Chapter 3
Composite Indicator of Poverty

As stated in the introduction, our main objective is to operationalize multidimensional poverty comparisons. After some clarification on this objective and on a first methodological choice in Section 3.1, Section 3.2 presents a quick review of the main methodologies used to build a composite indicator of poverty (CIP). Our second methodological choice takes us to a short presentation of different variants of factorial approaches and to the argument supporting our third methodological choice, the multiple correspondence analysis (MCA) technique (Section 3.3). Finally, Section 3.4 develops the MCA technique and illustrates it with a numerical case study on Vietnam.

3.1 Individual and Population Poverty Comparisons

For discussion, it is important to clarify the terminology regarding the three concepts of *poverty indicator*, *poverty measure,* and *poverty index*. Let I_{ik} be the value of indicator I_{k} for the elementary population unit i, called here individual i for simplification.[1] I_{ik} is then a *poverty indicator* value. The value I_{ik} can be transformed as $g_{\mathrm{k}}(I_{ik})$, with the function g_k, to better reflect a poverty concept relative to indicator I_{k}. This is frequently the case, especially with a quantitative indicator I_k to which is associated a poverty threshold (poverty line) z_k. A basic transformation is simply the censoring of I_{k} at z_k to get I^*_{k}. In this case, well-known transformations are $g_k(I^*_{\mathrm{ki}}) = (z_{\mathrm{k}} - I^*_{\mathrm{ik}})^{\alpha}$ or $g_k(I^*_{\mathrm{ki}}) = (1 - I^*_{\mathrm{ik}}/z_{\mathrm{k}})^{\alpha}$. Then, $g_k(I_{\mathrm{ik}})$ is called a *poverty measure* value, again defined for individual i. In the particular case where the function g_k is the identity function, the poverty indicator and the poverty measure are the same. Finally, poverty measure values can be aggregated over the units for the whole population U, as $W_k\{g_k(I_{\mathrm{ik}}), \mathrm{i} = 1, \mathrm{N}\}$. W_k is then called a *poverty index* relative to the indicator I_k for the population U. Obviously, this index W_k can be

[1] The term "elementary population unit" can refer to individuals and households as well as to villages, regions and countries.

L.-M. Asselin, *Analysis of Multidimensional Poverty,*
Economic Studies in Inequality, Social Exclusion and Well-Being 7,
DOI 10.1007/978-1-4419-0843-8_3, © Springer Science+Business Media, LLC 2009

defined on any subpopulation of U consisting of n individuals, $n \geq 1$. For $n = 1$, the poverty index is a poverty measure on each individual.

Poverty indices are required for population comparisons, while poverty indicators and poverty measures are sufficient for comparisons between individuals.

An interesting review in Maasoumi (1999) first distinguishes between the literature addressing the issue of computing a *composite index of poverty* from a multidimensional distribution of poverty indicators on a given population, and the literature aiming at defining a *composite indicator of poverty* on each unit of the given population. The first type of this literature is well represented by Bourguignon and Chakravarty (1999). The first distinction, referring to Sen (1976), is between the identification and the aggregation problems. Any individual who is below the poverty threshold for at least one of the poverty attributes included in the poverty vector is identified as poor. It is thus the union concept of poverty that is used here, in contrast with the intersection concept. The aggregation technique relies on an axiomatic approach to the desired properties of the composite index, largely based on standard axioms enunciated for a univariate poverty index, and on a composite poverty measure referring to a given poverty threshold for each primary indicator. The implicit context is thus a set of quantitative indicators and the resulting index is usually relevant only for that type of indicators. In fact, the composite poverty measure proposed by Bourguignon and Chakravarty is a CES function of the shortfalls (poverty gaps) in each of the primary poverty indicators. Since the direct focus of this approach is on a poverty index, it is called a *one-step* approach to multidimensional poverty indices.

It should be obvious, on the other hand, that solving in a first step the problem of building a numerical composite indicator of poverty opens the way to computing a composite poverty index based on the composite indicator, relying then on the univariate theory of poverty indices. This approach is designated as a *two-step* approach to multidimensional poverty indices, where the focus is mainly on justifying a methodology for the composite indicator, the most critical part of the whole process.

This two-step approach is our first methodological choice.

We are thus also taken away from the multidimensional stochastic dominance theory, extension of the well-known unidimensional one. It can be found in Duclos, Sahn, and Younger (2006). We can see it as another one-step approach, since the focus is on classes of multidimensional poverty indices. This theory releases poverty comparisons from having to make arbitrary choices of poverty lines and poverty indices by looking at the relative position of distribution functions and at identifying regions where a distribution "surface" is over or under another one. This ordinal approach presents the theoretical interest of clarifying necessary and sufficient conditions for the robustness of comparisons. Difficulties remain for the operationalization: since the identification of dominance regions is often uneasy in a two-dimensional case, we can expect difficulties when the number of primary indicators can amount to tens, and when sampling errors must be taken into account in applied work. There is obviously an important trade-off here between the degree of robustness, here placed at a high level, and the power of the dominance tests.

Due to its central role in a two-step approach, the rest of this chapter focuses on the first step, the construction of a composite indicator of poverty (CIP).

3.2 Overview of Methodologies for a Composite Indicator of Poverty (First Step)

In what follows, a *composite indicator of poverty* (CIP) C takes the value $C_i(I_{ik}, k = 1, K)$ for a given elementary population unit U_i.

3.2.1 CIP Based on Inequality Indices: Entropy Concepts, Shorrocks Index

Theil (1967) has first observed that Shannon's entropy $I_n(y)$

$$I_n(y) = \sum_{i=1}^{n} yi \log 2 \frac{1}{yi} = -\sum_{i=1}^{n} yi \log 2yi$$

where y represents the *income shares* in a population of n units, constitutes a natural measure of income equality, taking the maximal value $\log_2 n$ when every unit has the same income. The corresponding *inequality* measure is then taken as the difference between the maximal entropy (from a uniform distribution) and $I_n(y)$:

$$\log 2n - I_n(y) = \sum_{i=1}^{n} yi \log 2 \left(\frac{yi}{1/n} \right). \tag{3.1}$$

We thus observe that equation 3.1 is the Rényi information gain or divergence measure $I_1(q||p)$,[2] where we take q = y and p = {1/n}, the uniform distribution. It is called Theil's first inequality index.[3]

The pioneering work of Theil on entropy-based inequality indices has generated a search for larger classes of inequality indices, on the basis of desirable properties defined with respect to redistributions of income in a given population. In particular, the requirement of additively decomposable inequality indices has led to important results by Shorrocks (1980). He proved that the only admissible indices satisfying,

---

[2] The concept of « divergence »between two distributions belongs to information theory. It is not a metric as defined mathematically. The general expression of the Rényi divergence measure, which he calls «information gain », between two distributions q and p is $I_1(q||p) = \sum_{k=1}^{n} q_k \log_2 \frac{q_k}{p_k}$. For details, see Rényi (1966).

[3] To be more precise, Theil and other authors use the natural logarithm in base e instead of base 2; from now on, we will not specify the base.

among others, the decomposable additivity axiom belong to the following class[4]:

$$I_\gamma(y) = \frac{1}{n}\frac{1}{\gamma(\gamma-1)}\sum_{i=1}^{n}\left[\left(\frac{y_i}{1/n}\right)^{\gamma} - 1\right] \text{ for } \gamma \neq 0.1, \tag{3.2}$$

$$I_\gamma^1(y) = \sum_{i=1}^{n} y_i \log\left(\frac{y_i}{1/n}\right).$$

$$I_\gamma^0(y) = \sum_{i=1}^{n} 1/n \log\left(\frac{1/n}{y_i}\right)$$

Observe that equation 3.2 can be written as

$$I_\gamma(y) = \frac{1}{\gamma(\gamma-1)}\sum_{i=1}^{n} y_i\left[\left(\frac{y_i}{1/n}\right)^{\gamma-1} - 1\right]. \tag{3.2'}$$

Obviously, I_γ^1(y) is Theil's first inequality index, and I_γ^0(y) is his second inequality index. This γ-class of entropy-based inequality indices is called the class of *Generalized Entropy* indices.

What we highlight here is that this axiomatic development of inequality indices generates a class of divergence measures including, as a particular case, the Rényi's information gain measure I_1(q||p). In fact, the case γ =1 corresponds to I_1(q||p) where we take p= {1/n}, and the case $\gamma = 0$ corresponds to I_1(q||p) where we take q = {1/n}. The γ-class of inequality indices is an *asymmetric* measure of divergence between a distribution y and the uniform distribution p = {1/n}.

Maasoumi (1986) relies on these developments of information theory to propose his entropy approach to the composite indicator problem. He looks for a general inter-distributional distance as a basis to derive the composite indicator C from an optimization criterion. Let us observe that the Generalized Entropy index 3.2' generates a divergence measure between *any* two distributions x and y if we substitute a distribution x to the uniform distribution {1/n} appearing as the denominator. This is precisely the divergence measure taken by Maasoumi as the distance between the composite indicator we are looking for, C, and any one of the primary indicators I_k, k = 1, K. We thus have

$$D_\gamma(C, I_k) = \frac{1}{\gamma(\gamma-1)}\sum_{i=1}^{n} C_i\left[\left(\frac{C_i}{I_{ik}}\right)^{\gamma-1} - 1\right] \quad \text{for } \gamma \neq 0.1 \tag{3.3}$$

and obtain Theil's first and second measures for γ = 1 and 0 respectively.

[4] We write the indices directly in terms of income shares instead of using mean income μ, in order to keep more clearly the link with the theory of distributions.

