Paper presented at the International Conference on the "Many Dimensions of Poverty" organised by IPC, Brasilia, 29-31 August 2005

Preliminary Version
Not to be quoted

Multidimensional Poverty and Cluster Analysis: An Illustration with Switzerland

Giovanni Ferro Luzzi, Yves Flückiger
Sylvain Weber [§]

Abstract

The measurement of poverty has often been criticized for relying solely on measures of financial deprivation. Poverty being a multidimensional state, related to health, schooling, living environment, psychological state as well as social tides, care should be taken to integrate these various components to have a proper picture of poverty. This is also true for rich countries where often, poor financial conditions are alleviated by social policies, like minimum income, unemployment or housing benefits, whereas social exclusion and poor health can dominate the poverty feeling. We illustrate how some descriptive statistical tools can offer new insights in the context of multidimensional poverty. Factor analysis is used to construct poverty indicators based on many possible dimensions without posing too many a priori restrictions. These variables are examined to identify common factors which convey some aspect of multidimensional poverty. By ascribing individual scores on each factor, we then use cluster analysis to see what populations subgroups are more affected by the various dimensions of poverty. Finally, a survival analysis is run to find the determinants of falling into poverty.

[§] All authors University of Geneva, Switzerland.

I. Introduction

In the literature, the basic notion that poverty should be measured on the basis of as large a number of components (attributes) as is relevant and feasible has enjoyed increasing support. Since the seminal work of Sen (1976) and Townsend (1979) and others, it has been recognized that other aspects of life not necessarily related to income can impair human development, such as the access to public goods, health, or education. Many authors have come up with new approaches to provide poverty measures which account for its multidimensionality while maintaining desirable properties (Bourguignon and Chakravarty, 1999, 2002). One main conceptual issue is how to count multi-dimensional poverty. In other words, is multi-dimensional poverty the accumulation of deprivation in various components of what is considered "normal life" (the intersection approach) or should it be defined as the failure to access to one or more of the dimensions (the union approach)?

On the empirical side, a few studies have come out which aim at applying the idea of multidimensionality to the measurement of poverty. The UNDP human poverty index (HPI) is one such attempt which combines life expectancy, education, and health. This index, although widely used, has been criticized however for leaving out a monetary measure of poverty, while providing no clear basis for its weighting of the various sub-indices (see Bibi, 2002). The choice of what factors should be considered "poverty" or "deprivation" as well as the importance of each in the HPI has been said to be value laden.

The goal of this paper is threefold. One is to find a statistical tool that enables us to obtain a picture of poverty without too many a priori restrictions. Currently household panels include scores of variables which can be used to measure non-monetary poverty or deprivation. We shall describe how factor analysis can provide a meaningful description of poverty when many variables are possible candidates. This step leads to the construction of broader indicators of poverty which are common grounds to various subsets of variables, and which can be given a value for each individual.

The second goal of the paper is to identify the "poors" based on the newly constructed indicators of deprivations. To this end, we make use of another statistical tool, which is not very often favoured by economists, namely *cluster analysis*. With this method, we group individuals according to how similar they are with respect to the various scores of multiple deprivation. There, we attempt to see whether the "union" or "intersection" approach is more relevant to the data we use.

Finally, once we have identified the “multiply deprived”, we use survival analysis based on the panel dimension of the dataset to examine what affect the chances of remaining poor.

II. Pitfalls in the Measurement of Poverty

A large body of research exists on the measurement of monetary poverty. It is concerned with finding suitable indicators that best define poverty with all desirable features such indices should have to respect basic properties. Recently, some authors have attempted to extend this literature on the multidimensional direction. It is not our purpose to discuss such literature here, but it is worth mentioning some difficulties inherent to the multidimensional approach.

1. Choosing the indicators of poverty or deprivation

It is not easy to determine what and how many indicators should be taken into account for measuring deprivation. There is an obvious trade-off between the possible redundancy caused by overlapping information and the risk of obviating some important variables. Further, once the idea that poverty is a multidimensional state has been accepted, one still has to define how far poverty should be measured in the various directions. Some authors may argue that it is essential to include indicators relating to standards of living or social relations, while other may have a more restrictive view on what needs should be included.

2. “Absence by choice” or deprivation

Another practical problem is related to the distinction between “absence by choice” and deprivation. Preferences can affect the consumption choice of a good, service or activity that may be judged as “necessary”. Hence, an individual who does not have such goods or activities should be considered deprived only if she would consume them, could she afford them. This definition of deprivation can therefore only be used when the required information is collected in the household survey. One should however remain cautious in interpreting this kind of information, as the degree of value judgment can be very high and the subjective measure of own deprivation can be very different from one individual to another.

3. Aggregating and weighting the indicators

One further difficulty lies in the choice of aggregating or not the various indices of deprivation. Should they be averaged in a unique indicator of multidimensional poverty? This carries the clear advantage of summarizing the complexity of multiple dimensions in a simple way. On the other hand, such an aggregation causes a loss of information. Since a multidimensional phenomenon is studied, the search of a better description of such variety is an important goal (Sen, 1987).

One further problem with the choice of aggregating multidimensional poverty in a single index is the potential problem that arises when needing to define the weights of all subindices of deprivation.

Before aggregating indicators, it is necessary to establish a weighting structure for each one given their different features. If each one is considered as a deprivation indicator with different importance, then the researcher must assign a different weight to each variable to reflect their differences. The first option is an equal weighting for each element. It is used in some papers as Townsend (1979). Alternatively, we can compute the weightings from the data. One possible strategy relies on a weighting structure based on frequencies, so that they are calculated as a function of the relative frequencies of the variables. For example, Halleröd (1994) and Silber and Deutsch (2005) give more importance to deprivation of goods considered as necessary by larger groups of the population. The importance of each indicator can be also computed by means of different multivariate statistical methods, as factorial analysis as Nolan and Whelan (1996), principal components analysis, Ram (1982) and Maasoumi and Nickelsburg (1988), or cluster analysis (Hirschberg et al., 1991).

4. Threshold definition

This step is related to the aim of any poverty or deprivation analysis: the identification of the poor population. The main problem that arise any poverty analysis is the arbitrary nature of the threshold choice. Further, defining a poverty line implies that the population is divided into two groups, "poors" and "non-poors", which can be viewed as excessively restrictive in view of the multidimensional nature of poverty. Some authors have opted for an alternative methodology which relies on the concept of "fuzzy sets". In this case, different degrees of deprivation are assumed instead of a dichotomy between poor and non-poor.

III. Methodology

The vast majority of empirical studies on poverty use some index of financial deprivation, either the income or consumption of a person or a household. There are obvious advantages to using a money metric to measure poverty, as it is quite easily interpretable, transparent and more or less comparable across countries.

We are here interested in a more descriptive approach of multidimensional poverty, which hopefully, can bring some insights to this topic and can bring some answers to the pitfalls discussed above. Typically, in the empirical literature, the various components of poverty are treated as separate dimensions. One also find studies which construct an index incorporating the information from separate indicators, each of which reflecting deprivation in a specific field. As stated in the introduction, our ambition is to see whether some criticisms made on a poverty indicator like the HDI can be met by letting somehow the data speak for themselves. In other words, can the weights ascribed to each dimension of deprivation be computed from the data? The dimensions themselves will be selected on the basis of their relative importance in the data. The idea is very similar to that of Slottje's (1991), who has suggested that, when measuring the quality of life across countries, the indicators could be weighted by the variance of individual attributes. To this end, he uses the method of Principal Components Analysis.

Factor analysis has been used before in the study of poverty. Nolan and Whelan (1996) use it to select the most appropriate indicators of deprivation. Halleröd (1995) does not exclude any indicator but varies the weights. Here, we follow Dekkers (2004) in that Factor Analysis is used on all variables pertaining to some kind of deprivation and let the data determine how many *latent factors* are to be used, as well as the weights imposed on them. The approach is similar in spirit to Dewilde (2004) too, who uses Latent Class Analysis, a categorical variant of Factor Analysis. Finally, such an approach has also been used by Collicelli and Valerii (2000), who use principal component analysis on a set of developing countries.

