Identifying deprivation profiles in Spain: a new approach[1]

Jesús Pérez-Mayo

Department of Applied Economics, University of Extremadura, Elvas Av., 06071 Badajoz (Spain), fax: 34 924 272 509, jperez@unex.es

Abstract

Usually, an indirect approach for measuring deprivation or poverty is used with poverty lines. However, some studies have used a direct approach to measure deprivation or poverty. The aim of this paper is improving the identification of the poor people. The central point of the concept of deprivation we use is related to the opportunity to have or do something. Therefore, deprivation means an inability to get the goods, facilities and opportunities, which are usual in the household environment. Since all of the needed variables are categorical, we use the latent class model to solve this problem because is the best model to achieve this objective.

JEL classification numbers: I31, I32

Keywords: deprivation; latent class analysis; poverty

[1] This manuscript is a pre-print version of the article published in Applied Economics (2005), volumen 37 (8), pages 943-955, available in http://www.tandfonline.com/doi/abs/10.1080/00036840500060972.

1. Introduction

In contrast to much previous research on poverty, which has considered issues as the unemployment-inequality trade off, the effects of growth, dynamics of income distribution, etc. (see e.g. Ayala, Martinez and Ruiz-Huerta, 2002; Enders and Hoover, 2003; Cohen, 1998), in which the analysis and measurement of poverty is considered as a subsidiary issue, the analysis and measurement of poverty is the main focus of this paper.

According to the European Council (1984), quoted in Eurostat (2000),"the poor people are those individuals, families or groups whose material, cultural and social resources are so limited that they are excluded from the minimum standard of living of the society where they live". In the previous quote, this multidimensional concept of poverty is more related to the standard of living of the person or household, more than the simple disability of satisfying the maintenance needs.

Nevertheless, some problems appear when poverty is measured: how standard of living is measured, which is this "minimum standard of living", when someone is under such minimum. In the most of the empirical studies on poverty, the standard of living is measured by the household income adjusted to household size by means of equivalence scales. Thus, a household is poor if its equivalent income is under a threshold (called poverty threshold or poverty line). Depending on the studies, it is defined as the 50 or 60 per cent of the mean or median income. Although this method has some advantages, easy computation and comparison between different periods or territories, it also has some drawbacks, following Martínez and Ruiz-Huerta (1999):

- The length of the reference period.
- Some non - monetary variables need to be included.
- Wealth is not included.
- It is difficult to evaluate household necessities.
- Underestimation.

Expenditure is also proposed as an indirect indicator of the standard of living because of the lower underestimation and, furthermore, the distortions derived from the current feature of the income. This advantage is related to consumption theory. According to the classical consumption theory, current expenditure is a better approximation of the permanent income than current income. However, expenditure also presents some drawbacks. It is difficult to estimate the annual expenditure from weekly data and, besides, it depends on consumption patterns. Therefore, the relationship between a low expenditure and a shortage of resources is not always right.

Once the problems of indirect indicators are exposed, it is logical to think on direct measures. Ringen's criticisms (1988) to the usual methodology of poverty measurement support theoretically the decision of incorporating direct non-monetary indicators. He said that income is a good indicator for poverty only if low income and low standard of living are tightly related. Concretely, he argued the inconsistency of indirect measuring of a direct and multidimensional variable by means of income. Furthermore, resources are not always applied for achieving goods considered as necessary. Therefore, low-income levels are not very reliable for identifying the most deprived households. Other advantages of direct indicators are:

- They describe better the poor (by income) households. Here, we can speak on living conditions of the poor population.
- Without leaving the income criteria, these indicators allow us to improve poor individuals or households identification. If there is a strong relationship between income and standard of living, they can be useful to determine a poverty threshold[1]. Otherwise, as Ringen (1988) argues, a combination of both indicators can provide a correct identification of poor people if such hypothesis is rejected.
- Finally, they can be used as an alternative indicator to measure poverty. As Martínez and Ruiz-Huerta (1999) expose, the theoretical support is found in the “standard of

living" approach (Atkinson, 1989). Therefore, poverty is not measured only as a shortage of resources, but of the usual goods and activities in a given society and time.

Nevertheless, this methodology is not free of drawbacks. These problems come from the multidimensionality of data and non-monetary variables and they are related to indicators aggregation as well as the difficulty of combining or substituting indirect indicators by the direct ones.

In this analysis, deprivation means to have denied the opportunity to have or do something through an inability to obtain the goods, activities, and opportunities to participate identified as generally appropriate in the community in question.

1.1 The construction of deprivation indicators

We need to fulfil some steps before building these indicators. These steps are to choose a set of indicators, to evaluate the household situation for each indicator, to define a weighting structure, to aggregate the indicators and, finally, to determine a threshold that divide the deprived population from the non-deprived.

1.1.1 Choosing indicators

This selection depends on the research goals. If we try to analyse the general standard of living, we needed to take into account more indicators.

In any case, it is not easy to determine what and how many indicators we should have taken into account for deprivation measuring. This selection comes from a trade-off between the possible redundancy caused by overlapping information and the risk of obviating some important variables.

Furthermore, there are two different approaches in deprivation research: on the one hand, those authors who seek the intrinsic elements of poverty and, on the other, the authors that consider a most complex and complete (related to welfare) idea. The latter consider some aspects as health, activity status, educational level, social integration, and leisure... topics more related to social exclusion than to poverty or deprivation.

Once the previous issue is fixed, a new dichotomy appears. We must choose between a needs-restricted study as Mack and Lansley (1985) and a research with a larger set of indicators related to standard of living as Halleröd (1994). In the first case, information on non-necessary goods is not considered. However, the researcher must face an issue: how to distinguish if a good is necessary or not? Mack and Lansley (1985) propose a consensual method to avoid arbitrariness and value judgements. They call "necessary goods" those goods considered as necessary by the society. In their work, a good was qualified as necessary if an half of the population considered it as necessary. Nevertheless, the definition of the concept of need is the great drawback of this approach.

The second approach, "life style" approach, avoids the distinction between needs and non-needs considering more variables. In this case, indicators are more related to standard of living than to deprivation. Namely, poverty or deprivation are considered as a low standard of living. Nevertheless, the main risk in indicators selection is arbitrariness. For example, Townsend (1979) started with 60 indicators and, afterwards, it selected twelve.