Maasoumi then proposes to define the optimal indicator as the C that minimizes a *weighted sum of the pairwise divergences*, i.e., the C that minimizes

$$D_\gamma(C, I; \delta) = \sum_{k=1}^{K} \delta_k \left\{ \frac{\sum_{i=1}^{N} C_i \left[\left(\frac{C_i}{I_{ik}} \right)^{\gamma-1} - 1 \right]}{\gamma(\gamma - 1)} \right\} \tag{3.4}$$

where the δ_k are *arbitrary* weights on the divergence component relative to the indicator I_k, $\sum \delta_k = 1$.[5]

By minimizing the divergence $D_\beta(C, I, \delta)$ for the function C, Maasoumi finds the following functional form for the composite indicator:

$$C_i = \left(\sum_{k=1}^{K} \delta_k I_{ik}^{-\gamma} \right)^{-1/\gamma} \qquad \gamma \neq 0, -1 \tag{3.5}$$

We recognize here a CES function. For the two specific values $\gamma = 0, -1$, the functional forms are

$$C_i = \prod_{k=1}^{K} I_{ik}^{\delta_k}, \text{ for } \gamma = 0. \tag{3.6}$$

$$C_i = \sum_{k=1}^{K} \delta_k I_{ik}, \text{ for } \gamma = -1. \tag{3.7}$$

Conclusion on the entropy inequality indices approach

1. The whole context of entropy inequality indices, including the associated divergence concept, refers to probability distributions, i.e., to numerical measures taking values in the interval (0.1). Thus, and as can be seen particularly from the divergence measure generated by the Generalized Entropy Index, the natural domain of application for our problem is a set of meaningful numerical indicators, i.e., of quantitative poverty indicators, expressed in terms of "shares," so that the individual value I_{ik} is in the interval (0.1). The money-metric type of poverty indicators, once transformed in individual shares, appears as the domain of validity of a functional form like equation 3.5.
2. There is an important source of indetermination with the parametric nature of the Maasoumi composite indicator. On what basis should we choose the parameter

[5] The parametrization used by Maasoumi for the γ-class is slightly different from Shorrocks's one, followed here until now. Maasoumi's parameter γ is Shorrocks's -1. From now on, we will use Maasoumi's γ.

value for the γ-Generalized Entropy indices? A strong point can be made for the values $\gamma = 1$ and $\gamma = 0$, which provide a simple linear (log-linear) form.[6]

3. If the weighting approach is maintained for the optimization criterion, obviously there remains the problem of determining the weights δ_k in a nonarbitrary way. There is in fact an optimal system of weights for the functional form (B), as Maasoumi (1999) has himself observed: the basic factorial method of principal components. This is precisely the type of methods that is reviewed below.

3.2.2 CIP Based on Poverty Structure Analysis: Inertia Concepts, Factorial Approaches

To a K-dimensional poverty vector is associated a K-dimensional distribution. In some sense, the previous approach looks at the marginal distributions of the primary indicators I_k. A kind of distance between these marginal distributions, the divergence measure, serves as a basis for identifying a "mean" distribution which provides the CIP. It is like looking at the multidimensional distribution from outside, from an external viewpoint. Another viewpoint is to look at the distribution from inside, trying to identify the numerous associations between the poverty dimensions determining the global form of the poverty "mass" dispersion. It is a search for a poverty structure, an internal viewpoint. Intuitively, this is what any factorial technique tries to operationalize, relying on the central concept of inertia which is in fact a measure of the global dispersion of the distribution. Going through this structural analysis, we can hope to come out with a CIP summarizing the most relevant information identified in the distribution.

This structural approach to multidimensional poverty analysis can be seen as an empirical step to implement the analysis of interconnections between different freedoms that Sen calls for to assess the effectiveness of development.[7]

Let us consider an example with two numerical indicators, x_1, money income, and x_2, area of agricultural land. Simply by representing the data in the R^2-space of individuals (here, households), we could see figures such as Figs. 3.1, 3.2, 3.3, and 3.4.[8]

In Case 1, money income and land area are perfectly correlated. It could be approximately observed in a highly agricultural country with an easy access to markets. An obvious way to rank the households with just one number is to use for each of them their relative position on line Δ. This position, y, is given by the linear function

$$y = \beta_1 X_1 + \beta_2 X_2 \tag{3.8}$$

[6] On this, see Asselin (2002) for a more extensive review of the entropy and information theory approach.

[7] Sen (1999), p.4

[8] Variables are supposed to be centered.

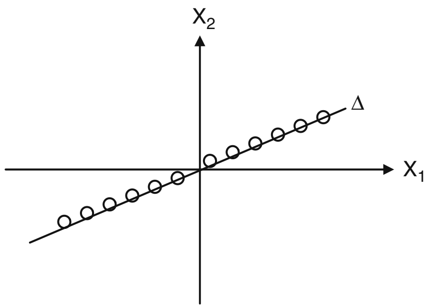


Fig. 3.1 Case 1

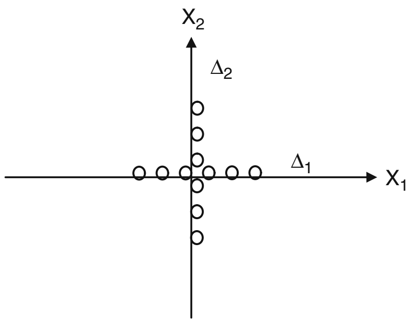


Fig. 3.2 Case 2

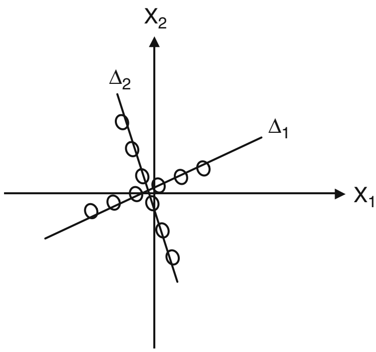


Fig. 3.3 Case 3

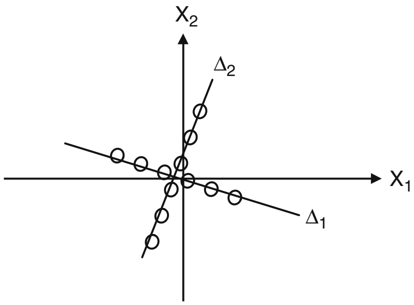


Fig. 3.4 Case 4

where the unit vector $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ identifies the support of line Δ, $\beta_1 > 0$, $\beta_2 > 0$. Equation 3.8 is thus a very relevant CIP. Cases 2, 3, and 4 are clearly situations where there are two groups of households. Looking further at the two groups of households on each line, we could discover that on Δ_1 lie urban households whereas on Δ_2 we find rural households. Lines Δ_1 and Δ_2 would then express two types of poverty: urban and rural poverty. The extreme situation of Case 2 could plausibly correspond to a country where the rural area is completely disconnected from markets. In Case 2, using the position on line Δ_1 as a global poverty indicator would formally be acceptable but would then not allow to discriminate between rural households. The same is true with line Δ_2 and urban households then being not discriminated. A better composite indicator should be proposed using both lines. Cases 3 and 4 are intermediate situations. In Case 3, the line Δ_1 could be eligible as a global poverty indicator, even if it does not discriminate between rural households. The vector β has the expected positive signs, and the fact is that rural households compensate a lower area of land with a higher money income, maybe by selling their labor force to larger farms. In case 4, the line Δ_2 could be eligible, with a positive vector β and no discrimination between urban households, these possibly leaving agricultural production for better opportunities in the labor market. But even in cases 3 and 4, a deeper analysis could suggest a better composite indicator than line $\Delta_1$or Δ_2. Finally, Case 5 shows a situation where the position on line Δ cannot be taken as a composite indicator of poverty, due to the negative sign of β_2. But it can be seen that there is more dispersion in x_1 than in x_2 and this fact can eventually be exploited (Fig. 3.5).

This internal visualization of the multidimensional distribution from the individual (household) viewpoint, i.e., from the line-points of the data matrix X[9] seen in the R^2-space, has a counterpart from the variable viewpoint, i.e., from the column-points seen in the R^N-space, where N is the number of individuals in the distribution. In our case, with $N = 12$ individuals, the two column-vectors determine a plan

[9] The convention used here is that in a matrix X, lines correspond to the statistical units (individuals) and columns to the variables (indicators).

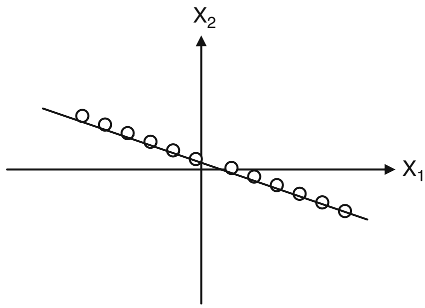


Fig. 3.5 Case 5

in the R^{12}-space, which degenerates in a single line in Cases 1 and 5. Figure 3.6 summarizes Cases 1–5.

