On measuring deprivation in Spain: comparing some techniques

JESÚS PÉREZ-MAYO

Department of Applied Economy, University of Extremadura, Campus Universitario, 06071 BADAJOZ (SPAIN)

Fax: 34 924 272 509 e-mail address: jperez@unex.es

Abstract

This paper analyses the deprivation in Spain based on ECHP data for 1996. 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 (Townsend 1988, Mayer and Jencks 1988, Muffels 1993, Callan *et al.* 1993, Dirven and Fouarge 1995, Layte *et al.* 1999, Whelan *et al.* 2000). 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 here an inability to get the goods, facilities and opportunities, which are usual in the household environment.

Among the strategies for the application of the capability approach (Brandolini and D'Alessio, 2000), the method we propose in this paper lies in the non-aggregative strategies, exactly, in the multivariate techniques. Since all of the variables needed to build the profiles are categorical, we use the latent class model to solve this problem because it is the best model to do it. Besides, the results are compared with those from another empirical analysis following Martinez y Ruiz-Huerta's (2000) and D'Ambrosio and Peragine's (2001) methodologies, respectively

1. Introduction

According to the European Council (1984), cited 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 citation, 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 going to be measured: how standard of living is measured, which this "minimum standard of living" is, 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). It is defined as the 50 or 60% of the mean or median income, depending on the studies. Although this method presents some advantages, easy computation and comparison between different periods or territories, it also has some drawbacks listed below, following Martínez and Ruiz-Huerta (1999):

a) the length of the reference period.

b) some non - monetary variables need to be included.

c) wealth is not included.

d) it is difficult to evaluate household necessities.

e) 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 last advantage is related to the consumption theory. According to the classical consumption theory, current expenditure is a better approximation of the permanent income than current income. However, expenditure as an indicator 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. 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:

a) They describe better the poor (by income) households. Here, we can speak on living conditions of the poor population.

b) Without leaving the income criteria, these indicators allows to improving 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.

c) Finally, they can be used as an alternative indicator to measure poverty. As Martínez and Ruiz-Huerta (2000) 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 usual goods and activities in a given society and time.

Nevertheless, this methodology is not free of drawbacks. These problems come derived 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 be denied the opportunity to have or do something through an inability to obtain the goods, facilities, 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. Such 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, two different approaches exist 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 poverty or deprivation.

Once the previous issue is fixed, a new dichotomy appears. We must choose between a needs-restricted study (Mack and Lansley, 1985) and a research with a larger set of indicators related to standard of living (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 necessary by 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 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 Household evaluation

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

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

where x_j it 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 (2000) say, since the aim is to detect deprivation situations and not a complete description of welfare, this issue is not so important.

Other 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 that it does not possess a necessary considered good by most of the population because they have decided not to have it? To solve this problem, Mack and Lansley define that deprivation is caused by an enforced disability to possess or accomplish the good or activity. According to this definition, a household 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 when they assert that the absence of a good is due to a lack of resources. Piachaud (1981, 1987) has expressed this issue when he criticized Townsend (1979) and Mack and Lansley (1985).

It can be possible that a household say that it can afford to satisfy a necessity and, simultaneously, it can get some non-necessary goods. Furthermore, the expectations reduction caused by poverty or deprivation persistence makes possible to find deprived households 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 such 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 and Jencks (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 that clearly have a 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 necessary by larger groups of the population or Desai and Shah (1988), when they built their deprivation index, give more a larger weight to the goods that are most widely owned in a society.

The former, consensual methodology, besides of having the advantage of being more closely to social views on what a decent minimum standard of living really means, is more stable since the perception of needs change slowly. Otherwise, the information required to know what goods are necessities is not always available.

Other studies, that used European Community Household Panel (in forward, ECHP) micro data, apply other weighting structures since this database does not collect the social views on the necessity of goods or activities. Martínez 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 and b) as well as Muffels and Fouarge (2001) weight each attribute by proportion of households that own the good. The latter justify their election with Runciman's (1966) definition of deprivation. According to this deprivation, 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 (Nolan and Whelan, 1996; Layte *et al.*, 1999, 2000), principal components analysis (Ram, 1982; Maasoumi and Nickelsburg, 1988; and Maasoumi, 1989) or cluster analysis (Hirschberg *et al.*, 1991).

A last methodology is to use market prices as weights. Nevertheless, prices are not available for each attribute.

