., & Stevanović, M. (2010). A fish stinks from the head: Ethnic diversity, segregation, and the collapse of Yugoslavia. *Demographic Research*, *22*, 1097-1142.

Howe, L. D., Hargreaves, J. R., & Huttly, S. R. (2008). Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging themes in epidemiology, 5*(1), 1.

Iceland, J., Weinberg, D., & Hughes, L. (2014). The residential segregation of detailed Hispanic and Asian groups in the United States: 1980-2010. *Demographic Research*, *31*, 593.

Jencks, C., & Mayer, S. E. (1990). The social consequences of growing up in a poor neighborhood. *Inner-city poverty in the United States, 111*, 186.

Jolliffe, I. T. (1986). Principal Component Analysis and Factor Analysis. *Principal component analysis* (pp. 115-128): Springer.

Kawachi, I., & Berkman, L. F. (2003). Neighborhoods and health. *Oxford University Press*.

Kessler, C., & Stroosnijder, L. (2006). Land degradation assessment by farmers in Bolivian mountain valleys. *Land Degradation & Development, 17*(3), 235-248.

King, E. M., & Hill, M. A. (1993). Women's education in developing countries: barriers, benefits, and policies. *Baltimore: Baltimore : Published for the World Bank by the John Hopkins University Press*.

Kolenikov, S., & Angeles, G. (2004). The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices. *Chapel Hill: Carolina Population Center, University of North Carolina*, 1-59.

Lagendijk, E., Assere, A., Derens, E., & Carpentier, B. (2008). Domestic refrigeration practices with emphasis on hygiene: analysis of a survey and consumer recommendations. *Journal of Food Protection, 71*(9), 1898-1904.

Leckie, G., Pillinger, R., Jones, K., & Goldstein, H. (2012). Multilevel Modeling of Social Segregation. *Journal of Educational and Behavioral Statistics, 37*(1), 3-30.

Lieberson, S., Peach, C., Robinson, V., & Smith, S. (1981). An asymmetrical approach to segregation. *Ethnic segregation in cities*.

Massey, D. S., & Denton, N. A. (1988). The dimensions of residential segregation. *Social Forces, 67*(2), 281-315.

Matthews, S. A., & Parker, D. M. (2013). Progress in spatial demography. *Demographic Research, 28*, 271-312.

Michels, A. (2011). Migration and inheritance practices in the Bolivian Altiplano. *Working paper/World Institute for Development Economics Research*.

Muthén, L., & Muthén, B. (2010). Mplus software (version 6). *Los Angeles, CA: Muthén & Muthén*.

Paulson, S., Gisbert, M. E., & Quitón, M. (1996). Case studies of two womens health projects in Bolivia. *USAid working papers; Document No.127864; https://www.popline.org/node/306664*

Perez-Heydrich, C., Warren, J. L., Burgert, C. R., & Emch, M. (2013). Guidelines on the use of DHS GPS data. *ICF International*.

Population Reference Bureau. (2013). 2013 World Population Data Sheet. *Retrieved from: www.prb.org/pdf13/2013-population-data-sheet_eng.pdf.*

Psacharopoulos, G. (1993). Ethnicity, education, and earnings in Bolivia and Guatemala. *Comparative Education Review, 37*(1), 9-20.

Punch, S. (2004). The impact of primary education on school-to-work transitions for young people in rural Bolivia. *Youth & Society, 36*(2), 163-182.

Robson, M.G., Stephenson, R., and Elfstrom, K.M. (2012) "Community Influences on Antenatal and Delivery Care in Bangladesh, Egypt, and Rwanda." *Public Health Reports* 127.1: 96-106.

Robusto, C. C. (1957). The cosine-haversine formula. *The American Mathematical Monthly, 64*(1), 38-40.

Runciman, W. G. (1966). Relative deprivation and social justice : a study of attitudes to social inequality in twentieth-century England. *London: Routledge & Kegan Paul*.

Rutstein, S., & Johnson, K. (2004). The DHS wealth index. *ORC Macro, Measure DHS*.

Schroeder, K. (2007). Economic globalization and Bolivia's regional divide. *Journal of Latin American Geography*, *6*(2), 99-120.

Schoen, R., & Nau, C. (2008). A behaviorally-based approach to measuring inequality. *Demographic Research*, *19*, 1727-1748.