Finally, even the threshold of poverty itself will be defined on the observation of the various sub-groups of the population in the dataset. This flexibility in the definition and measurement of multidimensional poverty carries the advantage that no subjective choice needs to be made. On the other hand, such an approach clearly also has some drawbacks. The indicators of multidimensional poverty we use have little to no linkage to the axiomatic approach. Further, some subjectivity cannot be avoided with the statistical methods we use, as we will see. The pattern of deprivation and the relations among variables, especially in cluster analysis are not always clear cut, so that some choices must be made based on judgement, rather than on strictly statistical tools.

The main idea to describe multidimensional poverty is based on the assumption that its various components translate into several variables, on which individuals accumulate deprivation. Each component therefore constitutes a given set of "capabilities", be it financial conditions, housing environment, social interactions, health or any other state that may hinder human development. Financial deprivation may translate into failure to repay debts, sacrificing vacations or unhealthy food purchases. Housing deprivation would imply smaller rooms, absence of central heating, or noisy living environment. In other words, each measured variable x_j is due to some *unobserved* common factors F_k and an idiosyncratic effect s_j:

$$x_j = \sum_k a_{jk} f_k + s_j \ ,$$

or, in matrix notation: $\mathbf{x} = \mathbf{A} \cdot \mathbf{f} + \mathbf{s}$

Where the **x** vector includes all observed (normalized) variables, **A** is the matrix of factor *loadings*, **f** is the vector of (latent) common factors, and **s** is the unique effects of the variables[1]. There are various methods to extract the factors, one simple way being through principal components[2], whereby the eigenvalues and eigenvectors of the correlation matrix on the observed variables are solved, with the added (scaling) constraint that the sum of the squared eigenvectors is equal to total variance.

One problem we must address is the fact that many of our deprivation indicators are dichotomous or ordinal with only a few scale steps. In such instances, it is known that the Pearson's correlation matrix is biased and will lead to biased estimates of the factor loadings if used as the basis for a factor analysis (Olsson, 1979). We will thus calculate the tetrachoric and polychoric correlation coefficients between our original indicators and use the resulting matrix as the starting point of our factor analysis. As Knol and Berger (1991) suggest, we will use unweighted least squares.

One problem encountered in the factor analysis is that the factor loadings matrix **A** defined above is not uniquely determined. To ensure a solution, one has to introduce constraints on the parameters in the original model. In general, one requires the first factor to have maximal contribution to the common variance of the observed variables, the second to have maximal contribution to this variance subject to being uncorrelated with the first, and so on. However, it is possible be that a more interpretable solution can be achieved using a transformed model, obtained by a process known as factor rotation. Various methods for the rotation of factors are available and we will make use of an oblique one (promax with power 3), which allows the factors to be correlated, rather than independent. In our case, this is indeed what we want, as we expect the different dimensions of poverty to be correlated.

The next step of the factor analysis is to decide how many factors are relevant to the model. As we shall see in the empirical part, this choice is not always clear and brings some additional difficulty.

---

1 The variables must be normalized as the procedure is sensitive to the units of measurement. Without this step, a variable with greater units could load higher on some factors, although the correlations are obviously the same.

2 See Everitt and Dunn (2001) for a detailed account of Factor Analysis.

Once a representation of the data in this form is considered adequate, each individual can be ascribed a "score" on the derived factors. These scores inform us on how each individual perform on each dimension of poverty. As all variables have been normalized, they indicate this performance relative to the mean of the population which is zero. Individuals with negative scores fare better than the average on these dimensions, while the opposite is true for positive scores. Obviously, with the four latent dimensions we found, some individuals may score negatively on some dimension and positively on some other.

We now want to see if we can identify some groups in the population which are more or less homogenous when using these measures of multidimensional poverty. To this end, we rely on Cluster Analysis. The latter is a technique which allows one to define relatively homogeneous groups on the basis of an original population of individuals having a priori very different characteristics. The main steps of the group identification procedure are as follows. Let there be n individuals with m characteristics (in our case the various scores of poverty). The goal is to determine groups of individuals having relatively similar characteristics, the opposite being true when individuals do not belong to the same group. The hierarchical agglomerative cluster method we will use leaves open the choice of the number of final groups. The distance between two clusters is the average distance between the scores in both clusters in a two-dimensional Euclidian space. The degree of similarity between two individuals is therefore measured by a proximity index, which is used to build a proximity matrix for all pairs of individuals. This matrix is then used for the classification procedure. In the first stage the values of the indices for all $n(n-1)/2$ potential pairs of individuals are compared and the two individuals who are found to be the closest are grouped. The same procedure is then applied on the basis of $(n-1)$ observations and hence of $(n-1)(n-2)/2$ distances. This procedure goes on until the desired number of groups is reached.

The number of clusters chosen should be such that the information loss is limited (the number of clusters is set as the number where pseudo-t^2 is maximal plus one) as well as that the difference between the clusters (the pseudo-F) is maximized.

Once the best groups are formed based on their similarity with respect to these dimensions of poverty, we are able to see if we can identify one or several groups of poors. As stated, it is theoretically possible that many groups are formed where only some dimensions are relevant to poverty. In other words, poors can be found either in the "intersection" sense (poverty in all dimensions) or the "union" sense, if some deprivation is compensated by some non-deprivation on another dimension. Against this prospect, we should also keep in mind that, by construction from Factor Analysis, some dimensions are more relevant than others since they capture covariation in the deprivation variables in decreasing order.

The final step of our analysis consists in finding the determinants of poverty. The group of poor individuals revealed by the Cluster Analysis is used and compared to the reference group. In order to find these determinants, we follow Dekkers (2004) by appealing to a simple panel logit model, where the dependent variable is the probability of falling into poverty, from one year to the next. "Poverty" will be here defined by taking individuals who do not belong to the major cluster of "non-poors".

IV. The Data

In this paper, we apply the technique presented in section 2 to Swiss data using the Swiss Household Panel (SHP). This panel dataset consists of 5 waves, from 1999 to 2003.

Questions are made to describe households as well as individuals and cover demographic, income, earnings, benefits, education, labour market status, description of housing and living conditions, possession of durables, mental and physical health, and so forth. It is very similar to the European Community Household Survey.

Table A1 in the Appendix is a list of the variables used for our analysis and Table 1 below provides basic statistics on these same variables. All these variables relate to some state of deprivation. As can be seen from Table A1 in the Appendix, many of these variables describe situation of *financial deprivation.* However, the data set also includes interesting information on the state of health, the housing conditions, the environment, as well as variables pertaining to "social exclusion". Further, we have also taken "subjective" variables indicating the level of satisfaction with the financial situation or life in general. It may seem strange to include such variables as they do not reflect deprivation of some kind, but how this (specific or general) deprivation is felt by the individual. Cheli and Lemmi (1995), as well as Dewilde (2004) use such variables too by reasoning that they may better reflect the everyday reality of poverty and deprivation.

In order to ease their interpretation, all variables have been constructed so that a greater value (or a value of one, if the variable is dichotomous) indicates a higher state of deprivation.

<Table 1 about here>

Table 1 gives the means, standard deviations, minimum and maximum value for each variable. In a headcount perspective, the mean deprivation levels vary quite substantially across “have/have not” variables from a low 2% almost 25%. If one looks at the variables close to what could be labelled “financial poverty”, the range is somewhat narrower, from around 2% to 13%. Similar values are observed for the other waves of the panel.