1.1.4 Aggregating indicators

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

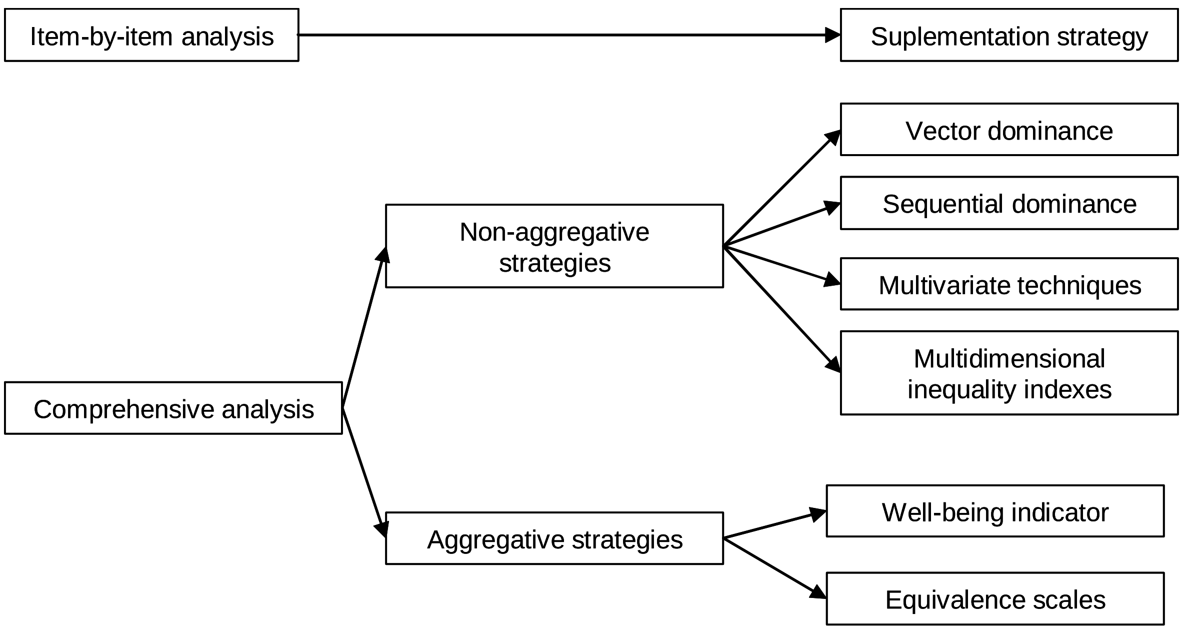

Figure 1. Strategies for measuring deprivation (Brandolini and D'Alessio, 2000)

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 measures or measures. Among the possible strategies, we emphasize the use of

- 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 trade-off between the synthesis and the better 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: 33) 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 and 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 determination

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

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

2) to identify to the poor 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) propose 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) opt for the weighted average of deprivation index.

3) 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 smaller values of both variables. As Martínez and Ruiz- Huerta (2000) say, this method has been applied in Halleröd

(1995) 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 Henry and Lazarsfeld (1968). Besides, Anderson (1954) y 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 (1982), 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.

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 individuals. 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:

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

b) 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.

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 *I*, *J*, *K* and *L* respectively. Therefore, we have a contingency table with *IxJxKxL* dimension. Resides, let *X* a latent variable with *T* 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_{ijklt} is the probability for (i,j,k,l,t) cell in the joint distribution *ABCDX*. Furthermore, π_t is the probability of belonging to latent class t and $\pi_{ijkl|t}$ is the probability of have a combination of observed variables given $X=t$. The rest of parameters are conditional probabilities.

Therefore, the parameters of latent class model are the conditional probabilities $\pi_{i|t}$, $\pi_{j|t}$, $\pi_{k|t}$, $\pi_{l|t}$ and the latent class probabilities π_t, under the following restrictions:

$$\sum_{i=1}^{I}\pi_{i|t} = \sum_{j=1}^{J}\pi_{j|t} = \sum_{k=1}^{K}\pi_{k|t} = \sum_{l=1}^{L}\pi_{l|t} = 1, \qquad [4]$$

and $\sum_{t=1}^{T}\pi_t = 1$.

Maximum-likelihood estimations of latent class models are more complex than totally observed log-linear models. Among the estimation models, Newton-Raphson and EM (Dempster, Laird y Rubin, 1977) algorithms are the most used. We are going to use the latter in this analysis.