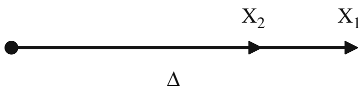


Fig. 3.6 Case 1

All these examples suggest that detecting poverty structures, let us say poverty types, through lines like Δ, Δ_1, and Δ_2 can be seen as a promising approach to know more about the real multidimensionality in a given population and to the emergence of a relevant composite indicator of this poverty. This is precisely what a factorial

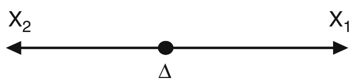


Fig. 3.7 Case 5

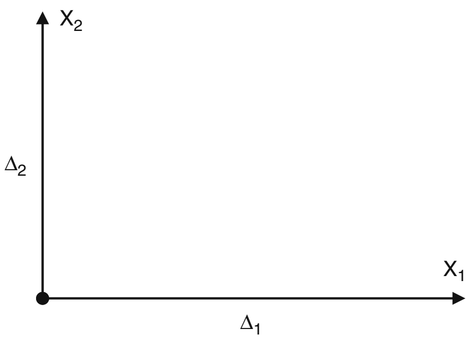


Fig. 3.8 Case 2, 3, 5

approach tries to systematize. Principal component analysis (PCA) is now taken as an example of what can be achieved in that direction (Figs. 3.7 and 3.8).

Essentially, PCA consists in building a sequence of uncorrelated (orthogonal) and normalized linear combinations of input variables (K primary indicators), exhausting the whole variability of the set of input variables, named "total variance" and defined as the trace of their covariance matrix, thus the sum of the K variances. These uncorrelated linear combinations, in fact the lines Δ above and their related unitary vectors β, are latent variables called "components." The optimality in the process comes from the fact that the first component looked for has a maximal variance λ_1, the basic idea being to visualize the whole set of data in reduced spaces capturing most of the relevant information.

Let X(N,K) be the data matrix giving the distribution of the K numerical, centered, primary poverty indicators, $K < N$. From now on, let W be the normalized (unitary) K-dimensional vector[10] previously identified as β, and let $\Sigma = X'X$ be the covariance matrix. The problem of estimating the first component consists in finding a linear combination XW such that $W'\Sigma W$ is maximal under the constraint $W'W = 1$. With λ as the Lagrange multiplier, the problem consists in solving the equation

$$(\Sigma - \lambda I)W = 0 \tag{3.9}$$

where I is the unit (K,K) matrix. There are different ways of solving equation 3.9, a frequent one being an iterative method.[11] The vector W is called an eigen or characteristic vector, and the value λ an eigen or characteristic value. The line whose support is given by W is called a factorial axis, and the word "factor" is also taken to be the same as "component." The K elements of W are called "factor-score coefficients."

All subsequent components α have decreasing variances λ_α whose sum is the total variance of the K indicators. This total variance is also named the *total inertia* of the distribution of the K indicators. The stepwise reduction process just described corresponds geometrically to a change in the Cartesian axis system (translation and rotation) of the K-dimension euclidean space R^K. It is neutral regarding the orientation of the factorial axis. The whole process relies on analyzing the structure of the covariance matrix of the K initial variables.

The first component F_1 is an interesting candidate for the composite indicator of poverty C, but it must satisfy obvious consistency conditions relative to the signs of the K elements of W. C has the following expression for the population unit i:

$$C_i = \sum_{k=1}^{K} W^{1,k} I_i^{*k}. \tag{3.10}$$

[10] W is a (K.1) column-vector.

[11] See Anderson (1958).

The I^{*k} are the standardized primary indicators, i.e., the columns of the data matrix X after standardization. The factor score coefficients $W^{1,k}$ must all be positive (negative) to interpret the first component as a decreasing (increasing) poverty indicator, depending on whether the primary indicators increase (decrease) when people become wealthier. At the end of the process, it comes out that the $W^{\alpha,k}$ are in fact the usual multiple regression coefficients between the component F_{α} and the standardized primary indicators. Built this way, the first component can be described *as the best regressed latent variable on the K primary poverty indicators.* No other explained variable is more informative, in the sense of explained variance.

3.2.3 The Fuzzy Subset Approach[12]

The fuzzy subset approach is motivated by the artificial dichotomization between the poor and the non-poor, which is determined by a poverty line whose definition is rarely uncontroversial. Let x be a welfare indicator, e.g., total expenditure per capita, which we want to use as a poverty indicator. The starting idea is then to transform x in x'= 1 − F(x), where F(x) is the distribution function of x. x', taking its values in [0.1), is then interpreted as a degree of poverty and the function 1 − F(x) is called a membership function. Clearly this definition can be applied also to any categorical discrete ordinal indicator, which is then recoded as a numerical indicator. In this categorical case, Betti et al. (2006) use instead the definition

$$x' = \{1 - F(x)/1 - F(x_1)\} \tag{3.11}$$

where x_1 is the smallest value taken by the indicator. The poorest individuals then take the value 1, and the richest, the value 0.

Suppose now there are K indicators with transformed values x'^{k} according to equation 3.11, and the value $x'^{k,I}$ for individual i. The composite indicator C is then defined as the weighted average:

$$C_i = \frac{\sum_{k=1}^{K} Wkx'k, i}{\sum_{k=1}^{K} wk} \tag{3.12}$$

where w_k is an indicator weight defined a priori from the average $\overline{x'k}$ of x'^{k}:

$$wk = \ln \frac{1}{x'k}. \tag{3.13}$$

[12] We follow here essentially Betti et al. (2006), focusing on the basic framework of the fuzzy approach.

It must be observed that, according to equation 3.11, $\overline{x'k}$ tends to be smaller with smaller frequencies of the most deprived ones (lower values of $x_1^{\prime k}$). In fact, for a dichotomous indicator, it gives $\overline{x'k} = \frac{N_1^k}{N}$, where N_1^k is the number of deprived individuals. Then the weight given to indicator k is larger with a smaller number of deprived people. "Thus deprivations which affect only a small proportion of the population, and hence are likely to be considered more critical, get larger weights; while those affecting large proportions, hence likely to be regarded as less critical, get smaller weights".[13]

We retain from the basic fuzzy approach that

a) it is immediately applicable to categorical ordinal indicators;
b) an important preliminary step before aggregation consists in a numerical rescaling of each primary indicator, based on marginal distributions; and
c) indicator weights are defined a priori from the marginal distributions allowing for greater importance given to less frequent deprivations.

At the end of this overview of methodologies for defining a composite indicator of poverty, our second methodological choice is to explore deeper the factorial approach, essentially since it seems a priori more promising, with its internal viewpoint, to articulate our understanding of multidimensionality, while offering at first sight an interesting proposal for a composite indicator. But some variants of the factorial techniques still need to be discussed.

3.3 Factorial Techniques

3.3.1 Factor Analysis (FA)[14] and Principal Component Analysis (PCA)

As seen above (Section 2.2.2), PCA is a factorial technique searching for a small set of independent *linear combinations of the K primary indicators*, called "components," to catch a maximal portion of the total variance of the distribution. When all possible components have been extracted, the whole variance is explained. The first component, accounting for the largest portion of the variance, is an interesting CIP candidate if some consistency conditions are met. This is the approach used by Filmer and Pritchett (1998) for their household asset index.

Factor analysis (FA) is the reverse way of exploring multidimensionality. It tries to identify *K linear combinations of $m < K$ latent (nonobservable) variables*, called factors or communalities, able to predict the K observed indicators with as small an error as possible. More precisely, the predictive model to be estimated is[15]

[13] loc. cit.

[14] A specific technique not to be confounded with Factorial Analysis, which is a generic term.

[15] See Anderson (1958), Section 14.7.

$$I = \Lambda f + U \tag{3.14}$$

where I is the vector of the K primary indicators,[16] Λ is a (K,m)-matrix of *factor loadings*, f is an m-component vector of nonobservable factor scores and U is a K-vector of error. A difficult decision has to be made on the number m of factors to retain in the model. Different estimation techniques can be used, including a principal component approach. Clearly this modeling factorial technique does not respond directly to our research objective to get a CIP. But the m latent factors can in fact be expressed as linear combinations of the K primary indicators through equation[17] 3.15 linking factor-score coefficients and factor loadings:

$$W = \Sigma^{-1}\Lambda \tag{3.15}$$

where W is the (K,m) matrix of the *factor-score coefficients* as defined above with PCA and Σ^{-1} is the inverse covariance matrix[18] of the K primary indicators. Once the matrix W is obtained through equation 3.15, as in PCA, the first factor is an interesting candidate for a CIP, again if consistency conditions hold with the first factor-score coefficients. This is in fact the way Sahn and Stifel (2000) proceed to build a household asset index from data sets provided by the Demographic and Health Surveys (DHS), taking m = 1 in the model, i.e., only one factor.

In comparison with PCA, it should be noted here that "in PCA, multicollinearity is not a problem because there is no need to invert a matrix. For most forms of FA and for estimation of factor scores in any form of FA, singularity or extreme multicollinearity is a problem."[19]

In addition to being theoretically developed for numerical variables as PCA, for the objective of defining a CIP, the FA approach appears to us as an unnecessary detour with possible technical difficulties.

3.3.2 Multiple Correspondence Analysis (MCA) and PCA

Interesting as it is, the PCA technique has some limitations:

a) the whole technique has been developed for a set of quantitative variables, measured in the same units. The optimal sampling properties for parameter estimation depend on the multivariate normal distribution and do not any more exist with categorical variables[20];

---

[16] We suppose here that the K indicators are standardized.

[17] See Tabachnick and Fidell (2001), Chapter 13, *Principal Components and Factor Analysis*.

[18] Here a correlation matrix since the indicators are supposed already standardized.

[19] Tabachnick and Fidell (2001), p. 589.

[20] See, among others, Kolenikov and Angeles (2004) for a similar critique of using PCA with discrete data. The authors ignore MCA as a possible solution. They use a parametric approach based

b) the operationalization of the composite indicator, for population units not involved in the sample used for estimation, is not very appealing since weights are applicable to *standardized* primary indicators. Particularly, standardization adds some ambiguity in a dynamic analysis where the base-year weights are kept constant, as we think they should.