Before analyzing the results of our multidimensional approach, we present evidence on the prevalence of poverty and its development over time in Switzerland based on a unidimensional definition of poverty. Table A5 displays some basic results with respect to financial poverty using the most traditional indices such as the headcount ratio, the income poverty gap ratio or the Foster-Greer-Thorbecke (FGT) index (1984). These results have been computed for a poverty threshold fixed to 50% of the median equivalized household income. It is worth noting that the nominal relative poverty line has substantially increased in 2001 and 2002 while decreasing in 2003.

Looking at table A5, we observe that, in 1999, 7.8% of the entire population fall below the relative poverty line. This proportion stays more or less constant throughout the entire period analyzed despite the increase in the nominal threshold. A quiet different picture arises when looking at the income gap ratio which shows that, in 1999, the poors were on average 22.1% away from the annual threshold value. In other words, in 1999 a transfer of 22.1% of 24’000 CHF was required to bring each poor to the limit of poverty. Four years later, the transfer needed to eradicate poverty has been reduced to 17.8% of the relative poverty line fixed at 25’500 CHF.

Looking at the different values taken by the FGT for higher values of α, reveals that the proportion of people living far away from the poverty line is very small. By increasing the poverty aversion from 2 to 5 reduces the FGT index from 0.626 to 0.115 in 1999. This conclusion is even more pronounced for 2003 which is characterized by an even more marked decrease of the FGT index.

These results confirm the conclusions obtained by former studies on poverty in Switzerland (see Leu and Burri, 1999) which furthermore show that poverty rates measured on a unique financial dimension are somewhat higher for women than for men but the difference is not very significant. They also highlight that poverty rates tend to be higher among single males and single parent households, among foreigners as well as in the French and Italian speaking regions. Finally, this study shows that the poverty rates are very sensitive to the threshold definition.

V. Estimation of latent poverty factors and clusters of poors

Here we proceed with Factor Analysis in order to unravel what common factors best capture the covariance in all variables. As previously stated, many variables are dichotomous, although the latent deprivation indicator would be considered continuous. In such cases, tetrachoric and polychoric correlation coefficients must be used to estimate the factor loadings, as Pearson correlation coefficients would lead to biased estimates. The polychoric correlation matrix itself is not presented as it is not out primary interest. Suffice is to say that, as should be expected, variables tend to be more strongly correlated when they belong to the same "dimension", although some exceptions can be observed. Only a few coefficients display a negative sign, but they are never significantly different from zero. This matrix of correlations is then used to extract the factors via Principal Component Analysis.

The next step involves choosing the number of appropriate latent factors. To this end, we rely on some standard visual and statistical tools, commonly used in factor analysis, although one should be aware that most of these rules are somehow *ad hoc* and cannot avoid value judgments. One criterion which has been put forth is to exclude factors which have eigenvalues smaller than one, as the variance explained is less than that of the original variables. However this criterion is usually considered too lax, and should only be taken as an upper limit on the number of factors. Another criterion is to keep those factors so that the cumulated variance explained is no less than 70%. This criterion depends on the fit of the model. In our case, this would also imply too large a number of factors. The Scree diagram reported in the Appendix is used to find an "elbow" in the eigenvalue curve. Only the year 2001 is provided, as the curves are very similar for the other years[3].

In our case, the Scree plot seems to suggest the presence of a general factor, as suggested by a large first eigenvalue (8.459) and a much smaller second one (2.523). But one might argue that a secondary elbow occurs at the 5th eigenvalue implying a four-factor solution. Another way to use the Scree plot is to draw a straight line from the lowest eigenvalues. The threshold is where this line separates from the eigenvalue line. We thus decided that 4 factors were appropriate to describe the data, although the variance explained by the last two factors is somewhat on the low side.

[3] Table A2 in the appendix contains the eigenvalues, as well as the associated proportion of variance explained by each latent factor for years 1999-2003.

Next, we apply a factor rotation to provide a more meaningful representation of the loadings. As stated in the previous section, it makes sense to hypothesize that the common factors of deprivation (our four dimensions of poverty) are correlated. Indeed, one can assume that, say, "social exclusion" is positively correlated to "health" or "financial poverty".

Therefore, we apply an *oblique* rotation of the factor loadings. The resulting loadings for year 2001 are presented in Table 2. Again, very similar results are found for the other years.

<Table 2 about here>

A glance at Table 2 shows some clearly distinctive patterns. Indeed, the first 13 variables load positively and quite high on the first factor. These variables all pertain to financial deprivation, or deprivations in basic goods and services that are due to the lack of financial resources, sacrificing housing concessions, durable goods, but also other activities usually taken for granted. It is worth noting that the three subjective indicators of satisfaction with the financial situation also have high loadings on this factor. Hence factor 1 clearly reflects the dimension of "*Financial Poverty*". Income tightness still reflects the main hidden factor of poverty in Switzerland.

The second factor is clearly related to physical (variables *Health, Medication, and Handicap)* and mental (variables *Optimism*, *Depression, Life satisfaction*) health together. This latent dimension could be labelled "*Poor Health*".

The next dimension which seems to have some importance is one which could be named "*Bad Neighbourhood*". Three variables, noise, pollution and violence loads pretty high on this factor.

Finally, the fourth factor has high loadings for variables that are mostly related to social life, whether being member of an association, seeing friends or family, or simply going out. This latent factor is clearly associated to the dimension known as "*Social Exclusion*". It is worth mentioning that the three variables relative to physical health also have pretty high loadings in this factor, indicating that they have a clear impact on social life.

Let us also mention that factors 3 and 4 appear in reversed order for year 2003. This only means that for this particular year, the *"Social Exclusion"* was more important than the *"Bad neighbourhood"*. As shown in the column "proportion" of Table A2, the proportions of variance explained by these two factors are always close. This inversion is thus not unexpected.

Dekkers (2004) only identifies three factors for Belgium, which are financial poverty, social exclusion and poor *mental* health[4]. Dewilde (2004) has four underlying factors when analyzing British and Belgian household's panels. However, they are different from ours, as she lists "Housing", "Financial stress" and "Limited financial means" which jointly roughly correspond to our "Financial deprivation" and "Housing environment", which is the same as our "Bad Neighbourhood". Nevertheless, it must be emphasized that all these names are only subjective labels based on the examination of the loadings resulting from factor analysis and the rotation performed.

The second part of Table 2 gives the correlation coefficients among the four factors. As implied by the oblique rotation we operated on the loadings. It appears that factor 1, 2 and 4 are moderately and positively correlated, while factor 3 has no correlation to the other factors. It seems therefore that financial poverty, poor health and bad environment move together to some extent, whereas social exclusion is unrelated to the other dimensions. This last result is somewhat unexpected, as we had anticipated a strong positive correlation, at least with financial poverty, as found in Dekkers.

There is however some inconsistency across the years for this factor. In 1999, "Social Exclusion" is positively (but still moderately) correlated to all other dimensions but "Bad Neighbourhood", while it is slightly positively correlated to "Bad Neighbourhood" in 2000 and 2003. We therefore prefer to remain inconclusive as to the correlation of "Social Exclusion" to the other dimensions of poverty.

We now turn to the results of the Cluster Analysis. As detailed in the previous section, individuals are now grouped according to their relative distance, and the appropriate number of groups or "clusters" is determined by looking at various statistics. As is well-known, such an exercise is highly subject to value judgement. We considered two different statistics, namely: the pseudo-F developed by Calinski and Harabasz (1974), and the pseudo-t^2 which is a transformation of the Je(2)/Je(1) presented by Duda and Hart (1973). Large values of the pseudo-F index indicate distinct clustering and one must therefore maximize this statistic. On the contrary, a small value of the pseudo-t^2 indicates distinct clustering, and one should choose the number of clusters so that this index is low and has much larger values next to it. It may be advisable to look for consensus among the two statistics, that is, local peak of pseudo-

4 Dekkers only includes variables pertaining to mental health in his analysis, as he considers physical health more of a determinant of poverty, rather than a dimension of poverty itself. Although we admit that health could be taken as a determinant, we see no reason to separate mental and physical health with this respect.