This algorithm is an iterative process of estimation with two stages. In the E(xpectation) step all the expected values are computed given observe values and “current” model parameters. In the M(aximisation), likelihood function of the complete table is maximised from expected data in the previous step. It implies the computation of updated estimations of models parameters. Iterations continue until convergence is achieved.

Thus, finally, we can obtain the maximum-likelihood estimations

$$\hat{\pi}_{i|t},\ \hat{\pi}_{j|t},\ \hat{\pi}_{k|t},\ \hat{\pi}_{l|t} \text{ and } \hat{\pi}_t; \qquad [5]$$

from which it is possible to compute the probabilities

$$\hat{\pi}_{ijklt} \text{ and } \hat{\pi}_{ijkl} = \sum_{t=1}^{T}\hat{\pi}_{ijklt}. \qquad [6]$$

The next step of the analysis is to assign every observation to each latent class. Thus, we compute the conditional probability of belonging to class t of latent class given i, j, k and l categories of observed variables A, B, C and D

$$\hat{\pi}_{t|ijkl} = \frac{\hat{\pi}_{ijklt}}{\sum_{t=1}^{T}\hat{\pi}_{ijklt}}. \qquad [7]$$

Given this probability, we assign the cells by modal probability.

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. Afterwards, we analyse the relationship between deprivation and income.

3.1 Database

The data we used in this study came from ECHP for the last wave (1996). This database is a longitudinal survey begun in 1994 for all the member countries 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 in its temporal feature. Since this panel is done along the time it is possible to observe, for example, 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 and 1998b) or persistence in states as studies by Stevens (1994 and 1999), Cantó (1996, 1998, 2000a and 200b), 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 before cited advantages, this database presents some bad points. 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 incorporation of some variables related to the 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 used methodology for getting household groups according to their deprivation level.

a) 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 indicator sets.

b) Following Martínez and Ruiz-Huerta (1999, 2000), some aspects as health, the 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 aspects are more nearer to “social exclusion".

c) We follow “enforced lack” criteria (Mack and Lansley, 1985) to determine deprivation in each variable. Consequently, deprivation in a variable 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].

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

Once these problems are explained, it is possible to show the methodology we used in this work. The intended goal is 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. The consequent contingency table had a large dimension and this dimension made impossible estimate any model. Therefore, the authors decide 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 and b) consider household financial situation and durable goods possession, calling them "basic needs" and "secondary needs", respectively, and, 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 room, leaking roof, dampness and rotting in windows frames and floors and housing facilities). However, the study we have done shows 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, Martínez 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. The latter is a combination of built a dimension more than the deprivation, related it to the life style.

Once the consideration of these three aspects is decided, the variables included in each dimension are shown. We have selected them after proving 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. While the four former variables measure the household ability to face those needs, not their fulfilment, the two latter has been elected as indicator the enforced lack 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 room 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

The analysis of quality of fit of the different models we can consider shows, firstly, that hypothesis of independence of variables must be rejected.

Table 1. Latent models for basic deprivation

Model	L^2	Prob.	G.L.	BIC
Independence	3222.3723	0.0000	120	2173.9935
Two classes	437.6076	0.0000	112	-540.8792
Three classes	146.4309	0.0039	104	-762.1639
Four classes	70.5768	0.9760	96	-768.1262
Five classes	58.4712	0.9936	88	-710.3398

Source: Author's elaboration.

The models with four and five classes should be accepted because the probability of explaining the situation is very high. Finally, we should recall that, with a large sample size, the addition of small differences causes a large difference and, so the model must be rejected or its fit will be worst. Therefore, we recommend using the BIC test in such situation. Following this contrast, we should accept three different groups in the population for basic deprivation.

Before commenting every class, we have to say that the variable concerning payment arrears shows "no arrears" as its modal category. The difference lays in probabilities, since the most deprived class has the biggest probability for that category.

The most deprived group is composed by 3.06% of households. These households only can afford eating meat every second day and to possess a telephone. Furthermore, it is estimated that they are not able of facing the rest of needs.