Since concepts of multidimensional poverty are frequently measured with categorical ordinal indicators, for which PCA is not a priori an optimal approach, looking for a similar but more appropriate factorial technique is justified. Here comes naturally into the picture multiple correspondence analysis (MCA), designed in the 1960s and 1970s[21] to improve the PCA approach when the latter loses its parametric estimation optimal properties and to provide more powerful description tools of the hidden structure in a set of categorical variables.

The most important technical difference between PCA and MCA is the use of the χ^2 metric (chi-square), instead of the usual Euclidean metric used in PCA, to measure distances between two lines or two columns of the data matrix being analyzed. This χ^2 metric has been introduced into the area of factorial analysis in the years 1960–1970 by the French school of statistics led by J.-P. Benzécri, and then appeared as factorial techniques specifically designed for categorical variables "Correspondence Analysis" (CA) and its extension "Multiple Correspondence Analysis" (MCA).

From now on, we will assume that the K primary indicators are categorical ordinals, the indicator I_k having J_k categories. This is a very general setting, applicable to any mix of quantitative and categorical poverty indicators, since a quantitative variable can always be redefined in terms of a finite number of categories. Let us associate with each primary indicator I_k the set of J_k binary variable 0/1, each corresponding to a category of the indicator. We introduce the following notation:

1. X(N,J): the matrix of N observations on the K indicators decomposed into J_k binary variables, where $J = \sum_{k=1}^{K} J_k$ is the total number of categories. X is named the *indicatrix matrix.*
2. N_j: the absolute frequency of category j, i.e., the sum of column j of X;
3. N': the sum of the elements of matrix X, i.e., N $\times$ K;
4. $f_j = \frac{N_j}{N'}$: the relative frequency of category j $f_j^i = \frac{X(i,j)}{X(i)}$, where X($i$) is the sum of line i of the matrix X. The set $f_j^i = \left\{ f_j^i, j = 1, J \right\}$ is named the profile of observation i.

on the multivariate normal distribution and the estimation of a polychoric correlation coefficient matrix.

[21] The French school of data analysis led by Benzécri has been particularly creative and influential in the development of correspondence analysis.

MCA is a PCA process applied to the indicatrix matrix X, i.e., to the set of the J binary variables in the R^N space, transformed into profiles, but with the χ^2 metric on row/column profiles, instead of the usual Euclidean metric.

The χ^2 metric is in fact a special case of the Mahalanobis metric developed in the 1930s and used in Generalized Canonical Analysis. It takes here the following form, for the distance between two observed profiles i and i' in the R^{J} space:

$$d^2\left(f_J^i, f_J^{i'}\right) = \sum_{j=1}^{J}\left(\frac{1}{f_j}\right)\left(f_j^i - f_j^{i'}\right)^2. \tag{3.16}$$

The only difference with the Euclidean metric lies in the term $\left(\frac{1}{f_j}\right)$, by which low-frequency categories receive a higher weight in the computation of distance.

The χ^2 metric has two important properties not possessed by the Euclidean metric[22]: the distributional equivalence property and the duality property. The χ^2 metric is directly linked to statistics used in very old statistical tests like the Pearson χ^2-test of the theoretical distribution of a given empirical distribution and the Pearson χ^2-test of the independence of two categorical variables presented in a two-way frequency table.

The distributional equivalence property means that the distance between two lines (individuals, households, etc.) of the profile matrix remains invariant if two identical columns (poverty variable) are merged, or if we add to the data matrix a column identical to another one. And symmetrically, for modifying lines and keeping invariant the distance between columns. Concretely, it means that the factorial analysis run with the χ^2 metric, as with MCA, is quite robust (stable, invariant), to the way a set of categorical variables, as poverty indicators, is built: extending a set of indicators with closely correlated additional indicators, defining categories within a same indicator, etc. PCA, with the Euclidean metric, is sensitive to such transformations. This theoretical property is empirically observed and illustrated in references given in the preceding footnote.

The duality property is explicitly presented with the duality equations 3.18a and 3.18b. These equations are also referred to in the literature as "transition" or "barycentric" equations.[23] This duality property is the theoretical basis (see the literature just referred to) allowing the simultaneous representation, in the same factorial plane, of the lines (individuals, households), often aggregated in socioeconomic groups, and of the columns (poverty attributes). This simultaneous representation, unique to MCA, is a very powerful exploration tool for the identifi-

[22] That these properties are specific to the χ^2 metric can be found in Benzécri J.P. and F. Benzécri (1980), pp. 37–40, Greenacre M. and J. Blasius (1994), p. 35, and Lebart L., A. Morineau and M. Piron (1990), p. 74.

[23] The duality equations can be found in Benzécri J.P. and F. Benzécri (1980), pp. 80–90, Lebart L., A. Morineau and M. Piron (1990), pp. 75–79, and Greenacre M. and J. Blasius (1994), p. 14.

cation of poverty determinants, associated with poverty types. In fact, this property, much more than the distributional equivalence one, is the main advantage of MCA for applying factorial concepts and methods to multidimensional poverty analysis.

To sum up, due to using the χ^2 metric, the difference between MCA and PCA shows up particularly in two properties which seem highly relevant for the poverty meaning of the numerical results.

Property #1 (marginalization preference)

Factorial scores produced by MCA overweight the smaller categories within each primary indicator. In fact, we have

$$W_{j_k}^{\alpha,k} = \frac{N}{N_{j_k}^k} \text{ Covariance } \left(F_\alpha^*, I_{j_k}^k\right) \tag{3.17}$$

where

$W_{\mathrm{j_k}}^{\alpha,\mathrm{k}}$ = the score of category $\mathrm{j_k}$ on the factorial axis α(non – normalised)

$I_{j_k}^k$ = the binary variable 0/1 taking the value 1 when the population unit has the category $\mathrm{j_k}$.

F_α^* = the normalized score on the factorial axis α

$N_{j_k}^k$ = the frequency of the category $\mathrm{j_k}$ of indicator k

Thus, in the case of a binomial indicator, the marginal category will receive a higher weight, since the covariance is the same for both categories.

In terms of poverty, if we think of (extreme) poverty in a given society as being more relative than absolute and characterized by social marginalization, i.e., by the belonging to a minority group within the population, the group of people characterized by a poverty category j_k, then this category will receive more weight in the computation of a composite indicator of poverty. If we interpret the factorial weights (regression weights) as expressing the social choice in poverty reduction, then these highly weighted poverty attributes represent those which this society tries to eliminate in priority. As noticed above (Section 3.2.3), this higher weight given to a smaller number of deprived people is looked for by the indicator weighting system defined a priori with the fuzzy approach.

Property #2 (reciprocal bi-additivity or duality)

The way it is defined, MCA can be applied on the indicatrix-matrix either to the row-profiles (observations) or to the column-profiles (categories), so that it has the following remarkable and unique duality property:

$$F_\alpha^i = \frac{\sum_{k=1}^{K} \sum_{jk=1}^{Jk} \frac{W_{jk}^{\alpha,k}}{\sqrt{\lambda_\alpha}} I_{i,jk}^k}{K} \tag{3.18a}$$

where

K = number of categorical indicators
J_k = number of categories for indicator k
$W_{jk}^{\alpha,k}$ = the score of category j_k on the factorial axis α(non – normalised)
I_{i,j_k}^{k} = the binary variable 0/1 taking the value 1 when the unit i has the category j_k.
F_{α}^{i} = the score (non - normalized) of observation i on the factorial axis α

and reciprocally

$$W_{jk}^{\alpha,k} = \frac{\sum_{i=1}^{N_{jk}} \frac{F_{\alpha}^{i}}{\sqrt{\lambda_{\alpha}}}}{N_{jk}^{k}}. \tag{3.18b}$$

Let us assume, for example, that the first factorial axis meets the consistency conditions to be considered as a poverty axis[24] and that we can take as the composite indicator of poverty $C_i = F_1^i$. Then the duality relationships stipulate

Equation 3.18a: the composite poverty score of a population unit is the simple average of the factorial weights (standardized) of the K poverty categories to which it belongs.
Equation 3.18b: the weight of a given poverty category is the simple average of the composite poverty scores (standardized) of the population units belonging to the corresponding poverty group.

We feel that these two properties, especially equation 3.18b for the reciprocal bi-addivity, are quite relevant for the poverty meaning of the numerical results coming out of this specific factorial analysis, MCA.[25] With the simultaneous graphical representation of population units and poverty attributes, MCA appears as an analytic tool particularly efficient for the study of multidimensional poverty represented in a set of categorical ordinal indicators.

It must also be observed that by breaking down each indicator I_k in as many variables, J_k, as there are categories, MCA allows for *non-linearity* in the categorical weights, contrary to a PCA which would be run on a numerical coding 1 to J_k of the indicator I_k , as some researchers could be tempted to do.

Having looked at some variants of factorial analysis, FA, PCA, and MCA, our third methodological choice is to go on with MCA, due essentially to its particular convenience for categorical variables, its remarkable duality properties and its operationality. This is why we explore more attentively in the following section a research strategy that is relevant in applying MCA to the problem of measuring multidimensional poverty.

---

[24] We come back to these consistency conditions in Section 3.4 below.

[25] A complete description of MCA can be found in Lebart et al. (2000) or Greenacre and Blasius (1994).