1.1.2 Evaluation

In most of the empirical studies, indicators are binary variables that express the possession of a given good or the participation in a given activity. With dichotomized indicators, the situation of a household or individual for each one of them can be evaluated according to the following function $z(x_{ij})$, where $z(x_{ij})$ is the amount of j good or activity owned or accomplished by the i household.

$$z(x_{ij}) = \begin{cases} 1 & \text{if } x_{ij} < x_j \quad \rightarrow \textbf{deprived} \\ 0 & \text{if } x_{ij} \geq x_j \quad \rightarrow \textbf{non deprived} \end{cases} \qquad [1]$$

where x_{ij} is the "social norm" or the more common quantity or value in the society.

A problem that these variables present is that they only inform on the presence of the good or the activity. There is no information about quantity or quality. To solve it, Desai and Shah (1988) generalize the function of the expression (1), considering a

distance or disparity function respect of the modal value of the variable *j*. Nevertheless, as Martínez and Ruiz-Huerta (1999) say, since the aim is to detect deprivation situations and not a complete description of welfare, this issue is not so important.

Another problem is related to the relationship between absence and deprivation. Preference structures and life styles can affect the consideration of a good as necessary and its acquisition given the available resources. For instance, how to qualify a household or individual that it does not possess a good that is considered as necessary by most of the population because they have decided not to have it? To solve this problem, Mack and Lansley (1985) define that deprivation is caused by an enforced inability to possess or accomplish the good or activity.

According to this definition, a household or individual that do not have a good or an activity is considered deprived only if it can afford them.

However, the former definition only can be used when the required information is collected. Although this information was available for each indicator, a new problem appears: the reliability of households or individuals when they assert that the absence of a good is due to a lack of resources. Piachaud (1981, 1987) has exposed this topic when he criticized Townsend (1979) and Mack and Lansley (1985).

It can be possible that a household say that it cannot afford to satisfy a necessity and, simultaneously, it can get some non-necessary goods. Furthermore, the reduction of expectations caused by poverty or deprivation persistence makes possible to find deprived households or individuals that argue not to need these basic goods those they lack.

We think that a combined analysis of objective and subjective lacks can describe the deprivation situation better. Other authors have opted for an alternative methodology[2]: fuzzy sets. In this case, different degrees of deprivation are assumed instead of a dichotomy between poor and non-poor. Consequently, the extreme values imply a deprivation situation or absence of deprivation and the other values in the interval (0, 1) express a partial deprivation.

About this methodology, we consider that our aim, the better identification of the poor population, is achieved better with a clear differentiation between the deprivation and its absence. If the identification is the first step for reducing poverty, it is important to know who must be the receiver of these policies.

1.1.3 Weighting indicators

Before aggregating indicators, it is necessary to establish a weighting structure for each one given their different features. For instance, are so important "to have arrears in the mortgage payment", "to possess a microwave" and "to have light problems in the housing"? 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), Mack and Lansley (1985) or Mayer (1989). This weighting structure can be justified, on the one hand, by reducing the researcher's interferences on the results and, on the other, for lack of information on the consideration as "necessary" of the goods or activities. However, the absence of discrimination between some components with a clearly different importance in deprivation measuring is an important problem.

Alternatively, we can compute the weightings from data. One of the possible strategies consists of 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) gives more importance to the absence of goods considered as necessary by larger groups of the population and Desai and Shah (1988), in their deprivation index, give a higher weight to the goods that are most widely owned in a society.

The former, consensual methodology, besides of having the advantage of being closer to social views on the meaning of an adequate minimum standard of living, is more stable since the social perception of needs change slowly. Otherwise, the information required to know which goods are necessities is not always available.

Other studies, where European Community Household Panel (in forward, ECHP) micro data are used, apply other weighting structures since this database does not collect the social views on the necessity of goods or activities. Martinez and Ruiz-Huerta (1999, 2000) weight each attribute by the ratio between the proportion of people who has the good *j* and the total of proportions for each indicator. On the other side, Whelan et al. (2001a, b) and Muffels and Fouarge (2001) weight each attribute by the proportion of households that own the good. The latter justify their election with Runciman's (1966) definition of deprivation. According to this definition, the better a person see the others, the poorer he or she feels.

The importance of each indicator can be also computed by means of different multivariate statistical methods, as factorial analysis as Nolan and Whelan (1996) or Layte et al. (1999, 2000), principal components analysis, Ram (1982) and Maasoumi and Nickelsburg (1988), or cluster analysis (Hirschberg et al., 1991). Moreover, a last methodology is to use market prices as weights. Nevertheless, prices are not available for each attribute and it can be difficult to estimate them.

1.1.4 Aggregating indicators

Once previous issues are done, the researcher faces the most important decision: how to work with the multidimensionality of poverty or deprivation. The greater the structure we impose on data, the closer we arrive at a complete cardinal measure. In Brandolini and D'Alessio (2000), the main strategies depending on the degree and method of aggregation are shown.

(a) Item-by-item analysis: supplementation strategy.

(b) Comprehensive analysis

i. Non-aggregative strategies

- Vector dominance

- Sequential dominance

- Multivariate techniques

- Multidimensional inequality indices.

ii. Aggregative strategies

- Well-being indicator

- Equivalence scales

Therefore, the possible methods go from the supplementation strategy to the computation of a synthetic welfare indicator. The former consists in considering all the indicators one by one, by studying their univariate characteristics and their correlation structure, with some information on income distribution. Its simplicity, an advantage, causes a great drawback if there is much information on households or individuals: this method does not summarize it and so, a good description cannot be done.

The alternative is to consider jointly the indicators, to aggregate them and to obtain a summary measure or some measures. Among the possible strategies, we emphasize the use of the following ones:

- Multivariate statistical techniques.

- Multidimensional poverty indexes, developed by Bourguignon and Chavrakarty (1999) as a valuation function of the attributes. This method is, practically, equivalent to the next strategy.

- Construction of a welfare indicator, indicator that it can be measured in monetary units or in another unit of "welfare". While, for the last option, we can use the multivariate statistical analysis to build it, we can adjust income to attribute values with equivalence scales.