Table 2. Latent and conditional probabilities for latent basic deprivation

		Classes			
Variables		1	2	3	4
HF003	1	0.0001	0.1690	0.0853	0.6667
	2	0.9999	0.8310	0.9147	0.3333
HF006	1	0.1962	0.5860	0.8202	0.9873
	2	0.8038	0.4140	0.1798	0.0127
HF007	1	0.5708	0.8550	0.9971	0.9999
	2	0.4292	0.1450	0.0029	0.0001
HF008	1	0.0627	0.1974	0.7915	0.9722
	2	0.9373	0.8026	0.2085	0.0278
HF010X	1	0.3739	0.0403	0.1311	0.0146
	2	0.6261	0.9597	0.8689	0.9854
HB001	1	0.4956	0.8538	0.6979	0.9616
	2	0.5044	0.1462	0.3021	0.0384
HB006	1	0.5814	0.9999	0.7831	0.9956
	2	0.4186	0.0001	0.2169	0.0044
Latent class	probability	0.0396	0.0664	0.2414	0.6526

Source: Author's elaboration.

In the other extreme, we have a 66.67% of households, those that can satisfy all the needs. It would be better call them "low deprived households" than "high life style" because we only measure the ability of satisfying needs, not the actual fulfilment or the extent of this satisfaction. For example, the question about new clothes only express the ability of buy them, not their price or quality.

Besides, a 6.64% of the families can afford new clothes (with a high probability for not affordability), eating meal every second day and the possession of durable goods. Finally, a 24.14% of the households can satisfy the most of needs, except affording its home adequately warm. Besides, it differs from the better off before in the conditional probabilities. For this group, the conditional probabilities of inabilities are higher than the ones for the previous class. These groups can be called "light deprived households" due to corresponding to an intermediate situation.

3.3.2 Housing

Independence of variables is rejected and the hypothesis of three different groups is accepted again following the BIC criteria.

Latent probabilities values show that the most frequent situation of Spanish households is good housing quality. This result agrees with the ones from previous studies on deprivation in Spain already cited.

An 82.75% of households, those that belong to the class 3, have not deprivation in any indicator. They have a separate kitchen, bath or shower, indoor flushing toilet as well as running water. Besides, they live in a free dampness-housing, variable that indicates that this accommodation has not leaks, dampness or rottenness in wooden windows or floors.

Table 3. Latent models for housing deprivation

Model	L^2	Prob.	G.L.	BIC
Independence	3092.8912	0.0000	247	934.9784
Two classes	743.5319	0.0000	238	-1335.7526
Three classes	244.1107	0.2351	229	-1756.5454
Four classes	230.2593	0.3039	220	-1691.7684
Five classes	201.3472	0.6717	211	-1642.0521

Source: Author's elaboration.

On the other hand, it is estimated that the class formed by the most deprived household is very small, almost a 2%. Some households that, except for a separate kitchen, have not the rest of facilities. Even more, the probability of lack of a separate kitchen is the higher in this group. Finally, these households live in homes with leaks, dampness or rottenness in wooden windows or floors.

Table 4. Latent and conditional probabilities for housing deprivation

		Classes		
Variable		1	2	3
HA008	1	0.7474	0.9785	0.9919
	2	0.2526	0.0215	0.0081
HA009	1	0.2105	0.9969	0.9983
	2	0.7895	0.0031	0.0017
HA010	1	0.5522	0.9893	0.9985
	2	0.4478	0.0107	0.0015
HA011	1	0.1471	0.9496	0.9916
	2	0.8529	0.0504	0.0084
HA014	1	0.4149	0.3603	0.1848
	2	0.5851	0.6397	0.8152
HA018	1	0.5559	0.5182	0.0265
	2	0.4441	0.4818	0.9735
HA019	1	0.6548	0.8999	0.0645
	2	0.3452	0.1001	0.9355
HA020	1	0.4989	0.2668	0.0175
	2	0.5011	0.7332	0.9825
Latent class	probability	0.0176	0.1549	0.8275

Source: Author's elaboration.

The last group, though a little greater (a 15.49%), is also small. Its deprivation levels are very similar to levels for the first class since they have all the needs, except the free dampness housing.

The shortage of room in the dwelling is not an important feature to differentiate between groups, but it is near a 50% for the most deprived group and, therefore, we can point it as a problem for these households.

3.3.3 Secondary deprivation

The analysis of this dimension reveals the rejection of independence hypothesis again. In addition, according to BIC test, the four-class model must be chosen.

Table 5. Latent models for secondary deprivation

Model	L^2	Prob.	G.L.	BIC
Independence	7311.6850	0.0000	57	6813.7051
Two classes	1648.6065	0.0000	50	1211.7821
Three classes	299.1265	0.0000	43	-76.5425
Four classes	118.3555	0.0000	36	-196.1581
Five classes	105.7257	0.0000	29	-147.6325

Source: Author's elaboration.