3.4 MCA Technique Applied to Multidimensional Poverty Measurement

Since MCA consists basically in exploring the internal structure of a covariance matrix while producing at the same time an additive decreasing disaggregation of the total variance (inertia) of the matrix, the rationale for using such a technique in the context of multidimensional poverty consists in searching the real multidimensionality of poverty reflected in a set of poverty indicators more or less correlated. And the specific by-product of such a search is a significant composite indicator of multidimensional poverty, as we will now see.

3.4.1 A Fundamental Consistency Requirement

We now consider more closely the conditions under which the factorial approach, and especially the MCA variant, can generate a truly relevant composite indicator of multidimensional poverty. We could have here a full axiomatic formulation so that the objective of poverty comparison is satisfactorily met. But the axiomatic requirements can be largely simplified with a two-step approach. If the first step has provided a relevant composite indicator of poverty, the axiomatic requirements for the second step, regarding the computation of aggregated poverty indices, can rely on standard requirements now generally accepted in the case of unidimensional poverty measurement, especially for the well-known case of money-metric poverty. For the first step of constructing a composite indicator C from K ordinal categorical indicators I_{k}, there is at least the following requirement:

Monotonicity axiom (M)[26]

The composite indicator of poverty must be monotonically increasing in each of the primary indicators I_{k}.

The axiom just means that if a population unit i improves its situation for a given primary indicator I_{k}, then its composite poverty value C_i increases: its poverty level decreases.

Let us see what it means to take the first factorial component F_1 as the composite indicator of poverty C. From equation 3.18a above, its expression would be

$$C_i = F_1^i = \frac{\sum_{k=1}^{K} \sum_{jk=1}^{Jk} \frac{W_{jk}^{1,k}}{\sqrt{\lambda_1}} I_{i,jk}^{k}}{K}. \tag{3.19a}$$

[26] We assume that the sign of the composite indicator is selected in such a way that a larger value means less poverty or, equivalently, a welfare improvement, and that the ordering relation A < B between two categories A and B of the same indicator means that B is preferable to A.

To simplify, let us write $W^{*\alpha,k} = \frac{W^{\alpha,k}}{\sqrt{\lambda_\alpha}}$ for the normalized category-score on the factorial axis α. Then we have

$$C_i = \frac{\sum_{k=1}^{K} \sum_{j_k=1}^{J_k} W_{jk}^{*1,k} I_{i,j_k}^{k}}{K}. \tag{3.19b}$$

The monotonicity axiom translates into two requirements:

M1: First Axis Ordering Consistency (FAOC-I) for an indicator I_k

For an indicator I_k for which the ordering relation between categories is noted $<_k$, the ordering relation $<_w$ of the weights $W_{j_k}^{*1,k}$ must be equivalent to either $<_k$ or to $>_k$.

M2: *Global First Axis Ordering Consistency (FAOC-G)*

For all indicators I_k, the FAOC-I condition is fulfilled with the same orientation: the ordering relation $<_w$ is equivalent to either $<_k$ for all indicators or to $>_k$ for all.

If and only if the monotonicity axiom is satisfied can $C = F_1$ be taken as a composite indicator of poverty, after eventually changing the sign of F_1 when $<_w$ is equivalent to $>_k$ for all indicators. But then the reciprocal bi-addivity property of MCA gives a very interesting consistency result for C_i. Due to equation 3.18b which says that the weight of an indicator category, $W_{j_k}^{1,k}$, is given by the average composite poverty score of the population group of size N_{j_k} having the category (attribute) j_k, we can state the following property of C:

Composite Poverty Ordering Consistency (CPOC)

With $C = F_1$ satisfying the monotonicity axiom (M), for a given indicator I_k, let the population group P_{j_1} have a category j_1 of I_k inferior to the category j_2 possessed by the group P_{j2}. Then the group P_{j_1} is also poorer than P_{j2} relative to the composite poverty.

In other words, the population ordering for a primary indicator I_k is preserved with the composite indicator. This is a remarkable consistency property specific to MCA, due to the dual structure of the analysis.

Clearly, there is no guarantee that MCA run on the K primary indicators will come out with the FAOC property, and then using the first factorial component as the composite indicator of poverty would be inconsistent and not acceptable. In fact, everything depends on the structure of the covariance matrix X'X.[27]

There are two ways of overcoming this unpredictable difficulty: minor adjustments to the set of the K primary indicators, or exploiting more than one factorial axis.

[27] We use X for the matrix of centered variables.

3.4.2 Positive Rescaling of the K Primary Indicators[28]

As seen above, due to the duality relationship equation 3.18b, the categorical weight $W_{jk}^{*1,k} = \frac{W_{jk}^{1,k}}{\sqrt{\lambda_1}}$ appearing in the CIP equation 3.19b has a strong meaning in terms of multidimensional poverty: it is the average multidimensional poverty level of the group of individuals having the category j_k of the primary indicator I_k. But the numerical value of $W_{jk}^{*1,k}$, either negative or positive since by construction the average is zero, is irrelevant inasmuch as the numerical scaling of I_k remains unchanged relative to the distances between categories. Developing this idea, it is possible to improve the meaning of the categorical weights by rescaling I_k with the gap between the worst-off individuals, $j_k = 1$, and any better-off group, $j_k = 1$. We are thus led to rescale the indicator I_k, on the factorial axis α, here supposed to satisfy the consistency requirements, with the following categorical weights:

$$W_{jk}^{+\alpha,k} = \frac{W_{jk}^{\alpha,k} - {W_1}^{\alpha,k}}{\sqrt{\lambda_\alpha}}. \tag{3.20}$$

Thus, the most deprived category for I_k always has a weight equal to zero, and the weight given to any superior category j_k, strictly positive, represents the gain in *total* poverty reduction, as measured on axis α, when an individual can get out of the most deprived status in the primary indicator I_k by accessing the status $j_k, k > 1$. Under the hypothesis that the first factorial axis satisfies the FAOC condition, the definition of equation 3.19b of the CIP is now transformed as

$$C_i = \frac{\sum_{k=1}^{K} \sum_{jk=1}^{J_k} W_{jk}^{+1,k} I_{i,jk}^{k}}{K}, \quad C_i \geq 0. \tag{3.21}$$

From this point of view, MCA appears as a technique of rescaling numerically, in a meaningful way, a set of categorical ordinal indicators and of providing at the same time the rationale for a consistent aggregation of the rescaled indicators.

3.4.3 Adjustments to the Set of the K Primary Indicators

It should be noted that a binomial indicator always meets the FAOC-I requirement. In the case in which a multinomial indicator does not satisfy this requirement, regrouping some categories can sometimes achieve the FAOC-I. If this operation does not succeed, a more radical solution is to eliminate the indicator. Obviously,

[28] This section assumes that the factorial axis referred to, usually the first axis, meets the FAOC-G condition, with the orientation chosen such that welfare increases (poverty decreases) from left (negative side) to right. A simple adaptation of this will be made below in a more general case.

if the primary indicators have been carefully selected, defined and tested, this is a high price to pay for satisfying a technical condition. Although we do not in general favor the elimination of indicators, the option does become more acceptable when the number of indicators K is large and there appears to be some duplication in a specific domain (or dimension) of poverty.

If all indicators satisfy FAOC-I but FAOC-G is not met, it means that relative to the first factorial axis there are two subsets of indicators with opposite ordering on this axis, thus negatively correlated. Two such disjoint subsets of indicators will always appear with K binomial indicators, this being in particular the case when applying MCA to asset poverty, where the indicator for each asset is usually binomial: ownership or not. In this last case, there is no consistency problem if one of the two subsets is the empty subset $\varnothing$, which is not unusual. Let us assume that both subsets are not empty. It means that the multivariate measurement of poverty cannot be shrunk into a unidimensional poverty measurement restricted to the first factorial axis, and that in spite of existing correlations, the poverty concept reflected in the K chosen indicators is really deeply multidimensional. If we stick to the first factorial axis, the only way to get out of this inconsistency would be to eliminate one of the two subsets of indicators, which does not seem a priori acceptable: the information loss would then be too important. We need a more appropriate research strategy going beyond the first factorial axis.

3.4.4 A Research Strategy Using More than the First Factorial Axis

We need some additional tools to design a research strategy that will not consider only the first factorial axis. Let L be the number of factorial axes, determined by the rank of the matrix X. We have $L \leq J - K$, where J is the total number of categories for the K indicators.

$$\text{Let } \Delta_l^k = \frac{\sum_{j_k=1}^{J_k} N_{j_k}^k W_{k,j_k,l}^2}{N} \tag{3.22}$$

be the *discrimination measure* of indicator I_{k} on the factorial axis I. It is in fact the variance of the distribution of the categorical weights on axis I, since the average weight is always 0.

We know from the theory of MCA that

$$\lambda_l = \frac{\sum_{k=1}^{K} \Delta_l^k}{K}, \tag{3.23}$$

i.e., the eigenvalue of axis I, is the average of the discrimination measures of the K indicators.

It follows from the basic factorial equation

$$\text{Total Inertia} = I_{tot} = \sum_{l=1}^{L} \lambda_l \tag{3.24}$$

that we have the equation 3.25 below:

3.4.4.1 Total Inertia Decomposition

$$I_{tot} = \frac{\sum_{l=1}^{L} \sum_{k=1}^{K} \sum_{j_k=1}^{J_k} N_{j_k}^{k} W_{k,j_k,l}^{2}}{K \times N} = \frac{\sum_{l=1}^{L} \sum_{k=1}^{K} \Delta_l^k}{K}. \tag{3.25}$$

In the case of MCA, $I_{tot} = \frac{J}{K} - 1$, i.e., it is the average number of categories per indicator minus 1.[29] If all indicators are binomial, total inertia is precisely 1. It is also shown that the contribution of indicator I_{k} to total inertia is $J_k - 1$.