There is a trade-off between synthesis and the best description. This issue has not defined yet in the literature. Although, on one hand, joining all the attributes in an index offers the advantage of summarizing the complexity in a simple way, such 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) exposes a reason to choose the non-aggregative alternative.

Nolan and Whelan (1996), Layte et al. (1999, 2000), Martínez and Ruiz-Huerta (1999, 2000) and Whelan et al. (2001a, b) consider different dimensions in poverty or

deprivation analysis, corresponding each one of them to different aspects as basic needs, secondary needs or housing conditions.

1.1.5 Threshold definition

This step is related to the aim of any poverty or deprivation analysis: the identification of poor population. Three approaches can be followed to achieve this goal:

- To establish an income threshold, for whose construction the information on the standard of living is used. Poverty line is the income value below which deprivation increases markedly. An example of this approach is Townsend's study (1979), based in a close relationship between standard of living and income. If such hypothesis is rejected, it is difficult to find a clear poverty line.

- To identify population with living conditions indicators. It is necessary, then, to establish a value for a deprivation index that divide to the population in two groups. However, this task it is not free of problems. For example, Mack and Lansley (1985) proposed two conditions to determine the threshold (poor population also lacks some non-necessary goods and usually its income is low) and Muffels and Fouarge (2001) opted for the weighted average of deprivation index.

- To identify poor population by means of a combination of monetary income and standard of living criteria. This method is based on Ringen's (1987) criticisms to the hypothesis of a strong association between monetary income and standard of living for the smallest values of both variables. As Martínez and Ruiz-Huerta (2000) exposed, this method has been applied in Halleröd (1994) and Nolan and Whelan (1996) the studies to identify the “real poor” and the “consistent poor”, respectively.

2. Latent Class Analysis

Latent class models were introduced by Lazarsfeld (1950) and Lazarsfeld and Henry (1968). Besides, Anderson (1954) and McHugh (1956) have been studied estimation and identification problems. Goodman (1974) connected these models with contingency tables theory and finally, we can present some authors who have developed

these techniques as Agresti (1990), Andersen (1993), Bartholomew (1987), Clogg (1993) or McCutcheon (1987).

Dependence relations between categorical variables in a contingency table are often caused by an underlying association between them and another variable that is not directly observed and it is called latent variable.

The latent class model is a statistical technique that allows to study the existence of one (or more) latent variable from a set of explanative and observed variables and to define, from the classes, a typology of analysed households. In latent class models, both observed and latent variables are categorical with two or more categories, so that the relation between indicators must fulfil two *a priori* hypotheses:

- Symmetrical relation: each observed variable can explain and be explained by the behaviour of any other categorical variable in the table.

- Local independence: observed variables are statistically independent given a category of the latent variables. That is, observed variables are conditionally independent given a class of latent variable.

The latent class model can be parameterised in two different ways: by conditional probabilities or a log-linear model.

Let a set of categorical variables, A, B, C and D, with a number of categories A, B, C and D respectively.

Therefore, we have a contingency table with $AxBxCxD$ dimension.

Besides, let X a latent variable with X classes. The basic equations of latent class model are:

$$\pi_{ijkl} = \sum_{t=1}^{T} \pi_{ijklt} \qquad [2]$$

where

$$\pi_{ijklt} = \pi_t \pi_{ijkl|t} = \pi_t \pi_{i|t} \pi_{j|t} \pi_{k|t} \pi_{l|t} \qquad [3]$$

Symmetrical relation hypothesis is fulfilled because every observed variable only depends on latent variable and, besides, observed variables are statistically independent given every latent class (local independence hypothesis).

Here, π_{xabcd} is the probability for (x,a,b,c,d) cell in the joint distribution $XABCD$. Furthermore, π_x is the probability of belonging to latent class x and $\pi_{abcd|x}$ is the probability of have a combination of observed variables given $X=x$. The rest of parameters are conditional probabilities.

Therefore, the parameters of latent class model are the conditional probabilities $\pi_{a|x}$, $\pi_{b|x}$, $\pi_{c|x}$, $\pi_{d|x}$ and the latent class probabilities π_x, under the following restrictions:

$$\sum_{a=1}^{A}\pi_{a|x}=\sum_{b=1}^{B}\pi_{b|x}=\sum_{c=1}^{C}\pi_{c|x}=\sum_{d=1}^{D}\pi_{d|x}=1, \qquad [4]$$

and $\sum_{x=1}^{X}\pi_x=1$.

2.1 Estimation procedure

As some latent variable is considered in the model, we can speak about complete and incomplete data. The former are related to the frequency distribution of all the variables (observed and non-observed) and the latter are the frequencies of observed variables.

The EM algorithm is an iterative procedure and each iteration consists in two steps. In the Expectation step, all the expected values are computed given the observed values and the "current" model parameters. In the Maximization step, the likelihood function for all the data is maximized by using the expected data computed in the step before. It involves the estimation of the model parameters as there are no missing data, that is, estimates are used as observed frequencies. For doing this, the same procedures for getting maximum likelihood estimations for a usual log-linear model are used: Newton-Raphson or Iterative Proportional Fitting algorithms. The obtained estimates are used in a new Expectation step to get new estimates for complete table frequencies. Iterations continue until convergence is reached.