The main feature of this model is the similarity of some latent classes. This phenomenon happens 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. Latent and conditional probabilities for secondary deprivation

		Classes			
Variables		1	2	3	4
HF004	1	0.0001	0.5378	0.1712	0.9191
	2	0.9999	0.4622	0.8288	0.0809
HF005	1	0.0001	0.5058	0.0706	0.7977
	2	0.9999	0.4942	0.9294	0.2023
HB002	1	0.9492	0.9762	0.9963	0.9996
	2	0.0508	0.0238	0.0037	0.0004
HB003	1	0.4500	0.5614	0.9392	0.9884
	2	0.5500	0.4386	0.0608	0.0116
HB004	1	0.0995	0.1081	0.9714	0.9931
	2	0.9005	0.8919	0.0286	0.0069
HB005	1	0.1294	0.2274	0.7624	0.9401
	2	0.8706	0.7726	0.2376	0.0599
Latent class	probability	0.1598	0.0732	0.3211	0.4459

Source: Author's elaboration.

The smaller group, formed by a 15.98% of households, is the most deprived. Except for the affordability of a colour TV, they cannot face the rest of needs.

On the other hand, a 44.59% of the families belong to the class with smaller deprivation, since can afford all the needs and buying all the goods. Namely, they have or choose not to have them.

Between them, there are two classes. One of them, composed by a 7.32%, can afford needs concerning holidays and furniture and the possession of some durable goods but microwave and dishwasher. That is, they have a basic set of housework facilities. On the other hand, a 32.11% of households, despite of affording the durable goods, cannot afford the needs.

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: HP001 (basic deprivation), HP002 (housing conditions) and HP003 (secondary deprivation) with three categories each one, since the models with three latent classes were selected in each dimension in the first stage of the study.

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

The analysis of the next table shows that there is some relationship between the variables, since the independence hypothesis is rejected. We only accept one model by the BIC test and this model is the one that considers three classes in the population for deprivation.

Table 7. Latent models for overall deprivation

Model	L2	Prob.	G.L.	BIC
Independence	2578.8434	0.0000	8	2238.1203
Two classes	212.8643	0.0000	30	-49.2304
Three classes	80.2490	0.0000	21	-103.2173
Four classes	32.6383	0.0011	12	-72.1996
Five classes	14.8322	0.0020	3	-11.3773

Source: Author’s elaboration.

The three determined classes, as we could expect, express the situations shown in the partial studies. Relationship between partial and overall categories is stronger for basic deprivation than secondary. This fact is caused by differences in belonging proportions in each dimension. Since they are very similar in secondary deprivation, there is not a clear differentiation in an intermediate degree of overall multidimensional poverty.

The same reason, belonging proportions, causes that conditional probabilities in housing dimension are higher for “low deprivation” category. We have to remind that the probability of belonging of this category is more than an 82.75%.

It is estimated that the first group, formed by 6.64% of the families, has a high basic and secondary deprivation levels, more than a 75% in any case. Furthermore, conditional probability estimates for “low deprivation" is nearly zero. Therefore, we can assert that there is an almost perfect identification between high deprivation levels in the overall variable and the respective partial variables.

Table 8. Latent and conditional probabilities for overall deprivation

		Classes		
Variables		1	2	3
HP001	1	0.4078	0.0455	0.0001
	2	0.0001	0.2237	0.0099
	3	0.5921	0.3673	0.0300
	4	0.0001	0.3635	0.9600
HP002	1	0.1049	0.0211	0.0019
	2	0.3325	0.2292	0.0659
	3	0.5625	0.7497	0.9322
HP003	1	0.9296	0.3353	0.0271
	2	0.0001	0.0897	0.0484
	3	0.0703	0.5207	0.1937
	4	0.0001	0.0543	0.7307
Latent class	probability	0.0664	0.2976	0.6360

Source: Author’s elaboration.

The analysis of the frequencies assigned to this group shows that basic and secondary dimensions are the most important ones. A high deprivation in at least one of them (category 1) causes the ownership to category 1 of overall deprivation if there is no deprivation in the other one. In fact, a household with high deprivation in both dimensions belongs to “most overall deprived households”, even if it is non-deprived for housing conditions.