Let us denote $\kappa = \{1, 2, \ldots, K\}$ as the set of integers from 1 to K.

We will now generalize the previous approach to the composite indicator of poverty.

First observe that there is an obvious one–one correspondence between the categorical coefficients appearing in the linear expression of the L possible factorial components and the categorical contributions to the disaggregated total inertia, as shown in matrices A and B. The general term of matrix B is the square of the matrix A general term. The usual approach restricted to consider only the first factorial axis to pick up from A the J coefficients of the composite indicator, conditional on the FAOC-G requirement, is based on the fact that in matrix B the first line has a maximal sum in the inertia decomposition. Equation 18a is then the functional frame of the composite indicator, as proposed by the factorial theory. But there is no specific reason why a maximal inertia criterion, conditional to ordering (poverty) consistency as revealed in matrix A, should be restricted to the first line of matrix B. A more efficient approach can be looked for, and this is the idea explored to generalize the usual factorial approach to the CIP (Figs. 3.9 and 3.10).

1	…	j_1	…	j_1	1	…	j_k	…	J_k	1	…	j_K	…	J_K
1														
:														
l							$W_{jk}^{l,k}$							
:														
L														

Fig. 3.9 Matrix A categorical coefficients in factorial components: Equation (3.18a)

For each factorial axis I, we can look for two *subsets* of indicators, each subset satisfying the axis ordering consistency condition (AOC) in one of the two axis

[29] See Lebart et al. (2000) p.120.

	1	…	j_1	…	j_1	1	…	j_k	…	J_k	1	…	j_K	…	J_K
1															
:															
l								$W^2_{k,\,jk,\,l}$							
:															
L															

Fig. 3.10 Matrix B categorical contributions in total inertia disagregation: Equation (3.25)

orientations, i.e., both requirements AOC-I and AOC-G, which now no more refer only to the first axis. The worst situation occurs when, for a given axis I, no indicator meets AOC-I; both subsets are then the empty subset Ø. Among these AOC subsets, we retain the one whose sum of discrimination measures is maximal. We will then consider that there is a *poverty type* specific to axis I if and only if the *sum of discrimination measures* of this AOC subset represents the larger part of the total discriminating power of axis I, i.e., is larger than 50% of $K \times \lambda_l$. Axis I will then be named a *poverty axis* and the sum of discrimination measures of this AOC subset is identified as the *poverty-relevant inertia* of axis I. To each factorial axis I, we can thus associate a unique subset of the K indicators, whose indices are a *subset* κ_l of κ.

Poverty Type Set of Axis l

The Poverty Type Set of the factorial axis I, $\{I_k\}_{k\in\kappa_l}$, is the most discriminating subset of AOC indicators satisfying $2 \times \sum\limits_{k\in\kappa l} \Delta_l^k > K\lambda_l$.

It should be clear that the set $\{I_k\}_{k\in\kappa_1}$ can be empty, which means that the factorial axis I does not represent any poverty type set.

It should also be clear that the poverty type sets from different axes are not necessarily disjoint: the same indicator can belong to many of them. The potential intersection between these sets can be eliminated by a sequential process starting with the first axis and continuing with the others as ordered by MCA, since the discriminating power of each axis is decreasing. The way to eliminate these intersections, while trying to retain at each step the maximal inertia, is naturally coming out of the total inertia decomposition (3.25): at each step, we keep a given indicator k into the poverty type set where its discrimination measure is larger. We refer to this sequential process as to *the algorithmic identification of independent poverty types*, more simply the *poverty types algorithm*. Let then $\kappa_l^* \subseteq \kappa_l$ be the subset of indicator indices at step $L^* \geq 1$ in the sequential process.

Normally, to ensure that the process retains a maximal proportion of I_{tot} in the disjoint poverty sets, the algorithm must be pursued until $L^* = L$. We then have built a complete sequence of poverty type sets.

3.4.4.2 Complete Sequence of Poverty Type Sets

The sequence of disjoint subsets of indicators $\{I_k\}_{k\in\kappa_l^*}$ resulting from the application of the poverty types algorithm until $L^* = L$, is called a complete sequence of poverty type sets. The number d of non-empty subsets is the number of independent poverty types provided by the set of the K primary indicators.

Two cases are then possible: all K indicators belong to the sequence, i.e., $\bigcup_{l=1}^{L} \kappa_l^* = \kappa$, or some indicators are not retained in the process. In this last case, they could simply be eliminated from the search of a composite indicator: in a simultaneous factorial analysis of all K indicators, they do not meet the minimal consistency requirement on any factorial axis. But again, we cannot necessarily assume that these rejected indicators are not good. A less radical approach would be to process them separately as a second set of indicators and to build with them a second composite indicator.[30] With two numerical CIP, any of the reviewed aggregation techniques well fitted to quantitative indicators could be used, including PCA.

The poverty types algorithm can rapidly become quite demanding with a large number K of primary indicators, let us say $K \geq 10$, which is not unusual in applied multidimensional poverty analysis. As an example, with 10 indicators having on average 3 categories, the process could involve the analysis of $L = 20$ factorial axis. Even if all well-known software allows such an analysis with some tedious work for the analyst, to facilitate the operationalization, it seems admissible, even if not optimal, to introduce the possibility of interrupting the algorithm when some kind of ideal situation is met, that is, as soon as all K indicators appear in a sequence of disjoint poverty type sets. This leads us to the following definition:

3.4.4.3 Minimal and Admissible Sequences of Poverty Type Sets

A *minimal sequence of poverty type sets* is obtained when the poverty types algorithm is interrupted at the smallest value $L^* \leq L$ for which either $\bigcup_{l=1}^{L^*} \kappa_l^* = \kappa$, i.e., all indicators are included in the sequence of disjoint poverty sets, or $L^* = L$.

Here also, the number d of non-empty subsets is the number of independent poverty types provided by the set of the K primary indicators.

It should be stressed that this definition allows, in particular, for stopping the process to the first factorial axis if the FAOC condition is achieved. To our knowledge, this has been the usual practice until now, unfortunately at the expense of frequently giving up a subset of the primary indicators or of merging relevant categories, which means an information loss.

If a minimal sequence of poverty type sets is reached for a small $L^* < L$, e.g., for $L^* = 1$ (first axis), there can still be an important loss of information with some indicators having a very low discriminating power. In that case, important improvements can be obtained by considering additional axes beyond L*, without necessarily going until $L^* = L$. It is clear that beyond L*, all K indicators remain in the disjoint sequence of poverty sets, but some indicators could be associated to a poverty set and axis I in which their discrimination measure is higher. It then seems better to extend the algorithm until some criterion is met. One possible criterion is to stop the process when the sum of the L* eigenvalues represent at least 50% of

[30] This means a rerun of the first factorial analysis without these indicators.

the total inertia, I_{tot}, given by $I_{tot} = \frac{J}{K} - 1$. That type of minimal sequence can then be called an *admissible sequence of poverty type sets*. Each axis that appears in the sequence then has an inertia (eigenvalue) larger than the average inertia per factorial axis, $1/K$. This application of the algorithm obviously requires analyzing less than half of the total number of factorial axes, possibly much less depending on the inertia captured by the first axes. When a minimal sequence exists, especially when it occurs immediately at $L^* = 1$, our proposal would thus be to pursue the algorithm until an admissible sequence has been reached.

We can now derive, from equation 3.21, a generalized definition of the composite indicator of poverty, which can be applied when the first factorial axis does not meet the FAOC-G requirement.

3.4.4.4 Generalized Definition of the Composite Indicator of Poverty

Let a complete or admissible sequence of complete poverty type sets be obtained, which is always possible with the poverty dimensions algorithm. Then the value C_i of the composite indicator of poverty for the population unit i is given by

$$C_i = \frac{\sum_{l=1}^{L^*} \sum_{k \in \kappa_l^*} \sum_{jk=1}^{J_k} W_{jk}^{+l,k} I_{i,jk}^{k}}{K}. \tag{3.26}$$

Definition of equation 3.21 is the special case where $L^* = 1$: all K indicators belong to the poverty type subset of the first factorial axis. This is the case where the multivariate measurement of poverty can be logically reduced to one aggregate poverty type, due to the structure of the correlation matrix: all K indicators are positively correlated. In the general case, there is more than one poverty type, in fact one for each poverty type set; the way to aggregate them is suggested by the structure of equation 3.19a and the fundamental equation of decomposition of the total inertia 3.23: instead of picking up the J_k weights attributed to the indicator I_k only from the set of weights provided by the first factorial axis, it takes them from the axis which define the poverty type subset to which it belongs with a maximal variance. The positive rescaling of the indicators (Section 3.4.2) is done only for the poverty type set, with the orientation of the axis chosen consequently from left (poorer) to right (less poor).

Coming back to matrices A and B above, this algorithmic approach to the CIP means that we move simultaneously in the whole matrices A and B, A to identify any existing poverty ordering consistency, B to keep the most relevant ones according to the discrimination measure, and, avoiding any overlapping, this optimization process is translated into a CIP according to the duality frame 3.18a.

Deliberately we did not use the term poverty *dimension* set in place of poverty *type* set. A poverty dimension is identified a priori as a subset of indicators relative to the same domain of basic needs or basic welfare. It is an a priori concept. A poverty type is a statistical concept defined from the multivariate distribution of the whole set of indicators in a given population. A poverty type can, and will usually be, poverty multidimensional. It is a concept that helps exploring, reducing, and clarifying the meaning of multidimensional poverty in a given population, according to a behavioral specificity of that population and/or to specific poverty reduction policies. Numerous poverty dimensions can thus shrink into just one poverty type, or some types, which obviously should simplify the analysis. This is what we try to achieve by the proposed generalized construction of the composite indicator of poverty.