The complete data n_{xabcd} follow a multinomial distribution $M(N,\pi_{xabcd})$. Therefore, the logarithm of the likelihood function is

$$\sum_{xabcd} n_{xabcd} \log \pi_{xabcd} \qquad [5]$$

It is possible to get sufficient statistics for parameters if, firstly, we decompose the probability π_{xabcd} by means of equations [2] and [3].

$$\sum_{xabcd} n_{xabcd} \log \pi_{xabcd} = \sum_{x} n_{x\cdots} \log \pi_{x} + \sum_{x,a} n_{xa\cdots} \log \pi_{a|x} + \sum_{x,b} n_{x\cdot b\cdot\cdot} \log \pi_{b|x} + \\ + \sum_{x,c} n_{x\cdot\cdot c\cdot} \log \pi_{c|x} + \sum_{x,d} n_{x\cdots d} \log \pi_{d|x} \qquad [6]$$

In the *E* step, the expected values for complete data are computed given the observed data and the parameters from the former iterations. Since the complete data follow a multinomial distribution, the conditional distribution of such complete data n_{xabcd} given the observed data n_{abcd} is a multinomial distribution $M(n_{abcd}, \pi_{xabcd}|\pi_{abcd})$, and, therefore,

$$E\left[n_{xabcd} \mid n_{abcd}, \pi_{xabcd}\right] = n_{abcd} \frac{\pi_{xabcd}}{\pi_{abcd}} = n_{abcd} \pi_{x|abcd} \qquad [7]$$

Thus, in the *E* step we have

$$E\left[n_{x\cdots} \mid n_{abcd}, \hat{\pi}_{xabcd}(p)\right] = \sum_{a,b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)$$

$$E\left[n_{xa\cdots} \mid n_{abcd}, \hat{\pi}_{xabcd}(p)\right] = \sum_{b,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)$$

$$E\left[n_{x\cdot b\cdot\cdot} \mid n_{abcd}, \hat{\pi}_{xabcd}(p)\right] = \sum_{a,c,d} n_{abcd} \hat{\pi}_{x|abcd}(p)$$

$$E\left[n_{x\cdot\cdot c\cdot} \mid n_{abcd}, \hat{\pi}_{xabcd}(p)\right] = \sum_{a,b,d} n_{abcd} \hat{\pi}_{x|abcd}(p)$$

$$E\left[n_{x\cdots d} \mid n_{abcd}, \hat{\pi}_{xabcd}(p)\right] = \sum_{a,b,c} n_{abcd} \hat{\pi}_{x|abcd}(p)$$

In the *M* step and given the equation [7], estimates for parameters in the p+1-th iteration are computed given the sufficient statistics of complete data in the p-th iteration.

$$E\left[n_{x\cdots} \mid \hat{\pi}(p+1)\right] = N\hat{\pi}_x(p+1)$$

$$E\left[n_{xa\cdots} \mid \hat{\pi}(p+1)\right] = N\hat{\pi}_x(p+1)\hat{\pi}_{a|x}(p+1)$$

$$E\left[n_{x\cdot b\cdot\cdot} \mid \hat{\pi}(p+1)\right] = N\hat{\pi}_x(p+1)\hat{\pi}_{b|x}(p+1)$$

$$E\left[n_{x\cdot\cdot c\cdot} \mid \hat{\pi}(p+1)\right] = N\hat{\pi}_x(p+1)\hat{\pi}_{c|x}(p+1)$$

$$E\left[n_{x\cdot\cdot\cdot d} \mid \hat{\pi}(p+1)\right] = N\hat{\pi}_x(p+1)\hat{\pi}_{d|x}(p+1)$$

From these expressions, the estimates of the probabilities are

$$\hat{\pi}_x(p+1) = \frac{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{N}, \qquad [8.a]$$

$$\hat{\pi}_{a|x}(p+1) = \frac{\sum_{b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}, \qquad [8.b]$$

$$\hat{\pi}_{b|x}(p+1) = \frac{\sum_{a,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \qquad [8.c]$$

$$\hat{\pi}_{c|x}(p+1) = \frac{\sum_{a,b,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \qquad [8.d]$$

$$\hat{\pi}_{d|x}(p+1) = \frac{\sum_{a,b,c} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \qquad [8.e]$$

The iteration procedure must continue until the growth of the logarithm of the likelihood function was less than a very low value, for instance 10^{-6}. Even if the iterations are repeated many times, it is possible to find a local optimum.

From the equations [8], it is possible to compute the probabilities

$$\hat{\pi}_{xabcd} \text{ and } \hat{\pi}_{abcd} = \sum_{x=1}^{X^*} \hat{\pi}_{xabcd}. \qquad [9]$$

3. A Study on Deprivation from ECHP Data

In this section, a deprivation analysis in Spain by using micro data of the European Community Households Panel. Along the section, different problems are commented on database, indicators and methodology used in this study as well as the results of the application of such methodology.

3.1 Database

The data we used in this study belong to the last available wave of ECHP (2000). Exactly, we have chosen an extended sample for Spain and deprivation is measured at household level and so, our sample is composed by 14614 households.

This database is a longitudinal survey begun in 1994 for every member country of the European Union. The objective pursued by EUROSTAT when this panel was created was the comparability of data and results between different countries. To achieve such comparability, survey questions, data collecting, codification and weighting structure were harmonized.

Its great advantage is its temporal feature. Since this panel is done along the time it is possible to observe, for example, the effects produced by income mobility or impoverishment processes. Furthermore, as the same sample units are followed along the waves, researchers can determined followed paths (Hills, 1998a, b) or persistence in states as Stevens (1994, 1999), Cantó (1998, 2000, 2002), Fouarge and Muffels (2000) or Devicenti (2001).

Furthermore, the database has been designed to collect detailed information on income of each household member as well as other important aspects related to material and demographic household features. This is the reason why it that will be preferable to the Household Expenditure Survey to do studies on deprivation or non-monetary poverty. This panel includes some useful variables to analyse poverty and even social exclusion.

In spite of the advantages before exposed, this database presents some drawbacks. No information on household expenditure is collected and so, description done by means

of income and living conditions cannot be improved. For instance, if consumption patterns were known, influence from preferences structure would be eliminated on some questions on financial situation.

Also, information on financial situation and living conditions only is referred to the capacity of purchasing or accomplishing, respectively, a good or an activity and it does not measure how many times is purchased or accomplished.

3.2 Building Deprivation Indicators

In the previous section, we have exposed that an advantage of ECHP is the inclusion of some variables related to household situation that allow improving the information provided by income. Among them, we have the ability of satisfying a set of needs or to purchase some goods, the arrears in some payments as mortgage or rent and housing conditions. We think that we need to do some comments before exposing the methodology we use for getting groups of households according to their deprivation level.

- In order to avoid effects produced by arbitrariness in choosing indicators, we will be used a criterion derived from multivariate statistical methods (latent class analysis). Thus, those attributes that divide the population in homogeneous groups are considered and if a variable seems to show a similar distribution in the subgroups according to standard of living or deprivation, it is eliminated of the indicator set.