Silver, H. (1994). Social exclusion and social solidarity: three paradigms. *International l Lab. Review, 133*, 531.

Singh, A., Vainchtein, D., & Weiss, H. (2009). Schelling's segregation model: parameters, scaling, and aggregation. *Demographic Research*, *21*, 341-366.

Skiles, M. P., Burgert, C. R., Curtis, S. L., & Spencer, J. (2013). Geographically linking population and facility surveys: methodological considerations. *Population health metrics*, *11*(1), 14.

Snijders, T. A. B., & Bosker, R. J. (2012). Multilevel analysis: an introduction to basic and advanced multilevel modeling. *International Encyclopedia of Statistical Science*. Springer Berlin Heidelberg, 2011. 879-882..

Sparks, P. J., Sparks, C. S., & Campbell, J. J. (2013). An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the US. *GeoJournal, 78*(2), 389-405.

StataCorp. (2013). Structural Equation Modelling Reference Manual, Release 13. *Stata Journal, College Station, TX: StataCorp LP*.

Stretesky, P. B., Schuck, A. M., & Hogan, M. J. (2004). Space matters: An analysis of poverty, poverty clustering, and violent crime. *Justice Quarterly*, *21*(4), 817-841.

Szwarcwald, C. L., Andrade, C. L. T. d., & Bastos, F. I. (2002). Income inequality, residential poverty clustering and infant mortality: a study in Rio de Janeiro, Brazil. *Social science & medicine, 55*(12), 2083-2092.

Theil, H., & Theil, H. (1972). Statistical decomposition analysis; with applications in the social and administrative sciences. *London North-Holland HA33 No4*.

Thomas, D. (1990). Intra- Household Resource Allocation: An Inferential Approach. *The Journal of Human Resources*, *25*(4), 635-664.

Townsend, P. (1979). Poverty in the United Kingdom: a survey of household resources and standards of living. *University of California Press*.

US Aid. (2012). Sampling and household listing manual. *DHS Toolkit of methodology for the MEASURE DHS Phase III project*.

Uthman O.A., Moradi T., Lawoko S. (2011) Are Individual and Community Acceptance and Witnessing of Intimate Partner Violence Related to Its Occurrence? Multilevel Structural Equation Model. PLoS ONE 6(12): e27738.

Vadez, V., Reyes-García, V., Godoy, R. A., Apaza, V. L., Byron, E., Huanca, T., Wilkie, D. (2004). Does integration to the market threaten agricultural diversity? Panel and cross-sectional data from a horticultural-foraging society in the Bolivian Amazon. *Human Ecology,* 635-646.

Vandemoortele, M. (2014). Measuring Household Wealth with Latent Trait Modelling: An Application to Malawian DHS Data. *Social indicators research 118, no.2: 877-891*

Weisbrot, M., & Sandoval, L. (2008). The distribution of Bolivia's most important natural resources and the autonomy conflicts. *Center for Economic and Policy Research*, 2.

WHO & UNICEF (2014). Progress on drinking water and sanitation: 2014 Update. *World Health Organization. Retrieved from: http://www.who.int/water_sanitation_health/publications/jmp_report-2012/en/*

Wiggins, S., Schejtman, A., & Gray, G. (2006). Bolivia case study: an interpretative summary. *Research programme consortium on improving institutions for pro-poor growth. Working paper series n. 1*.

Wight, R. G., Aneshensel, C. S., D., M.-M., Botticello, A. L., Cummings, J. R., Karlamangla, A. S., & Seeman, T. E. (2006). Urban neighborhood context, educational attainment, and cognitive function among older adults. *American journal of epidemiology, 163.12* 1071-1078.

Woods, R. (2003). Urban- Rural Mortality Differentials: An Unresolved Debate. *Population and Development Review, 29*(1), 29-46.

World Bank. (2014). Bolivian data. *Retrieved from: http://data.worldbank.org/country/bolivia*.

World Bank. (2015). Urbanization trends in Bolivia, opportunities and challenges. *Global Program Unit, May 2015*.

Zoomers, A. (2006). Pro-Indigenous Reforms in Bolivia: Is there an Andean Way to Escape Poverty? *Development and Change, 37*(5), 1023-1046.