It should be noted that the two very relevant properties of MCA, the marginalization preference equation 3.17 and the reciprocal bi-additivity, especially equation 3.18b, are valid in each of the L^* axes involved in the generalized definition and thus keep their meaning, in the relevant poverty type I, for the interpretation of the categorical weights of the κ_l^* indicators defining this type. Moreover, the composite poverty ordering consistency remains valid for each identified poverty axis, with obvious adaptation.

The whole generalization approach must be viewed as an effort to highlight the multidimensional poverty structure hidden in the K-variate measurement of poverty, and at the same time to integrate into the composite indicator of poverty the maximum amount of information from the full information contained in the K primary indicators, as measured by the total inertia.

3.5 MCA: A Numerical Illustration

To illustrate the MCA technique described in Section 3.4, we use a household data set provided by the poverty observatory experimented in Vietnam in 2002 for monitoring the National Programme for Hunger Eradication, Poverty Reduction and Job Creation. The household survey was run in 4,000 households drawn from 20 communes, following the CBMS methodology developed in Vietnam by the Vietnam Socio-Economic Development Centre in partnership with MOLISA.[31]

Thirteen nonmonetary poverty indicators have been aggregated into a composite indicator. These indicators come from the areas of education (schooling, literacy), health (sickness events, sick days), housing conditions (type of house, toilet, and electricity), and household equipment (bicycle/motorcycle, radio/TV). Education and health indicators are broken down according to gender. These indicators are presented in Table 3.1, which provides the main numerical results coming out of the MCA computation.

The two columns "Factorial scores" give the values $W_{jk}^{\alpha,k}$ and the columns show the values Δ_l^k defined in equation 3.22.

[31] MOLISA: Ministry of Labour, Invalides and Social Affairs.

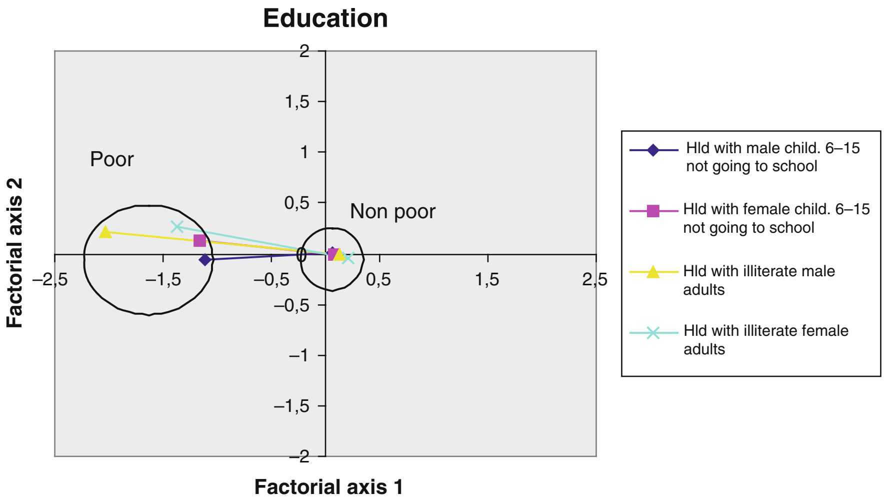
Table 3.1 Vietnam MOLISA poverty observatory; MCA results

		Factorial score		Discrimination measures	
INDICATOR	Category	axis # 1	axis # 2	axis # 1	axis # 2
Male child. 6–15 not going to school	Yes	−1,109	−0,056	0,07	0,00
	No	0,059	0,003		
Female child. 6–15 not going to school	Yes	−1,166	0,132	0,09	0,00
	No	0,074	−0,008		
Hld with illiterate male adults	Yes	−2,04	0,222	0,27	0,00
	No	0,134	−0,014		
Hld with illiterate female adults	Yes	−1,371	0,267	0,29	0,01
	No	0,211	−0,041		
Sickness events male per cap.	no sickness	0,03	0,518	0,00	0,61
	0–1/2	−0,071	−0,734		
	1/2 to 1	−0,065	−1,049		
	103>=1	−0,051	−1,596		
Sickness events female per cap.	no sickness	0,025	0,628	0,00	0,60
	0–1/2	−0,146	−0,54		
	1/2 to 1	0,021	−0,832		
	103>=1	0,019	−1,335		
Hld with radio or tv	With radio or tv	0,37	−0,021	0,43	0,00
	Without radio and tv	−1,153	0,067		
Hld with bicycle or motocycle	With bicycle/moto	0,265	−0,031	0,44	0,01
	Without bicycle/moto	−1,671	0,198		
Type of housing	Multi-storey permanent	1,063	0,38	0,40	0,01
	One-storey permanent	0,577	−0,037		
	Semi-permanent	0,08	0,013		
	Temporary	−1,125	−0,084		
	Not having owned house	−1,948	−0,035		
Type of toilet	Flush toilet septic tank	0,841	0,088	0,40	0,01
	Double vault compost latrine	0,584	−0,045		
	On fish ponds	0,236	0,186		
	Simple toilet/pit latrine	−0,482	0,016		
	On river, canal	−0,505	−0,122		
	Not having owned toilet	−0,905	−0,187		
Hld with electricity	Hld with electricity	0,284	−0,03	0,49	0,01
	Hld without electricity	−1,761	0,185		

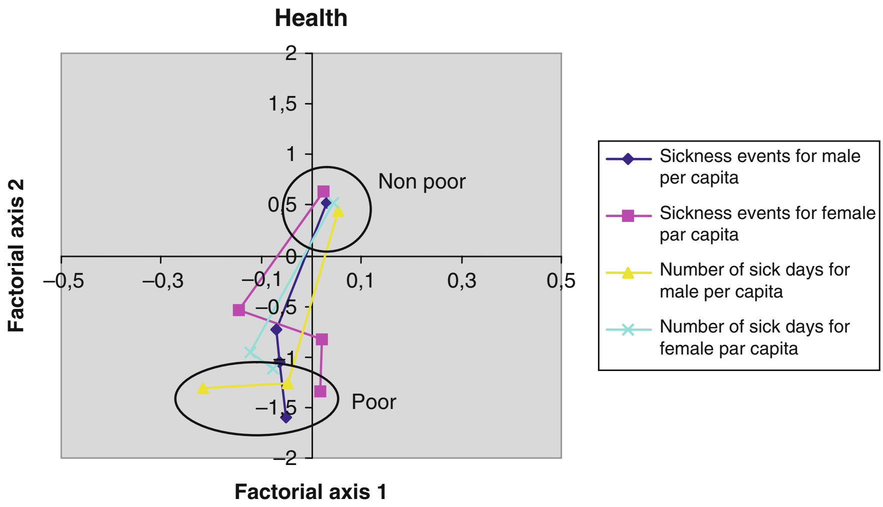
Table 3.1 (continued)

		Factorial score		Discrimination measures	
INDICATOR	Category	axis # 1	axis # 2	axis # 1	axis # 2
Number of sick days male per cap.	No sick days	0,053	0,445	0,01	0,57
	Less than 5 days	−0,046	−1,255		
	More than 5 days	−0,216	−1,305		
Number of sick days female per cap.	No sick days	0,045	0,528	0,00	0,56
	Less than 5 days	−0,121	−0,956		
	More than 5 days	−0,076	−1,122		
	average = eigenvalue			0,22	0,18

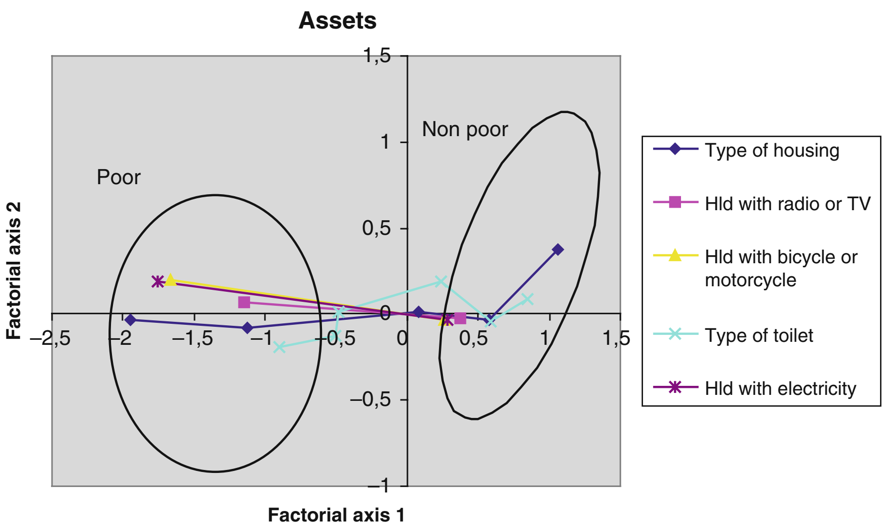
These results are summarized in Graphs 3.1, 3.2, 3.3, and 3.4.