- Following Martínez and Ruiz-Huerta (1999, 2000), some aspects as health, social relations or employment are not taken into account. They are excluded due to the consideration of poverty or deprivation as concepts related to standard of living and resources and the before cited topics are more nearer to “social exclusion”.

- We follow “enforced lack” criteria (Mack and Lansley, 1985) to determine deprivation in each variable. Consequently, deprivation in a variable is only considered if the absence of this attribute is due to lack of resources. This information is only collected in ECHP for durable goods owning and for the ability of doing some activities. Either it is possible to use a criterion "consensual", as Halleröd proposes, since information on the social view as necessity of a good or activity are not considered[3].

- We have considered different dimensions of deprivation as housing conditions, basic needs or durable goods.

- Each variable are codified as one (deprivation) and two (no deprivation).

Once these problems are explained, it is possible to show the methodology we used in this work. The intended goal is the identification of different groups in Spanish population according to their deprivation level. To achieve this identification and to summarize the collected information by the selected indicators we use a multivariate statistical method, the latent class analysis.

This technique is chosen because it is the most adequate for the pursued objective (to find homogeneous groups in the population with regard to an unobservable variable) and the type of indicators (categorical).

To select indicators, we started from a set of 33 questions related to financial situation, housing conditions and durable goods owing. This author decided to done firstly a partial latent analysis and, once latent groups for each dimension of deprivation are determined, we estimate a general latent variable that it would correspond with a theoretical concept of "general deprivation". That is, a two-stage process is followed in deprivation identification.

3.3 Different dimensions of deprivation

Some authors as Layte et al. (1999) or Whelan et al. (2001a, b) take into account household financial situation and durable goods possession, calling them "basic needs" and "secondary needs", respectively. Furthermore, they differentiate in housing conditions between, on the one hand, environment quality (pollution, noise, vandalism or crime) and, on the other, accommodation quality (inadequate light or space, leaking roof, dampness and rotting in windows frames and floors and housing facilities). However, a study we have done before shown that environment features do not seem to discriminate between households in Spain. Consequently, we have not considered such variables in this analysis.

On the other hand, Martinez and Ruiz-Huerta (1999, 2000) built an additional dimension related to life style from some variables related to financial situation and durable goods possession.

In this analysis, a previous exploratory study showed that variables concerning to deprivation could be grouped in three dimensions: basic needs, housing conditions and secondary needs or life style.

Once the consideration of these three aspects is decided, the variables included in each dimension are shown. We have selected them after testing their ability to discriminate between different deprivation situations.

- "Basic needs": include not to afford an adequate heating, buying new clothes, eating meal every second day, having friends or family for drink/dinner, to have arrears in ordinary payments[4] and to possess a car and telephone. With regard to the latter two variables, a household is assumed to be deprived if he or she cannot afford the possession of these goods.

- "Housing conditions" Among them, we consider the lack of separate kitchen, bath or shower, the presence of indoor flushing toilet, the lack of running water, the shortage of space and the absence of leaks or dampness. These variables only express the absence or presence of such features, not the ability of avoiding them.

- "Secondary or life style needs": Among the considered variables, there are not to afford paying for holiday, replacing worn-out furniture and to own colour TV, VCR, microwave and dishwasher.

3.3.1 Basic deprivation

Firstly, we must reject the hypothesis of independence following the quality of fit (Table 1). That is, it is possible to divide the population in some groups.

[Table 1 placed here]

At first sight, almost all the models should be rejected because of its probability. Only four and five class models could be accepted against the saturated one.

However, this is a typical case of "large sample size rejection". The sample size is very large, 15614 households, and we look for the most parsimonious model with the best fit. Hagenaars (1990: 56 – 68) provided some guidelines to choose the best model. The main role in selecting model should be the theory. Besides it, the parsimony principle must guide the selection procedure. Thus, *ceteris paribus*, it is better a model with fewer parameters (less complex) than one with more parameters (more complex). According to Hagenaars, it is possible to compute the "large sample size" effect testing the fit for relative frequencies better than absolute ones. In our case, the values of L^2 and χ^2 statistics show a large reduction and, consequently, it is right to assume that the sample size causes the rejection. Finally, in a latent class context, since once the classes are identified, each household is assigned to the latent class for which the conditional probability of belonging given the observed pattern is largest, it is important to take into account the probability of a misclassification for all the households.

$$E = \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} \sum_{d=1}^{D} \pi_{abcd} \, \varepsilon_{abcd} \qquad [10]$$

where ε_{abcd} is the household probability of misclassification.

Once these criteria are applied to the obtained results, we find three groups in the population according to their deprivation level.

Only one variable (Hf010x, "arrears in payments") seems not to differentiate between groups.

The results show (table 2) a low of deprived people is low because of its definition. The identification of low living conditions, more than the income levels, is the main goal of this paper. They can only hardly afford eating meat or fish every second day and to possess a telephone or a car.

[Table 2]

In the other extreme, we find a large group of households that can fulfil all the needs. Better then "rich" or "high life style" people, they should be called "low deprived people". The capability of satisfying the needs is the only issue to measure instead of the degree of their fulfilment. For instance, the question about new clothes only expresses the capacity to buy them, not their price, number or quality.

Finally, there an intermediate group of people whose difference with the group before is the ability of affording their home adequately warm. Therefore, it is possible to affirm that the basic deprivation level is low in Spain. However, it is important to remark again that deprivation is measured and not welfare or wellbeing.

3.3.2 Housing

The same comments on model selection argued before can be applied to this dimension.

The hypothesis of independence of variables is again rejected and so three groups of households are identified in the Spanish population.

[Table 3]

As the former dimension, the most likely situation of Spanish people is good housing quality. This conclusion agrees with the previous studies about deprivation in Spain.

[Table 4]

The households of class 3 have not deprivation in any indicator. They live in a household with a separate kitchen, a bath or a shower, an indoor flushing toilet as well as running water. Moreover, it is a dwelling without dampness, that is, there are no leaks, dampness or rottenness in wooden windows or floors. Finally, they report not to have lack of light or shortage of space.

The intermediate group lives in households whose main problem is the shortage of space. Thus, they have all the housing facilities and the dwelling where they live are free of dampness and darkness.