The following group, a 29.76% of the families, corresponds to a light or intermediate degree of multidimensional poverty. Again, the analysis of associated frequencies allows proving that the “accommodation” dimension does not affects the membership to this class, since cells of any category of that dimension appear in this class. Furthermore, the other dimensions are the most important. They have a high degree of affordability of durable goods, not of life-style needs.

Finally, the last and larger class, composed by a 63.60% of households, has the higher probabilities for low deprivation categories in any dimension. For basic dimension, the conditional probability of belonging to minimum overall deprivation category is practically one and the secondary dimension, presents a smaller probability, though high, a 73.07%. This latter value is due to the greater similarities of belonging probabilities to the latent classes in “secondary needs".

Once the Bayesian assignment process of frequencies to classes is done, we can observed that those cells where the less deprived category for at least one of basic or secondary deprivation appear belong to this class.

In conclusion, the profile of most overall deprived households has as main feature the high levels of deprivation in two of the dimensions, combined, basic and secondary. Furthermore, housing conditions does not discriminate between the different classes of overall deprivation.

On the other side, basic and secondary dimensions are the most important ones. In fact, the presence of the most deprived category of at least one of them caused a household to belong to the group with higher overall deprivation and if appears the highest category of, at least, one it is considered as a non-deprived household. Finally, if both facts happen simultaneously, this household lies in light deprivation group.

4. Other methods for measuring deprivation

In this section, we use two different methodologies for measuring deprivation: Martinez y Ruiz Huerta’s (2000, henceforth MR) y D’Ambrosio y Peragine’s (2001, henceforth DP). Both studies consider deprivation as a metric variable. Therefore, they present problems in threshold definition. An important issue is where to fix the poverty line.

They differ from the procedure used before in weighting and aggregating indicators. Both papers build a composite index as a function of a set of binary variables.

Before commenting the results, we are going to introduce briefly the methodology of both approaches.

MR builds a partial composite index for each dimension: basic, housing and secondary. These indices are the weighted and normalised sum of the indicators considers in each dimension. For a household j $(j = 1,\ldots,n)$ and a dimension *m*, represented by a set of J^m items, the partial deprivation index can be computed as

$$I_j^m = 100 * \sum_j w_j^m d_{ij}^m \qquad [8]$$

where d_{ij}^{m} is the degree of deprivation experimented by a household i for a item j of dimension m and w_j^m is the weight for this item. d_{ij}^{m} is a binary variable that takes the value 0 if the household is not suffering deprivation concerning the jth item and the value 1 if it suffer it. The normalised weight w_j^m given to the jth item comes from the following expression

$$w_j^m = \frac{v_j^m}{\sum_j v_j^m} \qquad [9]$$

where v_j^m is the proportion of households not lacking the jth item. Therefore, the weights for each item are a function of the spread of the good or activity compared with the spread of the rest of goods or activities.

These indices vary between 0 and 100 and, so, each value can be interpreted as the percentage of deprivation suffered by a household related to the maximum value. This value corresponds to a household that suffers deprivation in all the items.

On the other hand, before showing DP's methodology we must introduce some notation. Given a set of functionings X and a set of households N, they denote by E_i the set of functionings from each household i is excluded. First, they identify the groups of households with a similar degree of deprivation. They consider that households $i, j \in N$ suffer the same deprivation if and only if $\#E_i=\#E_j$. Thus, they divide the population into classes according to the cardinality of the set E.

The first step is to measure the household deprivation, *HD*, the feeling of being alienated from the part of the society characterized by access to a wider set of functionings. This feeling can be expressed as

$$HD(i) = \sum_{j=0}^{i} (e_i - e_j)\pi_j \qquad [10]$$

where π_j is the proportion of people belonging to group j.

After defining this household deprivation, it should be taken into account that deprivation feeling depends not only on social distance from the other but, besides, on the amount of people with whom it feels similar. The higher this number, the lower household deprivation is. Thus, DP computed an effective household deprivation that includes the proportion of households with a wider set of functionings.

$$EHD(i) = \sum_{j=0}^{i} (e_i - e_j)\pi_j \sum_{j=0}^{i-1} \pi_j \qquad [11]$$

Moreover, we have computed the *EHD* index for each dimension.

In the table below, we show a statistical summary of the measures above. The results support the same conclusions based on latent deprivation index. Housing deprivation (MR and DP) has the lowest value and secondary deprivation the higher one.