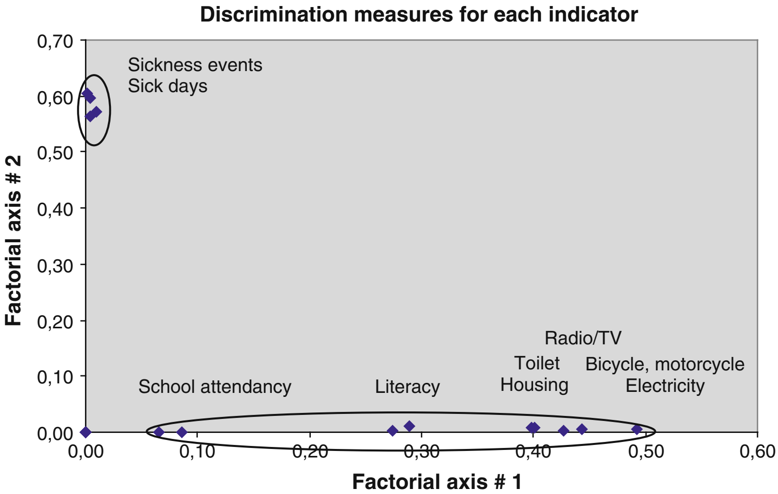
Graph 3.1 Factorial scores (categories), first two axes, education



Graph 3.2 Factorial scores (categories), first two axes, health



Graph 3.3 Factorial scores (categories), first two axes, household assets



Graph 3.4 Discrimination measures, first two axes

From Graphs 3.1 and 3.3 we see that the subsets of the four education indicators and the five household assets indicator meet the FAOC-I consistency requirement, and since the orientation is the same, altogether these nine indicators are globally consistent. But first-axis consistency problems appear with the health indicators. In fact, three of these four indicators are not consistent with the first axis, as can be seen from Graph 3.2. Only "number of sick days for male per capita" is FAOC-I and could be kept in a composite indicator with the nine others since it is globally consistent with them. Thus, considering only the first factorial axis, three health indicators should be eliminated from the composite indicator.

But we observe immediately from Graph 3.2 that these three health indicators are consistent relative to the second axis (as well as the fourth one). In fact, the four health indicators have a high discrimination power on the second axis, and almost no power relative to the first axis. Graph 3.2 illustrates very clearly this observation: education and asset indicator discriminate highly on axis #1 and very little on axis #2, with the reverse situation for health indicators. This is precisely the case where, to avoid an important loss of information, the poverty types algorithm of Section 3.4.4 can be expected to be more efficient in the construction of a composite indicator by translating numerically what is immediately revealed graphically. Computations are presented in Table 3.2.

The 13 indicators are listed in lines, with their discrimination measures on all factorial axes as columns. Here we limit the table to the first three axes since the algorithm, version "admissible sequence," stops at axis #2. The algorithm has been pursued until factorial axis #8, where 50% of the total inertia 2 (39/13 −1) is reached, without providing any additional poverty set beyond axis #2.

The first step identifies, on axis # 1, only one subset of indicators satisfying the AOC condition, the 10 indicators colored[32] in dark gray in the relevant column. This identification requires going back to Graphs 3.1, 3.2, and 3.3 or to Table 3.1. The other three indicators (health) are not consistent. Then we check if the inertia explained by these 10 indicators, 2.886, is at least 50% of the first axis total inertia, which is 1.448. We thus accept axis #1 as describing a poverty type, the poverty type set being the 10 consistent indicators.

The second step, relative to axis #2, identifies a first subset of five consistent indicators, dark gray colored: the four health indicators plus the male school attendance indicator.[33] Their discriminating power is 2.339. A second subset, light gray colored, consistent in the direction opposite to the first subset, includes five indicators with a discriminating power of 0.028. The dark gray subset has more inertia than the required 50% of 1.192 and can thus be accepted as a poverty type subset of indicators; it is completely dominated by the four health indicators. Then arises the elimination of overlapping between the poverty type subsets of axes #1 and #2. "Hld with male child. 6–15 not going to school" is kept in axis #1 where its discriminating power 0.066 is much higher than the 0.000 on axis #2. "Number of sick days for male per capita" is kept in the poverty type subset of axis #2 with a discriminating power of 0.572 instead of 0.010 on axis # 1. The two overlapping indicators are now labeled in white color on the axis where they are eliminated.

We then have two nonoverlapping poverty type subsets, one with nine indicators, another with four, covering the whole set of the 13 indicators. According to the minimal sequence algorithm, we can stop here. Just for curiosity, in Table 3.2 we present the third axis. As can be seen, there are two subsets of AOC indicators, but none of them can reach the requested inertia value 0.749. A similar situation is met for all subsequent axes until axis #8, where at least 50% of the total inertia is explained, and thus an admissible sequence is achieved. Only the first two axes provide poverty sets.

There is a large information gain by using the poverty types algorithm:

- with 13 indicators and 39 categories, the total inertia is $(39/13) - 1 = 2$;
- with reference to equation 3.25, "Total Inertia Decomposition," and Table 3.2, the inertia relative to the 10 indicators consistent on axis #1 is $2.886/13 = 0.222$, 11.1% of total inertia;
- the inertia collected by the 13 indicators coming out of the first two axes is $(2.876 + 2.339)/13 = 0.401$, 20.0% of total inertia, i.e., 81% more than with only the first axis. But the most important fact is that all 13 primary indicators appear in the composite indicator.

[32] As a convention, the dark gray color identifies the left to right (bottom-up) axis orientation, and the light gray color shows the reverse orientation.

[33] Its very low discrimination measure is 0.000169.

Table 3.2 Poverty types algorithm, minimal sequence

Indicators	Discrimination measures Dimension (factorial axis)		
	1	2	3
Hld with male child. 6–15 not going to school	0.066	0.000	0.047
Hld with female child. 6–15 not going to school	0.087	0.001	0.039
Hld with illiterate male adults	0.274	0.003	0.022
Hld with illiterate female adults	0.289	0.011	0.027
Sickness events for male per capita	0.002	0.605	0.271
Sickness events for female per capita	0.003	0.598	0.342
Hld with radio or TV	0.427	0.001	0.030
Hld with bicycle or motorcycle	0.443	0.006	0.007
Type of housing	0.399	0.008	0.020
Type of toilet rec	0.401	0.009	0.058
Hld with electricity	0.492	0.005	0.001
Number of sick days for male per capita	0.010	0.572	0.220
Number of sick days for female per capita	0.004	0.564	0.416
13*50% eigenvalue	**1.448**	**1.192**	**0.749**
Before eliminating intersections	2.886	2.339	0.135
		0.028	0.038
After eliminating intersections- 3 axis	2.876	2.339	

Table 3.3 summarizes the computation of the final rescaled weights appearing in the generalized CIP formula 3.26 above.

Table 3.3 Final categorical weights

INDICATOR	Category	Factorial score Axis # 1	Factorial score Axis # 2	Rescaled weights° 1000 Axis # 1	Rescaled weights° 1000 Axis # 2	Final weight
Male child. 6–15 not going to school	Yes	−1.109	−0.056	0		**0**
	No	0.059	0.003	2475		**2475**
Female child. 6–15 not going to school	Yes	−1.166	0.132	0		**0**
	No	0.074	−0.008	2627		**2627**
Hld with illiterate male adults	Yes	−2.04	0.222	0		**0**
	No	0.134	−0.014	4606		**4606**
Hld with illiterate female adults	Yes	−1.371	0.267	0		**0**
	No	0.211	−0.041	3352		**3352**
Sickness events male per cap.	No sickness	0.03	0.518		4936	**4936**
	0–1/2	−0.071	−0.734		2013	**2013**
	1/2–1	−0.065	−1.049		1277	**1277**
	I03>=1	−0.051	−1.596		0	**0**

Table 3.3 (continued)

INDICATOR	Category	Factorial score Axis # 1	Factorial score Axis # 2	Rescaled weights° 1000 Axis # 1	Rescaled weights° 1000 Axis # 2	Final weight
Sickness events female per cap.	No sickness	0.025	0.628		4584	**4584**
	0–1/2	−0.146	−0.54		1856	**1856**
	1/2 to 1	0.021	−0.832		1175	**1175**
	I03>=1	0.019	−1.335		0	**0**
Hld with radio or TV	With radio or TV	0.37	−0.021	3227		**3227**
	Without radio and TV	−1.153	0.067	0		**0**
Hld with bicycle or motorcycle	With bicycle/moto	0.265	−0.031	4102		**4102**
	Without bicycle/moto	−1.671	0.198	0		**0**
Type of housing	Multi-storey permanent	1.063	0.38	6379		**6379**
	One-storey permanent	0.577	−0.037	5350		**5350**
	Semi-permanent	0.08	0.013	4297		**4297**
	Temporary	−1.125	−0.084	1744		**1744**
	Not having owned house	−1.948	−0.035	0		**0**
Type of toilet	Flush toilet septic tank	0.841	0.088	3699		**3699**
	Double-vault compost latrine	0.584	−0.045	3155		**3155**
	On fish ponds	0.236	0.186	2417		**2417**
	Simple toilet/pit latrine	−0.482	0.016	896		**896**
	On river, canal	−0.505	−0.122	847		**847**
	Not having owned toilet	−0.905	−0.187	0		**0**
Hld with electricity	Hld with electricity	0.284	−0.03	4333		**4333**
	Hld without electricity	−1.761	0.185	0		**0**
Number of sick days male per cap.	No sick days	0.053	0.445		4086	**4086**
	Less than 5 days	−0.046	−1.255		117	**117**
	More than 5 days	−0.216	−1.305		0	**0**
Number of sick days female per cap.	No sick days	0.045	0.528		3853	**3853**
	Less than 5 days	−0.121	−0.956		388	**388**
	More than 5 days	−0.076	−1.122		0	**0**
			eigenvalue	0.22	0.18	