F statistic combined with a small value of the pseudo-t^2 statistic and a larger pseudo-t^2 for the next cluster fusion.

Both of these statistics are displayed in Table 3, where the first 15 groups can be examined. Taking 1999 as an example, we see that the pseudo-F is maximized for 3 clusters, whereas the pseudo-t^2 is maximal for 8 groups, indicating the presence of 9 clusters. But notice that the pseudo-t^2 is also high for 2 groups, so that the solution of 3 clusters seems to be the best compromise. Applying the same reasoning to each year gives 2 clusters in 2000 and 2001, and 4 clusters in 2002 and 2003.

<Table 3 about here>

The dendrogram (or cluster tree) in Figure A2 of the Appendix graphically shows the information concerning which observations are grouped at various level. At the bottom of the dendrogram, each observation would be considered its own cluster. As one climbs up in the tree, observations are combined until all are grouped together, the vertical height indicating the similarity (or dissimilarity) of two groups. Creating 2 clusters tantamounts to cutting the tree horizontally where it has only two branches. Since they are among the longest branches, the 2 clusters we formed are actually very dissimilar.

<Table 4 about here>

Table 4 shows average score values for the various clusters found in each year. Typically, a very large cluster is found that contains most of the sample, and which can undoubtedly be defined as the "non poors" cluster. The mean scores are found to be negative on all dimensions of poverty, indicating that most persons are not deprived along these dimensions. A smaller second cluster is then found to have positive mean scores on every dimension. The individuals belonging to this cluster can thus be called "multidimensional poors", as they suffer from multiple deprivations. For the years where we formed more than 2 clusters, we see that they can be considered as outliers, as very few individual compose them. We finally obtain the following proportions of poors: 4.38% in 1999, 1.42% in 2000, 1.48% in 2001, 2.88% in 2002 and 2.98% in 2003.

VI. The Determinants of Poverty

Our goal is here to assess the determinants of multidimensional poverty. The clusters we found enable us to build a dichotomous variable stating whether a person belongs to a group of poors or non poors, for each year. One could in principle imagine that such a model be estimated as a multiple outcomes one, whereby individuals end up as poor, "partially poor" or non poor, if one accept the "union" approach that being deprived along some but not all dimensions can also be considered as poverty. However, because our results from the Cluster Analysis did not provide such clusters in a consistent manner for each year and with sufficient observations in each group, we reasoned that it would not be very meaningful. One important step is the choice of the determinants of poverty. Clearly, all variables that are assumed to be potential causes of poverty should be included. We therefore selected variables that pertain to human capital, and labour market status as well as variables that may capture discrimination in the labour market, such as age, gender, and nationality. We also included variables that may be more causes of social exclusion such as household composition, marital status and the like. Income is included as well (in equivalized units) so as to explain the financial dimension of poverty. Finally, we introduce a set of time dummy variables to capture the effect of being poor in a given year.

Our dependent variable being binary, we will use limited-dependent-variable models. Indeed, what we want to explain is the state of being poor, which can only be either true (1) or false (0). Several binary response models are available, such as probit, logit or complementary log-log. The latter is the most appropriate to analyze our data, since unlike the two others, it is asymmetric. This model is typically used when the positive outcome is rare, which is obviously our case with about 3% of poor individuals. Another desirable feature of the complementary log-log model is that it is the discrete-time equivalent of the Cox proportional hazard model which is widely used in duration analysis.

<Table 5 about here >

A glance at Table 5 reveals that most coefficients have the expected sign, though some variables appear to be insignificantly different from zero[5]. Gender, for instance, has no effect on poverty, per se. Dekkers (2003, 2004) also finds a slightly negative effect with other European countries. On the other hand, the marital status dummy variables clearly indicate that divorced persons have a higher probability of falling into poverty with respect to both unmarried, and married people. Also very much in line with expectations is the effect of education, which unambiguously lowers the chances of falling into poverty.

[5] Because of non-linearity, the coefficient itself is only qualitatively related to the impact of the variable on the probability of falling into poverty. The latter is shown in the "marginal effect" column. For dummy variables, the effect is computed by taking the difference in probability, when the dummy is raised from 0 to 1. All other variables are set to their mean sample value.

Unemployment is also a strong predictor of poverty, as well as being a foreigner. Somewhat unexpectedly, being retired does not have a significant effect on poverty, while age does increase the chances of being poor (albeit in a non linear fashion). It may be that poverty increases with age, especially when households have children and therefore greater needs, and decreases progressively as children become less of a burden for their parents.

In order to have a broader picture of poverty, we ran the same model on a simple headcount indicator of financial poverty (with poverty line set at half the median income)[6]. It appears that the estimation gives, broadly speaking, similar results. The coefficients on age are more difficult to interpret as they switch sign and only the quadratic term is positive and significant, and the dummy variable for being retired becomes positive and significant too. This suggests that age may have different effect, whether one looks at poverty in a strictly financial perspective, or more broadly defined on various indicators of deprivation. Other variables like the nationality, marital status and unemployment still have the same expected effect.

Concluding Comments

This paper has attempted to put forth some ideas to address well-known problems in the measurement of multidimensional poverty. The advantages of this approach can be summarized as follows. First, the number of dimensions as well as their relative importance is not determined *ex ante* but chosen on the basis of empirical regularities in the data. To this end, we have used factor analysis, although other statistical tools could be used alternatively. Their relevance is therefore directly dictated by their power in explaining the variance of various deprivation indicators, and we have found that such a method provides a parcimonious representation of multidimensional poverty.

Secondly, no poverty threshold needs to be set, since the population of multiply deprived persons is identified by looking at their similarities with respect to their scores on the various dimensions through Cluster Analysis. One concomitant advantage is that more than one group of poors can theoretically be identified, if clusters are found with different mean scores on the poverty dimensions. Based on the "union" approach of multidimensional poverty, some people could be identified as poors solely on some but not all dimensions. This evidently may call for different policy measures, depending on the degree of deprivation on each dimension. In our case, the clusters found showed only one relatively small group of poors, which would

[6] For obvious reasons of collinearity, the income variable is dropped from the model.

actually fit better with the "intersection" approach, since they were found to have positive mean scores on all dimensions.

Still, this approach also has some limits. No proper index if poverty aggregated over all dimensions may be computed, thus comparison is made difficult if one were to analyze different countries. Further, the statistical tools used (Factor Analysis and Cluster Analysis) may be subject to some arbitrariness, notably in the selection of the initial set of deprivation variables.

Finally, our approach does not distinguish a possible sequence of multidimensional poverty. Do people fall into poverty sequentially in a similar fashion along the various dimensions, or do they become poor in no clearly distinguishing pattern? Such an issue should be addressed from an empirical point of view, as it may provide precious guidelines to policymakers concerned with poverty. We plan to further investigate this theme in future research.

References

BIBI, S. (2002). “Mesures de la pauvreté dans une perspective multidimensionnelle: Une revue de la littérature”, Mimeo.

BOURGUIGNON, F. and S.R. CHAKRAVARTY (1999). “A Family of Multidimensional Poverty Measures”, in D.J. Slottje (ed.), *Advances in Econometrics, Income Distribution and Methodology of Science*, Essays in Honor of C. Dagum, Springer-Verlag, London.

BOURGUIGNON, F. and S.R. CHAKRAVARTY (2002). ”Measurement of Multidimensional Poverty”, *Journal of Economic Inequality*, forthcoming.

CALINSKI T. and J. HARABASZ (1974), “A dendrite method for cluster analysis”, *Communications in statistics*, 3, pp. 1-27.

DEKKERS, G. (2003), “Financial and multidimensional poverty in european countries: Can the former be used as a proxy of the latter?”, IRISS working paper, No. 2003-13.