Finally, as we have exposed before, the class formed by the most deprived people is very small, around a 1%. They live in households that, except for a separate kitchen and an indoor flushing toilet, have not the rest of facilities. Even more, the probability of non-deprivation in an indoor flushing toilet is almost equal to the opposite. Finally, these households live in homes with leaks, dampness or rottenness in wooden windows or floors. An apparently absurd result is related to the shortage of space. It is important to note that the unit of analysis is the household and these variables are related to their households. Other studies have shown that old people living alone and lone mothers are more likely to be deprived.

3.3.3 Secondary deprivation

The analysis of this dimension reveals again the rejection of independence hypothesis. In addition, we have the same problem: the effect of sample size on model selection.

[Table 5]

The main feature of this model is the size of the latent classes because it is not so different from in the dimensions above. This fact appears because this deprivation dimension is not related to basic needs or maintenance, but to issues related to life style as being able to afford paying for holidays or having dishwasher.

[Table 6]

The smaller group is the most deprived. Except for the affordability of a colour TV, they cannot face the rest of needs.

On the other side, a half of the population belongs to the class with smaller deprivation, since can afford all the needs and buying all the goods. That is, they have them or choose not to have them.

Between them, two classes show different kinds of deprivation. The first and larger is related to needs and the other to durable goods. Thus, it is not possible to order these categories, only to express an intermediate situation between "totally deprived" and "totally non-deprived".

3.4 Overall Deprivation

Once different dimensions are analysed, the following step is combine them and to identify different groups in population for this overall definition. Thus, in this second step we have three variables: basic deprivation, housing conditions and secondary deprivation with four, three and four categories respectively, since these models were selected in each dimension in the first step of the study.

Again, we look for the existence of subgroups in the population, not a priori established, that have homogeneous features. Besides, these groups would be mutually differentiated.

The analysis of the next table (table 7) shows that there is some relationship between the variables, since the independence hypothesis is rejected. Following the same criteria assumed in the partial analyses (correcting the effect of sample size and choosing a low misclassification error), we should select the model that considers two classes in the population for deprivation: deprived and non-deprived households.

[Table 7]

The classes, as we could expect, expose the same situation shown in the partial studies (table 8). Relationship between partial and overall categories is stronger for basic deprivation than secondary. This fact is caused by differences in membership proportions for each dimension.

[Table 8]

The same reason, membership proportions, causes that conditional probabilities in housing dimension are higher for “low deprivation” category in both classes of aggregate deprivation. We must remind the expected high quality of Spanish households.

Regarding the basic dimension, the two groups with a higher deprivation level are more related to the “aggregate” deprived class. Besides, even if a household lays in the low basic deprivation category, it can be assigned to the high aggregation deprivation class due to its values in the other dimensions. Therefore, basic deprivation seems to be a concept very similar to minimum standard of living.

Finally, with regards to secondary deprivation, conditional probabilities given each aggregate class shows that, while deprived households are expected to present a high level of secondary deprivation or, at least, not to pay holidays or replacing furniture, non deprived ones can afford all the commodities and activities.

After assigning each response pattern to a latent class by a Bayesian procedure, the main feature of deprived households is the high levels of deprivation in two of the dimensions, basic and secondary. On the other side, housing conditions does not discriminate between the different classes of overall deprivation. If a household is in the most deprived category for basic or secondary dimensions, it is expected to belong to the deprived aggregate class. Even more, the only households with problems to assign (because the Bayesian probabilities are very similar for both aggregate classes) are those very deprived in a dimension and non-deprived in the other.

4. Conclusions

We have shown that latent class analysis is a useful tool for classifying the households by their deprivation level. This, we overcome the issues derived from using an indirect and multidimensional indicator, income, to measure a multidimensional phenomenon, deprivation.

We include a set of direct indicators on living conditions. Besides, considering deprivation as a categorical variable avoids threshold identification problem.

Different dimensions in deprivation have been taken into account: basic needs, secondary needs and housing conditions. Basic deprivation refers to ability for keeping the home adequately warm, buying new clothes, eating meal every second day, having friends or family for drink/dinner, having a car o a telephone and not to have arrears in payments.

The results for 2000 shows that basic needs can be satisfied by the most of households, since only a small proportion of households suffer a situation where they can afford eating meal every second day and having a telephone.

This fact appears again in housing deprivation where only a small proportion of households belong to the "most deprived" category. That is, a large proportion of households live in an accommodation without problems. Despite of this apparently shocking result, we have to recall the kind of households that have been sampled in this panel. Therefore, homeless households are less represented in the sample.

Finally, secondary deprivation is related to life style and, therefore, the proportions are more similar for each category (except the residual group that cannot afford some special appliances) than in other dimensions. Among durable goods, the most deprived category only can afford a colour TV.

Once each deprivation dimension has been studied, we combine them. We found two clusters: deprived and non-deprived households. Basic and secondary deprivations are the most important variables to decide the membership to an aggregate deprivation category. We can conclude that the proposed model could be an adequate procedure for identifying deprived households from the comparison of these results with some from alternative methodologies.

Acknowledgements

This paper is based on analysis of the European Community Household Panel survey for Spain in 2000. The data are used with the permission of the Spanish Statistical Institute, who bears no responsibility for the analysis or interpretations presented here. The research was carried out as a part of the work of a Research Project (2PR02A102) funded by the regional government of Extremadura (Spain) and started visiting CEPS/INSTEAD (Differdange-Luxembourg) in 2000 as an IRISS visitor.

References

Agresti, A. (1990) *Categorical data analysis,* John Wiley, New York.

Anderson, T.W. (1954) On estimation of parameters in latent structure analysis. *Psychometrika* **19**, 1-10.

Atkinson, A.B. (1989) *Poverty and social security*, Harvester Wheatsheaf, London

Ayala, L., Martinez, R. and Ruiz-Huerta, J. (2002) Institutional determinants of the unemployment-earnings inequality trade-off. *Applied Economics* **34**, 179-195.

Bartholomew, D.J. (1987) *Latent variables models and factor analysis*, Griffin, London

Brandolini, A. and D'Alessio, G. (2000) Measuring well-being in the functionings space. 26ª IARIW General Conference. Krakow.