Although there are some households with the maximum MR deprivation index (100), they are not very common as we can see in the percentiles. The maximum value for 75 percentile is the 38.20 for secondary deprivation, very far from 100.

On the other side, it is possible to find some households with an index equal to 0, that is, they do not suffer any deprivation. Exactly, a 18.90% of the households are in this situation, a 40.20% are free of basic deprivation, a 59.70% do not suffer housing deprivation and, finally, a 30.50% have some degree of secondary deprivation. Again, housing appears as the dimension with lesser deprivation and secondary deprivation is the higher one due to its nature, life style.

Table 9. Mean statistics for MR and DP deprivation indices

		MR_G	MR_B	MR_V	MR_S	EHD	EHBD	EHHD	EHSD
Mean		13.22	12.53	7.65	23.79	1.310	0.48	0.39	0.65
Median		9.63	7.89	0.00	21.67	0.43	0.16	0.00	0.38
Variance		170.37	273.63	154.00	593.91	4.03	0.74	0.63	0.85
Minimum		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum		94.36	100.00	100.00	100.00	16.50	5.916	7.33	4.22
Percentiles	**25**	2.67	0.00	0.00	0.00	0.04	0.00	0.00	0.00
	50	9.63	7.89	0.00	21.67	0.43	0.16	0.00	0.38
	75	19.86	22.13	10.85	38.20	1.49	0.81	0.36	1.07

Source: Author's elaboration from SPSS.

Moreover, it could be interesting to compute the correlation between the different dimensions. The results show that basic and secondary dimensions are the most important in order to get a high overall deprivation index. The housing dimensions appears again as the least important to what extent that the correlation coefficient between this dimension and the other are very low, specially with secondary dimension.

Moreover, when DP indices are analysed, this relationship is much looser. It is less than 0.5.

Table 10. Correlations between dimensions

	MR_G	MR_B	MR_V	MR_S
MR_G	1	0.831	0.674	0.829
MR_B	0.831	1	0.328	0.600
MR_V	0.674	0.328	1	0.297
MR_S	0.829	.600	0.297	1
	EHD	**EHBD**	**EHHD**	**EHSD**
EHD	1	0.817	0.654	0.806
EHBD	0.817	1	0.319	0.556
EHHD	0.654	0.319	1	0.287
EHSD	0.806	0.556	0.287	1

Source: Author's elaboration from SPSS.

Finally, if both indices are compared with the ones from the former section, we find that there is a good identification for the extreme categories, that is, highest and lowest deprived households are equally identified whatever method we use.

Since the intermediate categories of our model depend on some subsets of indicators and not on the cardinality of exclusion set or the amount of indicators expressing deprivation, this identification is more difficult.

Anyway, the biserial correlation coefficients show that there is a strong relationship between the categories and the values of MR and DP's indices, except for the secondary deprivation, where only the extreme groups are well identified.

5. 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 1996 shows that basic needs can be satisfied by the most of households, since only a 3.96% 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 1.76% of households belong to "most deprived" category. That is, a large proportion of households live in an accommodation without problems. Despite 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 very similar for each category (except the residual group that cannot afford some special facilities), 15.98%, 32.11%, and 44.59%, respectively. Among durable goods, the most deprived category only can afford a colour TV.

Once each deprivation dimension has been studied, they are combined. We found three groups again. Basic and secondary deprivations are the most important variables to decide the membership to an overall deprivation category. Besides, a 6.64% of households are the most deprived.

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.

---

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

[2] Cerioli and Zani (1990), Cheli *et al.* (1994), Cheli and Lemni (1995), Lemni *et al.* (1996) and Betti and Cheli (2001).

[3] For instante, when a household is asked about running water, there is no difference between 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.

References

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

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

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

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

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

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

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

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

Cantó, O. (2000b) *Climbing out of poverty, falling back in: low incomes' stability in Spain*. Documento de trabajo nº 13, Departamento de Economía Aplicada. Universidad de Vigo.

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

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

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

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

Fouarge and Muffels (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: Atkinson, A.B. y Hills, J. (eds.) *Exclusion, employment and opportunity*. CASE paper nº 4. London. CASE – London School of Economics.

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

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. Colchester. University of Essex.

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. Colchester. University of Essex.

Lazarsfeld, P.F. (1950). “The logical and mathematical foundation of latent structure analysis”. In: Stouffer, S.A. *et al.* (eds.), *Measurement and prediction*, 362-472, Princeton. Princeton University Press.

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

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*. London. Allen and Urwin.