DEKKERS, G. (2004). La perception de la pauvreté face à la réalité. Mesure de la pauvreté multi-dimensionnelle d’après les données du PSBH. In : Doutrelepont, R, Mortelmand, D. Casman, M.-Th. (Eds) : *Onze Ans de Vie en Belgique. Analyses Socio-économiques à partir du Panel Démographie Familiale.* Gent : Academia Press. pages 131-156.

DEUTSCH, J. and SILBER, J. (2005), “Measuring multidimensional poverty: An empirical comparison of various approaches”, *Review of income and wealth*, 51(1), pp. 145-74.

DEWILDE, C. (2004), “The multidimensional measurement of poverty in Belgium and Britain: A categorical approach”, *Social indicator research*, 68, pp. 331-69.

DRASGOW, F. (1969), “Polychoric and polyserial correlations”, In Kotz L, Johnson NL (Eds.), *Encyclopedia of statistical sciences*, Vol. 7, pp. 68-74, New York: Wiley, 1988.

DUDA R.O. and P.E. HART (1973), “Pattern classification and scene analysis” New York: John Wiley & Sons.

EVERITT, B.S. & DUNN, G. (2001) *Applied Multivariate Data Analysis*. Edward Arnold, London.

FILMER, D. and PRITCHETT, L.H. (2001), “Estimating wealth effects without expenditure data – or tears: An application to educational enrollments in states of India”, *Demography*, 38(1), pp. 115-32.

HALLERÖD, B (1995). The Truly Poor : Direct and Indirect Consensual Measurement of Poverty in Sweden. *European Journal of Social Policy,* 5(2), pp. 111-129.

HIRSCHBERG, J. MAASOUMI, E. and SLOTTJE, D. (1991). Cluster Analysis of Measuring Welfare and Quality of Life across Countries. *Journal of Econometrics,* 50(3), pp. 131-150.

KHAWAJA, M. (2003), “Migration and the reproduction of poverty: The refugee camps in Jordan”, *International migration*, 41(2), pp. 27-57.

KNOL, D. BERGER, M. (1991). Empirical Comparison between Factors Analysis and Multidimensional Item Response Models. *Multivariate Behavioral Research*, 26(3), pp. 457-477.

KOLENIKOV, S. and ANGELES, G. (2004), "The use of discrete data in PCA: Theory, simulations, and applications to socioeconomic indices", Working paper, University of North Carolina.

LAYTE, R.; MAÎTRE, B.; NOLAN, B., and WHELAN, C.T. (2001), "Persistent and consistent poverty in the 1994 and 1995 waves of the european community househopld panel survey", *Review of income and wealth*, 47(4), pp. 427-49.

LEU, R.; BURRI, S. (1999). Poverty in Switzerland. *Swiss Journal of Economics and Statistics* 135(3), pp. 303-328.

MILLIGAN, G.W. and COOPER, M.C. (1985), "An examination of procedures for determining the number of clusters in a data set", *Psychometrika*, 50(2), pp. 159-79.

NOLAN, B. WHELAN, C. (1996). Measuring Poverty Using Income and Deprivation Indicators: Alternative Approaches. *Journal of European Social Policy,* 6(3), pp. 225-240.

OLSSON, U. (1979), "Maximum likelihood estimation of the polychoric correlation coefficient", *Psychometrika*, 44(4), pp. 443-60.

OLSSON, U.; DRASGOW, F. and DORANS, N.J. (1982), "The polyserial correlation coefficient", *Psychometrika*, 47(3), pp. 337-47.

SLOTTJE, D. (1991). Measuring the Quality of Life across Countries, *Review of Economics and Statistics,* 73(4), pp. 648-754.

TOWNSEND (1979). *Poverty in the United Kingdom,* Penguin, Harmondsworth.

TSUI, K. (2002), "Multidimensional poverty indices", *Social choice and welfare*, 19, pp. 69-93.

Tables and Figures

Table 1: Descriptive statistics for the variables used in factor analysis, SHP 2001

Variable	Mean	Std. Dev.	Min	Max
Bills	0.079	0.270	0	1
Save 100CHF	0.120	0.325	0	1
Private retirement scheme	0.099	0.299	0	1
Ends meet	2.582	2.054	0	10
Income below needs	0.099	0.298	0	1
Housing small	0.119	0.324	0	1
Heating	0.072	0.258	0	1
Vacation	0.060	0.237	0	1
Invite	0.030	0.170	0	1
Restaurant	0.125	0.331	0	1
Car	0.023	0.151	0	1
Dishwasher	0.020	0.139	0	1
Dentist	0.023	0.149	0	1
Computer	0.026	0.158	0	1
Financial satisfaction	2.700	2.043	0	10
HH financial satisfaction	2.484	1.871	0	10
Living standards satisfaction	0.209	0.407	0	10
Noise	0.145	0.352	0	1
Pollution	0.118	0.323	0	1
Violence	0.877	0.662	0	1
Health	1.362	2.511	0	4
Medication	1.867	2.026	0	10
Depression	1.892	1.456	0	10
Life satisfaction	2.462	1.709	0	10
Optimism	0.186	0.389	0	10
Handicap	0.247	0.431	0	1
Affiliation	2.947	0.840	0	1
Cinema	3.187	0.994	0	4
Sports	1.748	0.969	0	4
Bar	3.047	0.777	0	4
Theatre	22.980	7.848	0	4
Contacts	0.079	0.270	0	30
Nb obs	6416			

Table 2: Rotated Factor Loadings (Oblique Rotation), SHP 2001

Variable	Factor 1	Factor 2	Factor 3	Factor 4
Bills	0.673	0.072	-0.087	-0.165
Save 100CHF	0.816	-0.028	-0.064	0.087
Private retirement scheme	0.702	-0.070	0.058	0.023
Ends meet	0.788	0.062	-0.034	-0.112
Income below needs	0.518	0.001	0.041	0.151
Housing small	0.166	0.046	0.103	-0.153
Heating	0.269	0.147	0.039	-0.055
Vacation	0.735	-0.019	-0.114	0.139
Invite	0.726	-0.098	0.069	0.161
Restaurant	0.717	-0.145	-0.006	0.143
Car	0.583	0.016	0.140	0.090
Dishwasher	0.556	-0.079	0.217	0.032
Dentist	0.744	0.049	-0.036	-0.053
Computer	0.715	0.030	0.043	0.112
Financial satisfaction	0.499	0.257	-0.005	-0.172
HH financial satisfaction	0.768	0.102	-0.022	-0.176
Living standards satisfaction	0.476	0.162	0.038	-0.154
Noise	0.025	0.007	0.800	-0.049
Pollution	0.006	-0.011	0.878	-0.008
Violence	0.075	0.046	0.418	0.018
Health	0.069	0.488	0.009	0.357
Medication	-0.032	0.400	0.010	0.508
Depression	0.050	0.624	0.006	0.027
Life satisfaction	0.246	0.542	0.009	-0.111
Optimism	-0.081	0.684	0.002	0.067
Handicap	-0.007	0.384	0.004	0.347
Affiliation	0.181	0.050	-0.010	0.163
Cinema	0.058	-0.043	-0.048	0.615
Sports	-0.064	0.149	0.077	0.348
Bar	0.142	-0.058	-0.015	0.330
Theatre	0.252	-0.095	-0.039	0.235
Contacts	-0.095	0.053	-0.025	0.314

Inter-factor correlations

	Factor 1	Factor 2	Factor 3	Factor 4
Factor 1	1.000	0.323	0.226	0.032
Factor 2	0.323	1.000	0.198	–0.004
Factor 3	0.226	0.198	1.000	0.080
Factor 4	0.032	–0.004	0.080	1.000

Note : The *Promax* method of oblique rotation has been used, with a power of 3.