Bourguignon, F. and Chavrakarty, S.R. (1999) The measurement of multidimensional poverty in *Advances in econometrics, income distribution and scientific methodology*, D.J. Slottje, Physica, Heidelberg.

Cantó, O. (1998) *Income mobility in Spain: How much is there?* Working paper FEDEA EEE 17, Madrid.

Cantó, O. (2000a) Income mobility in Spain: How much is there? *Review of Income and Wealth*, **46**(1), 85-101.

Cantó, O. (2002) Climbing out of poverty, falling back in: low incomes' stability in Spain. *Applied Economics*, **34**, 1903-1916.

Clogg, C.C. (1993) Latent class models: recent developments and prospects for the future. In *Handbook of statistical modelling in the social sciences,* (Eds.) G. Arminger, C.C. Clogg and M.E. Sobel, Plenum, New York.

Cohen, R. (1998) An analysis of the dynamic behaviour of earnings distributions. *Applied Economics*, **30**, 1-17.

D'Ambrosio, C. and Peragine, V. (2001) *Measuring social exclusion. mimeo*. Università Bocconi. Milan (Italy)

Dempster, A.P., Laird, N.M., Rubin, D.B. (1977) Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* **39**, 1-38.

Desai, M. and Shah, A. (1988) An econometric approach to the measurement of poverty, *Oxford Economic Papers*, **40**(3), 505-522.

Devicenti, F. (2001) *Poverty persistence in Britain: a multivariate analysis using the BHPS, 1991-1997*. ISER Working Paper 2001-02, Colchester, University of Essex.

Enders, W. and Hoover, G. (2003) The effect of robust growth on poverty: a nonlinear analysis. *Applied Economics*, **35**, 1063-1071.

EUROSTAT (2000) *European social statistics. Income, poverty and social exclusion*. Luxembourg.

Fouarge, D. and Muffels, R. (2000) *Persistent poverty in the Netherlands, Germany and the UK. A model-based approach using panel data for the 1990s*. European Panel Analysis Group Working Paper nº 15, Colchester, University of Essex.

Goodman, L.A. (1974) Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215-231.

Halleröd, B. (1994) *A new approach to the direct consensual measurement of poverty*. Social Policy Research Centre Discussion Paper nº 50, New South Wales University.

Hills J., (1998a) Does income mobility mean that we do not need to worry about poverty? in *Exclusion, employment and opportunity* (Eds.) A.B. Atkinson and J. Hills, CASE paper nº 4, CASE – London School of Economics, London.

Hills, J. (1998b) What do we mean by reducing lifetime inequality and increasing mobility? in *Persistent poverty and lifetime inequality: the evidence*. CASE report nº 5, CASE-London School of Economics, London.

Hirschberg, J.G., Maasoumi, E. and Slottje, J. (1991) Cluster analysis for measuring welfare and quality of life across countries, *Journal of Econometrics*, **50**, 131-150.

Layte, R., Maître, B. Nolan, B, and Whelan, C.T. (1999) *Income deprivation and economic strain,* European Panel Analysis Group Working Paper nº 5, University of Essex, Colchester.

Layte, R., Maître, B. Nolan, B, and Whelan, C.T. (2000) *Explaining levels of deprivation in the European Union,* European Panel Analysis Group Working Paper nº 12, University of Essex, Colchester.

Lazarsfeld, P.F. (1950) The logical and mathematical foundation of latent structure analysis in *Measurement and prediction*, (Eds.) S.A. Stouffer, Princeton University Press, Princeton, 362-472.

Lazarsfeld, P.F. and Henry, N.W. (1968) *Latent structure analysis,* Houghton Mifflin, Boston

Maasoumi, E. and Nickelsburg, G. (1988) Multivariate measures of well-being and an analysis of inequality in the Michigan data, *Journal of Business and Economic Statistics*, **6**, 327-334.

Mack and Lansley (1985) *Poor Britain,* Allen and Urwin, London.

Martínez, R. and Ruiz-Huerta, J. (1999) Algunas reflexiones sobre la medición de la pobreza. Una aplicación al caso español in *Dimensiones de la desigualdad. III Simposio sobre igualdad y distribución de la renta y la riqueza* J.M. Maravall (Ed.), Fundación Argentaria and Visor Editorial, Madrid, vol. 1, 367-428.

Martínez, R. and Ruiz-Huerta, J. (2000) Income, multiple deprivation and poverty: an empirical analysis using Spanish data. 26ª IARIW General Conference. Krakow.

Mayer, S.E. and Jencks, C. (1989) Poverty and the distribution of material resources, *Journal of Human Resources,* **21**, 88-113.

McCutcheon, A.L. (1987) *Latent Class Analysis,* Sage, Beverly Hills.

McHugh, R.B. (1956): Efficient estimation and local identification in latent class analysis. *Psychometrika*, **21**, 331-347.

Muffels, R. and Fouarge, D. (2001) Do European welfare regimes matter in explaining social exclusion? Dynamic analyses of the relationship between income poverty and deprivation: a comparative perspective. ESPE Conference. Athens.

Nolan, B. and Whelan, C.T. (1996) *Resources, deprivation and poverty,* Clarendon Press, Oxford.

Pérez-Mayo, J. (2002) *Modelos dinámicos de variables latentes aplicados a la construcción de indicadores económicos y sociales*. Ph. Thesis. *mimeo*.

Piachaud, D. (1981) Peter Townsend and the Holy Grail, *New Society*, **57**, 419-421.

Piachaud, D. (1987) Problems in the definition and measurement of poverty, *Journal of Social Policy*, **16**(2), 147-164.

Ram, R. (1982) Composite indices of physical quality of life, basic needs fulfilment and income. A principal component representation, *Journal of Development Economics*, **11**, 227-247.

Ringen, S. (1988) Direct and Indirect Measures of Poverty, *Journal of Social Policy*, **17**(3), 147-164.

Runciman, W.G. (1966) *Relative deprivation and social justice,* Routledge and Kegan Paul, London.

Sen, A.K. (1987) *The standard of living,* Cambridge University Press, Cambridge.