Appendix 1: sensitivity analysis for covariate *Female education*

In the model including *Female education*, the between-community variance and the intra-community correlation are slightly lower than those of the model including *Male education* (Table 7). The coefficients of both between and within effects in the multivariate model are very similar to those referring to *Male education* (Table 8).

Table 7: variance decomposition, model including community-*level mean female education*

	Between-community variance	**SE**	**Within-community variance**	**SE**	**Total variance**	**Intra-community correlation**
All (including *Female education*)	4.42	0.37	1.17	0.11	5.59	0.79

Table 8: results for the structural multivariate model including community-*level mean female education*

Model	**Variable**	**Coeff.**	**95% CI**
Indigenous	Indigenous	-0.41	[-0.76; -0.07]
Female education	Community-level mean years of female education	0.93	[0.85; 1.01]
	Group mean centred years of female education	0.19	[0.17; 0.20]
Distance to the closest municipal capital	Distance	-0.07	[-0.08; -0.05]
Risk of drought [ref. High]	Very low	0.69	[-0.64; 2.03]
	Low	2.70	[1.35; 4.05]
	Medium	1.97	[0.66; 3.27]

Appendix 2: sensitivity analysis for covariate *Urban*

In the model including *Urban*, the intra-community correlation is lower than that of the model including *Distance to municipal capital* (Table 9), but the coefficients of all the covariates included in the multivariate model are very similar to those of the model including *Distance to municipal capital* (Table 10): urban communities present a significantly lower mean deprivation.

Table 9: variance decomposition, model including *Urban* instead of *Distance to municipal capital*

	Between-community variance	**SE**	**Within-community variance**	**SE**	**Total variance**	**Intra-community correlation**
Female education	3.44	0.29	1.26	0.12	4.70	0.73

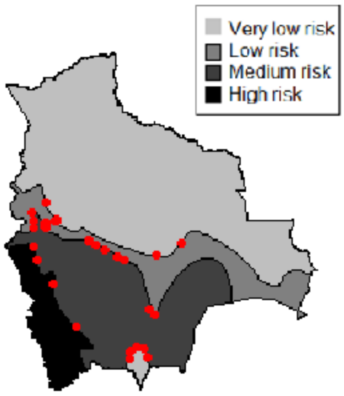
Table 10: results for the structural multivariate model including *Urban* instead of *Distance to municipal capital*

Model	Variable	Coeff.	95% CI
Indigenous	Indigenous	-1.09	[-1.41; -0.78]
Male education	Community-level mean years of male education	0.59	[0.52; 0.66]
	Group mean centred years of male education	0.19	[0.17; 0.20]
Rurality [ref. Rural]	Urban	3.83	[3.41; 4.26]
Risk of drought [ref. High]	Very low	-0.54	[-1.74; 0.65]
	Low	1.30	[0.11; 2.50]
	Medium	0.60	[-0.55; 1.76]

Appendix 3: sensitivity analysis changing categorization of *Risk of drought* for communities within 5 kilometres from the border between different areas

Figure 3 shows the position of the 46 communities that lie within 5 kilometres from the borders between areas at different risk of drought.

Figure 3: comunities within 5km from the border between areas at different risk of drought



Tables 11 and 12 present the result from the multivariate model in which the categorization of these communities has been modified to the adjacent area of risk of drought. The intra-community correlation is very similar to that of the original model, and no substantial difference is observed in the magnitude and significance of the coefficients of all the covariates included in the multivariate model.

Table 11: variance decomposition, model changing categorization of *Risk of drought* for communities within 5km from the border between different areas

	Between-community variance	SE	Within-community variance	SE	Total variance	Intra-community correlation
All (including modified *Risk of drought*)	5.12	0.42	1.25	0.12	6.37	0.80

Table 12: results for the structural multivariate model changing categorization of *Risk of drought* for communities within 5km from the border between different areas

Model	Variable	Coeff.	95% CI
Indigenous	Indigenous	-1.47	[-1.84; -1.10]
Male education	Community-level mean years of male education	0.92	[0.84; 1.01]
	Group mean centred years of male education	0.19	[0.17; 0.20]
Distance to the closest municipal capital	Distance	-0.07	[-0.08; -0.05]
Risk of drought (modified) [ref. High]	Very low	0.90	[-0.32; 2.12]
	Low	2.48	[1.26; 3.70]
	Medium	2.13	[0.96; 3.31]