Table 3: Statistics for determining the number of clusters

Number of clusters	1999		2000		2001	
	Pseudo-F	Pseudo-t^2	Pseudo-F	Pseudo-t^2	Pseudo-F	Pseudo-t^2
1	-	27.61	-	574.06	-	506.23
2	27.61	1263.16	574.06	5.13	506.23	27.44
3	647.66	24.48	292.51	27.84	284.76	19.96
4	451.26	113.86	204.92	36.99	197.10	16.64
5	391.79	23.60	168.37	18.93	156.02	27.66
6	322.52	54.67	139.45	9.86	133.27	22.90
7	282.30	28.62	118.00	909.70	115.43	1997.17
8	247.32	1820.89	243.86	5.76	413.69	18.50
9	489.08	115.99	214.26	29.06	363.99	91.57
10	454.63	609.26	194.46	48.52	340.44	226.49
11	498.27	7.53	183.61	2.96	337.42	6.50
12	453.90	43.63	167.24	2.99	307.70	779.43
13	421.28	22.33	153.58	10.73	380.13	451.25
14	391.76	431.94	142.18	15.57	408.72	9.15
15	407.55	16.01	132.97	135.15	380.83	20.25

Number of clusters	2002		2003	
	Pseudo-F	Pseudo-t^2	Pseudo-F	Pseudo-t^2
1	-	68.67	-	56.16
2	68.67	90.49	56.16	57.21
3	80.11	604.58	56.99	575.27
4	260.95	143.4	233.96	47.60
5	235.36	68.17	197.54	78.41
6	211.59	10.92	176.02	13.31
7	179.48	6.75	149.81	2081.03
8	154.68	60.85	472.72	3.24
9	143.63	19.26	414.19	51.42
10	130.16	1653.05	377.2	40.39
11	316.03	559.06	346.15	13.27
12	365.91	34.26	317.27	7.47
13	340.12	27.48	291.87	39.23
14	317.25	27.22	274.79	477.91
15	298.34	141.83	311.29	3.93

Table 4: Mean Scores on the 4 Factors, by Cluster, 1999-2003

	Factor 1	Factor 2	Factor 3	Factor 4	Observations	%
1999						
Cluster 1	-0.092	-0.074	-0.026	-0.040	7397	95.58
Cluster 2	1.970	1.578	0.562	0.828	339	4.38
Cluster 3	3.374	2.777	1.557	3.429	2	0.03
2000						
Cluster 1	-0.050	-0.048	-0.015	-0.004	6684	98.58
Cluster 2	2.829	2.182	0.506	0.013	96	1.42
2001						
Cluster 1	-0.049	-0.038	-0.012	-0.011	6321	98.52
Cluster 2	2.863	1.885	0.564	0.217	95	1.48
2002						
Cluster 1	-0.076	-0.052	-0.031	-0.041	5376	96.71
Cluster 2	2.338	1.147	0.682	1.008	160	2.88
Cluster 3	-0.291	2.634	2.195	1.217	17	0.31
Cluster 4	2.879	4.004	2.222	-0.148	6	0.11
2003						
Cluster 1	-0.089	-0.051	-0.016	-0.020	4943	96.81
Cluster 2	2.575	1.213	0.387	0.574	152	2.98
Cluster 3	1.991	3.433	-1.712	0.148	8	0.16
Cluster 4	3.737	5.515	1.771	1.445	3	0.06

Table 5: Complementary log-log model of the probability of falling into multidimensional poverty

Variable	Coeff.	Std. Err.	Marginal Effect [b]	Std. Err.
Year 2000	-1.419	0.140	-0.001935	0.00033
Year 2001	-1.335	0.141	-0.001816	0.00032
Year 2002	-0.218 [a]	0.114	-0.000384 [a]	0.00020
Year 2003	-0.172 [a]	0.118	-0.000306 [a]	0.00021
Age/10	0.599	0.249	0.001126	0.00049
$(Age/10)^2$	-0.057	0.026	-0.000107	0.00005
Gender (female=1)	-0.063 [a]	0.135	-0.000118 [a]	0.00026
Married	-0.418	0.188	-0.000820	0.00040
Divorced	0.829	0.208	0.002189	0.00080
Single parent family	0.999	0.179	0.003029	0.00094
Educ2	-0.305 [a]	0.247	-0.000502 [a]	0.00036
Educ3	-1.000	0.153	-0.001997	0.00044
Educ4	-1.843	0.257	-0.002020	0.00035
Educ5	-2.077	0.266	-0.002071	0.00035
Part-time	0.476	0.171	0.001042	0.00046
Student	-0.650	0.303	-0.000976	0.00039
At home	0.669	0.222	0.001667	0.00074
Retired	0.986	0.282	0.002788	0.00123
Unemployed	2.598	0.189	0.021602	0.00507
Other occupation	0.280 [a]	0.432	0.000606 [a]	0.00108
EU15	0.896	0.168	0.002525	0.00076
NonEU	2.041	0.326	0.012164	0.00443
Other nationality	1.969	0.433	0.011342 [a]	0.00581
French Speaking Region	0.825	0.264	0.001913	0.00080
German Speaking Region	-0.393 [a]	0.261	-0.000796 [a]	0.00058
Intercept	-6.241	0.642		
Ln σ^2_u	1.597	0.090		
σ_u	2.222	0.100		
Rho	0.750	0.017		
Log L	–2741.702			
Observations		27495		
Groups (individuals)		8575		

Notes: [a] Coefficient is not significant at the 0.05 level.

[b] Marginal effect evaluated at the mean of every variable and assuming that the random effect for that observation's panel is zero. For binary variables, variation of the probability of a "positive" outcome is calculated for a discrete change from 0 to 1.

Table 6: Complementary log-log model of the probability of falling into financial poverty (Equivalized Income less than half the median income)

Variable	Coeff.	Std. Err.	Marginal Effect	Std. Err.
Year 2000	0.155	0.078	0.001493 [a]	0.00079
Year 2001	0.143 [a]	0.080	0.001372 [a]	0.00081
Year 2002	0.258	0.083	0.002588	0.00092
Year 2003	0.161 [a]	0.086	0.001567 [a]	0.00089
Age/10	-0.347 [a]	0.190	-0.003198 [a]	0.00175
$(Age/10)^2$	0.052	0.020	0.000483	0.00018
Gender (female=1)	-0.017 [a]	0.103	-0.000155 [a]	0.00095
Married	0.354	0.154	0.003185	0.00137
Divorced	0.822	0.179	0.010589	0.00316
Single parent family	1.245	0.144	0.020932	0.00409
Educ2	-0.105 [a]	0.199	-0.000925 [a]	0.00168
Educ3	-0.749	0.115	-0.007165	0.00131
Educ4	-1.628	0.194	-0.009212	0.00098
Educ5	-1.949	0.206	-0.009835	0.00098
Part-time	0.518	0.144	0.005638	0.00187
Student	1.073	0.200	0.015405	0.00428
At home	1.402	0.155	0.024031	0.00457
Retired	1.600	0.202	0.029420	0.00695
Unemployed	1.620	0.190	0.035212	0.00813
Other occupation	1.727	0.281	0.040819	0.01356
EU15	-0.006 [a]	0.165	-0.000056 [a]	0.00151
NonEU	1.237	0.263	0.021933	0.00806
Other nationality	1.768	0.323	0.043183	0.01647
French Speaking Region	-0.373 [a]	0.202	-0.003187 [a]	0.00164
German Speaking Region	-0.598	0.198	-0.006204	0.00240
Intercept	-4.169	0.449		
$\ln \sigma^2_u$	1.567	0.058		
σ_u	2.189	0.063		
Rho	0.744	0.011		
Log L	–5569.238			
Observations	27495			
Groups (individuals)	8575			

Notes: [a] Coefficient is not significant at the 0.05 level.

[b] Marginal effect evaluated at the mean of every variable and assuming that the random effect for that observation's panel is zero. For binary variables, variation of the probability of a "positive" outcome is calculated for a discrete change from 0 to 1.