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

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*. Oxford. Clarendon Press.

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*. London. Routledge and Kegan Paul.

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

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*. Harmondsworth. Penguin Books.

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

Whelan, C.T., Layte, R. y Maître, B. (2001b) *Persistent deprivation in the European Union*. European Panel Analysis Group Working Paper nº 19. Colchester. University of Essex.

Annex 1. ECHP list of variables

HF003	Can the household afford keeping its home adequately warm?
HF004	Can the household afford paying for holiday?
HF005	Can the household afford replacing worn-out furniture?
HF006	Can the household afford buy new clothes?
HF007	Can the household afford eating meal every second day?
HF008	Can the household afford having friends or family for drink/dinner?
HF009	Has the household been unable to pay scheduled rent for accommodation during the past 12 months?
HF010	Has the household been unable to pay scheduled mortgage payments during the past 12 months?
HF011	Has the household been unable to pay scheduled utility bills during the past 12 months?
HF012	Has the household been unable to pay other loan repayments rent for accommodation during the past 12 months?
HA008	Does the dwelling have separate kitchen?
HA009	Does the dwelling have bath or shower?
HA010	Does the dwelling have indoor flushing toilet?
HA011	Does the dwelling have running water?
HA014	Has the dwelling the problem of shortage of room?
HA018	Does the dwelling have leaky roof?

HA019	Does the dwelling have damp walls or floors?
HA020	Does the dwelling have rot in window frames or floors?
HB001	Possession of a car
HB002	Possession of a colour TV
HB003	Possession of a VCR
HB004	Possession of a micro wave
HB005	Possession of a dishwasher
HB006	Possession of a telephone

Annex 2. 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 *E*xpectation step all the expected values are computed given the observed values and the "current" model parameters. In the *M*aximization 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, $\hat{n}_{xabcd}$ estimates are used as observed frequencies. For doing this, the same procedure 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 *E*xpectation step to get new estimates for complete table frequencies. Iterations continue until convergence is reached.

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

$$\sum_{xabcd} n_{xabcd} \log \pi_{xabcd} \qquad [12]$$

It is posible to get suffiecient statistics for parameters if, firstly, we decompose the probability π_{xabcd} by means of [1] and [2] equations.

$$\begin{aligned}\sum_{xabcd} n_{xabcd} \log \pi_{xabcd} &= \sum_{x} n_{x\cdots\cdot} \log \pi_{x} + \sum_{x,a} n_{xa\cdots} \log \pi_{a|x} + \sum_{x,b} n_{x\cdot b\cdot\cdot} \log \pi_{b|x} + \\ &+ \sum_{x,c} n_{x\cdot\cdot c\cdot} \log \pi_{c|x} + \sum_{x,d} n_{x\cdots d} \log \pi_{d|x}\end{aligned} \qquad [13]$$

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,

$$E[n_{xabcd} \mid n_{abcd}, \pi_{xabcd}] = n_{abcd} \frac{\pi_{xabcd}}{\pi_{abcd}} = n_{abcd} \pi_{x|abcd} \qquad [14]$$

So, in the *E* step we have

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

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

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

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

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

In the *M* step and given the equation [14], estimates for parameters in the p+1-th iteration are computed given the sufficient statistics of complete data in the p-th iteration.

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

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

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

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

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

From these expressions, the estimates of the probabilities are

$$\hat{\pi}_x(p+1) = \frac{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{N}, \quad [15.a]$$

$$\hat{\pi}_{a|x}(p+1) = \frac{\sum_{b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}, \quad [15.b]$$

$$\hat{\pi}_{b|x}(p+1) = \frac{\sum_{a,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \quad [15.c]$$

$$\hat{\pi}_{c|x}(p+1) = \frac{\sum_{a,b,d} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \quad [15.d]$$

$$\hat{\pi}_{d|x}(p+1) = \frac{\sum_{a,b,c} n_{abcd}\hat{\pi}_{x|abcd}(p)}{\sum_{a,b,c,d} n_{abcd}\hat{\pi}_{x|abcd}(p)} \quad [15.e]$$

Iteration procedure must continue until the growth of the logarithm of the likelihood function was less than an 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 [15] it is posible to compute the probabilities

$$\hat{\pi}_{xabcd} \text{ and } \hat{\pi}_{abcd} = \sum_{x=1}^{X*} \hat{\pi}_{xabcd} \,. \qquad [16]$$