Stevens, A.H. (1994) Persistence in poverty and welfare: the dynamics of poverty spells: updating Bane and Ellwood. *American Economic Review (Papers and Proceedings)*, **84**, 34-37.

Stevens, A.H. (1999) Climbing out of poverty, falling back in: measuring the persistence of poverty over multiple spells, *Journal of Human Resources*, **3**, 557-588.

Townsend, P. (1979) *Poverty in the United Kingdom,* Penguin Books, Harmondsworth.

Whelan, C.T., Layte, R. and Maître, B. (2001a) *What is the scale of multiple deprivation in the European Union?* European Panel Analysis Group Working Paper nº 19, University of Essex, Colchester.

Whelan, C.T., Layte, R. and Maître, B. (2001b) *Persistent deprivation in the European Union*, European Panel Analysis Group Working Paper nº 23, University of Essex, Colchester.

Table 1. Latent models for basic deprivation

Model	L^2	*Prob.*	*df*	*E*
Independence	5402.6175	0.0000	120	0.0000
Two classes	864.7468	0.0000	112	0.0425
Three classes	204.4839	0.0000	104	0.1127
Four classes	124.3822	0.0273	96	0.1157
Five classes	112.4434	0.0406	88	0.1356

Source: Author's elaboration.

Table 2. Latent and conditional probabilities for basic deprivation

		Classes		
Variables		1	2	3
Keeping home warm	1	0.8157	0.8805	0.3325
	2	0.1843	0.1195	0.6675
New clothes	1	0.7694	0.1666	0.0093
	2	0.2306	0.8334	0.9907
Eating meal every second day	1	0.4080	0.0001	0.0005
	2	0.5920	0.9999	0.9995
Having friends or family to having drink/dinner	1	0.9117	0.1656	0.0244
	2	0.0883	0.8344	0.9756
Arrears in payment	1	0.1645	0.0893	0.0187
	2	0.8355	0.9107	0.9813
Car	1	0.1796	0.2466	0.0187
	2	0.8204	0.7534	0.9813
Telephone	1	0.0799	0.1039	0.0001
	2	0.9201	0.8961	0.9999
Latent class	probability	0.0445	0.1766	0.7789

Source: Author's elaboration.

Table 3. Latent models for housing deprivation

Model	L^2	*Prob.*	*df*	*E*
Independence	2297.8916	0.0000	120	0.0000
Two classes	831.3006	0.0000	112	0.0509
Three classes	193.1697	0.0000	104	0.1312
Four classes	161.8952	0.0000	96	0.1732
Five classes	142.4487	0.0002	88	0.2729

Source: Author's elaboration.

Table 4. Latent and conditional probabilities for housing deprivation

		Classes		
Variables		1	2	3
Separate kitchen	1	0.2308	0.0246	0.0037
	2	0.7692	0.9754	0.9963
Bath or shower	1	0.5796	0.0001	0.0008
	2	0.4274	0.9999	0.9992
Indoor flushing toilet	1	0.4428	0.0005	0.0001
	2	0.5572	0.9995	0.9999
Running water	1	0.7420	0.0175	0.0064
	2	0.2580	0.9825	0.9936
Shortage of space	1	0.2687	0.4882	0.0563
	2	0.7313	0.5118	0.9437
Lack of light	1	0.2899	0.3524	0.0513
	2	0.7101	0.6476	0.9487
Leaks or dampness	1	0.6568	0.3514	0.1143
	2	0.3432	0.6486	0.8857
Latent class	probability	0.0104	0.1821	0.8076

Source: Author's elaboration.

Table 5. Latent models for secondary deprivation

Model	L^2	*Prob.*	*df*	*E*
Independence	17039.8967	0.0000	57	0.0000
Two classes	5484.0135	0.0000	50	0.0557
Three classes	653.2718	0.0000	43	0.0874
Four classes	115.5849	0.0000	36	0.1048
Five classes	84.5903	0.0000	29	0.1050

Source: Author's elaboration.

Table 6. Latent and conditional probabilities for secondary deprivation

		Classes			
Variables		1	2	3	4
Paying for holiday	1	0.9754	0.2610	0.8532	0.0544
	2	0.0246	0.7290	0.1468	0.9446
Replacing furniture	1	0.9085	0.2341	0.8694	0.1340
	2	0.0915	0.7659	0.1306	0.8660
Colour TV	1	0.0275	0.0094	0.0016	0.0001
	2	0.9725	0.9906	0.9984	0.9999
VCR	1	0.5610	0.2685	0.0266	0.0061
	2	0.4390	0.7315	0.9734	0.9939
Micro wave oven	1	0.8509	0.7226	0.0145	0.0001
	2	0.1491	0.2774	0.9855	0.9999
Dishwasher	1	0.9501	0.8401	0.1813	0.0500
	2	0.0499	0.1599	0.8187	0.9500
Latent class	probability	0.1138	0.0459	0.3275	0.5129

Source: Author's elaboration.

Table 6. Latent models for aggregate deprivation

Model	L^2	*Prob.*	*df*	*E*
Independence	3676.8849	0.0000	28	0.0000
Two classes	64.8944	0.0000	20	0.0630
Three classes	34.5437	0.0006	12	0.1544
Four classes	8.7189	0.0685	4	0.1781

Source: Author's elaboration.

Table 7. Latent and conditional probabilities for aggregate deprivation

		Classes	
Variables		1	2
Basic deprivation	1	0.1934	0.0039
	2	0.5052	0.0153
	3	0.3015	0.9808
Housing conditions	1	0.0380	0.0015
	2	0.1604	0.0584
	3	0.8017	0.9401
Secondary deprivation	1	0.4079	0.0288
	2	0.0414	0.0395
	3	0.5348	0.2800
	4	0.0160	0.6517
Latent class	probability	0.1986	0.8014

Source: Author's elaboration.

Notes

[1] This is the method proposed by Townsend (1979) and Muffels (1993).

[2] Cerioli and Zani (1990), Cheli et al. (1994), Cheli and Lemni (1985) and Lemni (1996).

[3] For instance, when a household is asked about running water there is no difference between a lack caused by not affording it and a decided lack.

[4] A household has arrears in ordinary payments if it has arrears in, at least, one of the following payments: rent for accommodation, mortgage, utility bills and other loan repayments