Appendix

Table A1: Description of the variables used in the factor analysis, SHP

Variable	Label
Bills	Bills unpaid in the last 12 months (1=yes, 0 = no).
Save 100CHF	Cannot afford to save CHF 100 per month (1=cannot; 0=can).
Private Retirement scheme	Cannot afford to apply for retirement saving scheme (1=cannot; 0=can).
Ends meet	Difficulties in making ends meet with current income (from 0= “no difficulty” to 10= “highest difficulty”).
Income below needs	Household incomes are below necessary income (1=yes, 0=no).
Housing small	House or flat is too small (1=yes, 0=no).
Heating	Heating in house is bad (1=yes, 0=no).
Vacation	Cannot afford one week of vacation (1=cannot; 0=can).
Invite	Cannot afford to invite friends once a month (1=cannot, 0=can).
Restaurant	Cannot afford restaurant once a month (1=cannot, 0=can).
Car	Cannot afford a private car (1=cannot, 0=can).
Dishwasher	Cannot afford a dishwasher (1=cannot, 0=can).
Dentist	Cannot afford visit to the dentist if necessary (1=cannot, 0=can).
Computer	Cannot afford a computer at home (1=cannot, 0=can).
Financial satisfaction	0=very satisfied with financial situation, 10= not at all.
Household fin. Satisfaction	0=very satisfied with household financial situation, 10= not at all.
Living standards satisf.	Satisfaction with living standards (0=very satisfied, 10=not at all)
Noise	Noisy environment (0=no, 1=yes)
Pollution	Problems with polluted environment (0=no, 1=yes).
Violence	Problems with delinquency or vandalism around the house (0=no, 1=yes)
Health	State of health (0=very good, 4=very bad).
Medication	Needs of medication (0=no,10=very high).
Depression	Frequency of negative feelings (0=never, 10=always)
Life satisfaction	Satisfaction with life in general (0=very satisfied, 10=not at all)
Optimism	Optimism feeling frequency (0=always, 10=never)
Handicap	Long term health problem or disability of a psychological or physical nature (0=no, 1=yes)
Affiliation	Passive or active member of whatever association (0=yes, 1=no)
Cinema	Frequency of going to the cinema (0=every day, 1=at least once a week; 2 =at least once a month, 3=less than once a month, 4=never)
Sports	Frequency of going to sporting events (0=every day, 1=at least once a week; 2 =at least once a month, 3=less than once a month, 4=never)
Bar	Frequency of going to a bar, pub, restaurant (0=every day, 1=at least once a week; 2 =at least once a month, 3=less than once a month, 4=never)
Theatre	Frequency of going to the theatre(0=every day, 1=at least once a week; 2 =at least once a month, 3=less than once a month, 4=never)
Contacts	Contacts with close friends per month (range is from 0=more than 30, to 30= no contact)

Figure A1 Scree Diagram for 2001 factor analysis

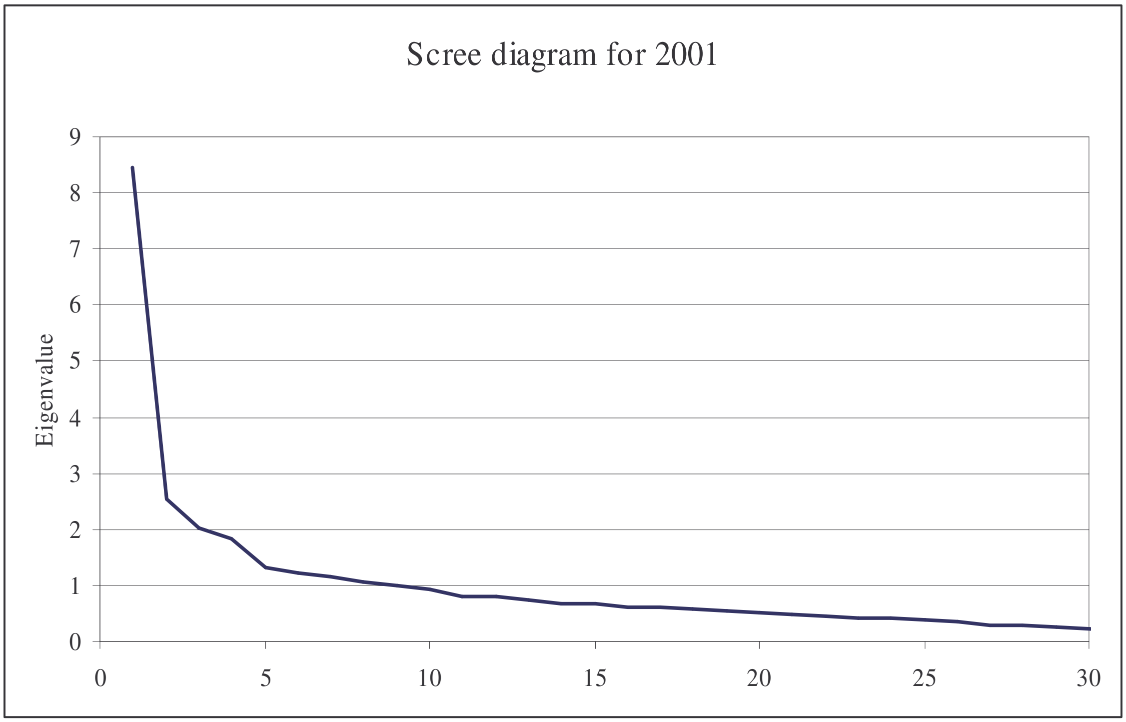

Table A2 Eigenvalues and Proportion of Variance Explained, SHP 1999–2003

Factor	1999			2000			2001		
	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative
1	8.548	0.285	0.285	8.417	0.263	0.263	8.459	0.264	0.264
2	2.309	0.077	0.362	2.474	0.077	0.340	2.523	0.079	0.343
3	1.901	0.063	0.425	2.094	0.065	0.406	2.013	0.063	0.406
4	1.665	0.056	0.481	1.793	0.056	0.462	1.828	0.057	0.463
5	1.153	0.038	0.519	1.245	0.039	0.501	1.308	0.041	0.504
6	1.132	0.038	0.557	1.189	0.037	0.538	1.224	0.038	0.542
7	1.085	0.036	0.593	1.134	0.036	0.573	1.145	0.036	0.578
8	1.016	0.034	0.627	1.066	0.033	0.607	1.057	0.033	0.611
9	0.929	0.031	0.658	0.953	0.030	0.636	0.990	0.031	0.642
10	0.834	0.028	0.686	0.921	0.029	0.665	0.945	0.030	0.672
11	0.809	0.027	0.713	0.860	0.027	0.692	0.817	0.026	0.697
12	0.726	0.024	0.737	0.772	0.024	0.716	0.794	0.025	0.722
13	0.699	0.023	0.760	0.763	0.024	0.740	0.748	0.023	0.745
14	0.655	0.022	0.782	0.701	0.022	0.762	0.690	0.022	0.767
15	0.616	0.021	0.803	0.686	0.021	0.783	0.667	0.021	0.788
16	0.581	0.019	0.822	0.641	0.020	0.803	0.612	0.019	0.807

Factor	2002			2003		
	Eigenvalue	Proportion	Cumulative	Eigenvalue	Proportion	Cumulative
1	8.553	0.267	0.267	8.765	0.274	0.274
2	2.517	0.079	0.346	2.544	0.080	0.353
3	2.113	0.066	0.412	2.072	0.065	0.418
4	1.885	0.059	0.471	1.910	0.060	0.478
5	1.335	0.042	0.513	1.333	0.042	0.520
6	1.207	0.038	0.550	1.188	0.037	0.557
7	1.179	0.037	0.587	1.135	0.036	0.592
8	1.146	0.036	0.623	1.132	0.035	0.627
9	0.965	0.030	0.653	1.073	0.034	0.661
10	0.896	0.028	0.681	0.953	0.030	0.691
11	0.823	0.026	0.707	0.844	0.026	0.717
12	0.808	0.025	0.732	0.768	0.024	0.741
13	0.712	0.022	0.754	0.735	0.023	0.764
14	0.667	0.021	0.775	0.721	0.023	0.787
15	0.650	0.020	0.796	0.660	0.021	0.807
16	0.612	0.019	0.815	0.623	0.020	0.827

Figure A2 Dendrogram for 2001 cluster analysis

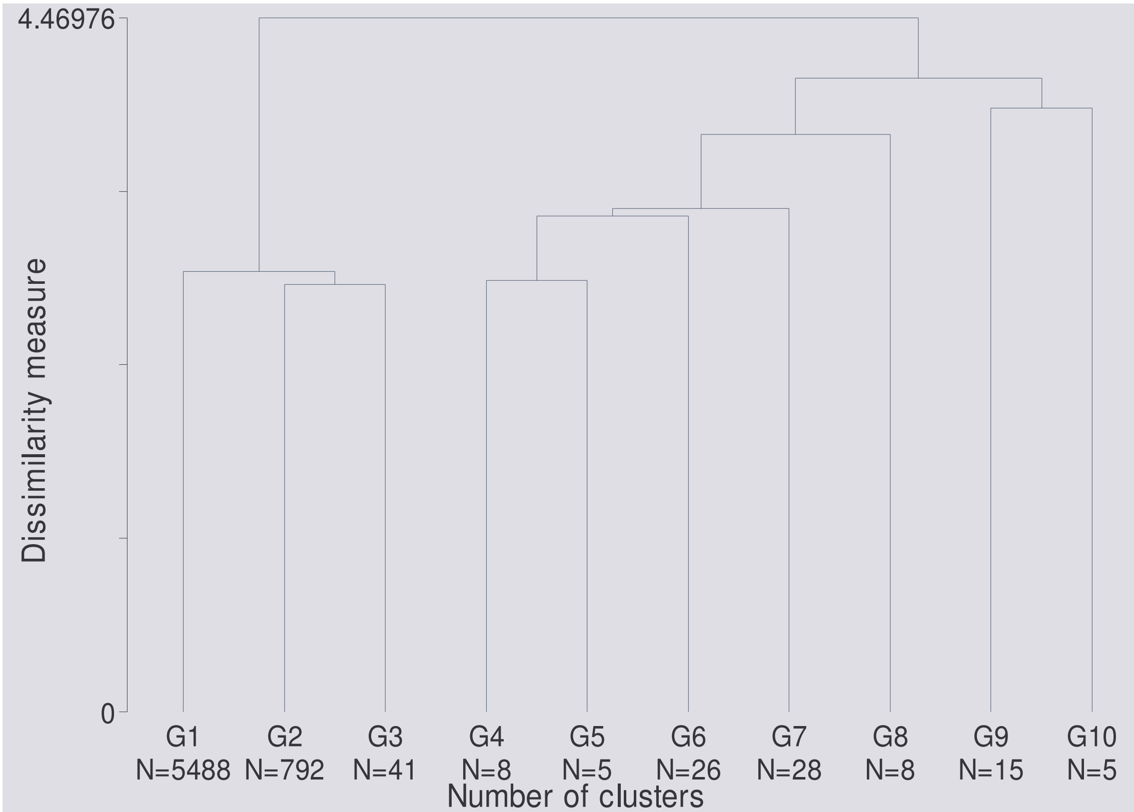

Table A3 Description of the variables used in the cloglog estimation, SHP

Variable	Label
Income	Yearly household net income equivalized (OECD), in thousands CHF
Age/10	Age in ten years
(Age/10)^2	Age in ten years squared
Gender	Gender (0=male, 1=female)
Single	Civil status: single
Married	Civil status: married
Divorced	Civil status: divorced separated or widow
Single parent	Parent living alone with one or more children
Educ1	Education: compulsory school or less
Educ2	Education: domestic science/general training course
Educ3	Education: maturity/apprenticeship
Educ4	Education: technical/vocational school.
Educ5	Education: university, higher specialised school
Fulltime	Occupation: fulltime job
Part-time	Occupation: part-time job
Student	Occupation: student, apprentice
Athome	Occupation: housekeeping
Retired	Occupation: retired
Unemployed	Occupation: unemployed or invalid insurance
Other occupation	Occupation: other
Swiss	Nationality: Switzerland
EU15	Nationality: European Union 15
NonEU	Nationality: Europe but outside EU15
Other nationality	Nationality: outside of Europe
French	French speaking region
Italian	Italian speaking region
German	German speaking region

Table A4 Descriptive statistics for the variables used in cloglog estimation, SHP 2001

Variable	Mean	Std. Dev.	Min	Max
Income	57.443	45.621	6	1264.8
Age/10	4.260	1.655	1.3	9
Gender	0.543	0.498	0	1
Single	0.295	0.456	0	1
Married	0.589	0.492	0	1
Divorced	0.116	0.321	0	1
Single Parent	0.060	0.238	0	1
Educ1	0.190	0.392	0	1
Educ2	0.051	0.220	0	1
Educ3	0.514	0.500	0	1
Educ4	0.135	0.342	0	1
Educ5	0.110	0.313	0	1
Fulltime	0.410	0.492	0	1
Part-time	0.201	0.401	0	1
Student	0.131	0.337	0	1
Athome	0.108	0.311	0	1
Retired	0.118	0.323	0	1
Unemployed	0.020	0.141	0	1
Other occupation	0.011	0.104	0	1
Swiss	0.894	0.308	0	1
EU15	0.084	0.278	0	1
NonEU	0.013	0.113	0	1
Other nationality	0.009	0.095	0	1
French	0.275	0.447	0	1
Italian	0.047	0.211	0	1
German	0.678	0.467	0	1
Nb of Obs.	5567			

Note: the number of observations in not the same as in Table 1 because income was not available for some individuals.

Table A5 Various Indices of Financial Poverty, SHP 1999-2003

	1999	*2000*	*2001*	*2002*	*2003*
Poverty Line[a]	24'000	24'120	24'857	25'714	25'500
Headcount ratio %	7.8	7.533	7.329	7.908	7.521
Aggregate poverty gap[b]	413.14	295.83	279.98	360.96	340.69
Poverty gap ratio %	1.721	1.227	1.126	1.404	1.336
Income gap ratio %	22.071	16.281	15.369	17.750	17.763
Watts index	2.210	1.559	1.398	1.797	1.643
Index FGT(0.5)*100	3.355	2.542	2.495	3.011	2.820
Index FGT(1.5)*100	0.996	0.692	0.604	0.763	0.727
Index FGT(2.0)*100	0.626	0.428	0.360	0.467	0.430
Index FGT(2.5)*100	0.416	0.281	0.231	0.312	0.269
Index FGT(3.0)*100	0.289	0.194	0.156	0.223	0.174
Index FGT(3.5)*100	0.208	0.138	0.110	0.167	0.116
Index FGT(4.0)*100	0.153	0.102	0.080	0.131	0.079
Index FGT(4.5)*100	0.115	0.077	0.059	0.105	0.055
Index FGT(5.0)*100	0.089	0.059	0.045	0.087	0.039
Clark et al. Index (0.10)*100	2.149	1.517	1.365	1.746	1.607
Clark et al. index (0.25)*100	2.064	1.459	1.318	1.675	1.555
Clark et al. index (0.50)*100	1.936	1.372	1.247	1.571	1.475
Clark et al. index (0.75)*100	1.822	1.295	1.183	1.481	1.402
Clark et al. index (0.90)*100	1.760	1.253	1.148	1.434	1.362
Sen index *100	2.474	1.915	1.728	2.064	1.988
Thon index *100	3.367	2.412	2.214	2.749	2.620
Takayama index *100	1.675	1.201	1.100	1.364	1.302
Observations	6321	5907	5567	5020	4680

[a] *Poverty line is set at half the median equivalized income.*
[b] *Units of income per observation.*