UNIVERSITÀ
DI SIENA
1240

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Giovanni Marano
Gianni Betti
Francesca Gagliardi

Latent class Markov models for addressing measurement problems in poverty dynamics

n. 695 – Aprile 2014

Abstract - The traditional approach to poverty measurement utilises only monetary variables as indicators of individuals' intensity of the state of deprivation, causing measurement errors of the phenomenon under investigation. Moreover, when adopted in a longitudinal context, this approach tends to overestimate transition poverty. Since poverty is not directly observable, a latent definition can be adopted: in such a conception is possible to use Markov chain models in their latent acceptation. This paper proposes to use Latent class Markov models which allow taking into account more observed (manifest) variables. We define those variables via monetary and non-monetary fuzzy indicators.

JEL Classification: I32, C13

Keywords: Poverty dynamics, Measurement errors, LCMM

Giovanni Marano, Department of Economics and Statistics, University of Siena; email: giovanni.27@alice.it;
Gianni Betti, Department of Economics and Statistics, University of Siena; email: gianni.betti@unisi.it;
Francesca Gagliardi, Department of Economics and Statistics, University of Siena; email: gagliardi10@unisi.it

1. Introduction

The conceptualisation and assessment of poverty involves at least four dimensions: multiple aspects covering both monetary and non-monetary deprivation; diverse statistical measures in order to capture different facets of poverty; the time-dimension; and comparability over space and time. This paper addresses the time-dimension of poverty, which involves two types of measures: (1) measures of poverty trend over time at the aggregate level; (2) measures of persistence or otherwise of poverty at the micro level. In particular, attention is focused on the distinction between transitory and persistent poverty, which cannot be distinguished in a traditional cross-sectional analysis.

In the recent literature, several papers have focussed on constructing measures of lifetime or intertemporal poverty; see, for example, Porter and Quinn (2008), Calvo and Dercon (2009), Hoy and Zheng (2011), Bossert, Chakravarty and D'Ambrosio (2012), Gradin, del Rio and Cantó (2012), Mendola and Busetta (2012). Duclos, Araar and Giles (2010) propose measures of chronic and transient poverty but these categories apply to aggregates and not to individuals.

Hulme and Shepherd (2003) have described the chronically poor as follows: "… *intuitively, we are talking about people who remain poor for much of their life course, and who may 'pass on' their poverty to subsequent generations*…". However, the specific criterion used to identify the chronically poor (and hence the transiently poor) is a subject of continuing debate. Foster and Santos (2013) distinguish two identification approaches: a *counting approach* and a *permanent income approach*.

Such rich literature rarely focuses on the problem of measurement errors and their effects on poverty dynamics. According to Verma and Gagliardi (2013) " *…The main measurement problem in the assessment of poverty trends is the definition of the poverty threshold and its consistency over time. The main measurement problem in the assessment of persistence of poverty concerns the effect of random measurement errors on the consistency of the individuals' observed poverty situation at different times…*"

In fact, poverty is not directly observable from sample surveys, especially when monetary variables are collected as indicators of individuals' intensity of the state of deprivation: this causes measurement errors in the phenomenon under investigation. Moreover, when adopted in a longitudinal context, the traditional approach tends to overestimate transition poverty, since small changes in the measured income or consumption expenditure can result in an individual crossing the poverty line.

To overcome these problems Betti (1996) adopted a latent definition for poverty: in such a conception it is possible to use Markov chain models to correct measurement errors in panel data.

The novelty of the paper consists of: i) using Latent class Markov models which allow taking into account more observed (manifest) variables; and ii): defining those variables via monetary and non-monetary fuzzy indicators.

The paper is composed of five sections. After the present introduction, Section 2 describes the latent class Markov models recently proposed in the literature and used for analysing poverty dynamics in the paper. Section 3 describes the fuzzy set approach to measure monetary and non-monetary (multidimensional) poverty cross-sectionally. Section 4 reports the results of the empirical analysis conducted on the 4-year balanced panel form EU-SILC waves 2006-2009 for France, Italy and United Kingdom, while Section 5 concludes the paper.

2. Latent class Markov models

A Markov chain is a discrete-time stochastic process characterized by the basic assumption that the next state only depends on the current state, regardless of the whole past history: these processes are suitable whenever one wants to model a situation where the future is independent of the past, given the present, as follows:

$$P(X_t = x_t \mid X_0 = x_0, X_1 = x_1 \ldots X_{t-1} = x_{t-1}) = P(X_t = x_t \mid X_{t-1} = x_{t-1}) \qquad (1)$$

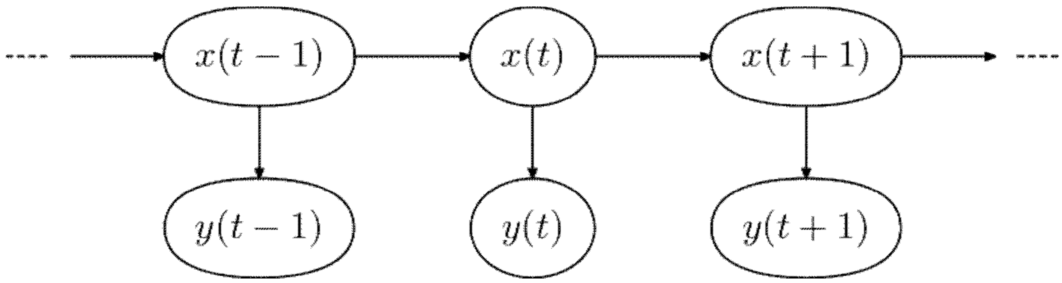
A natural extension of Markov chains are the Hidden Markov Models or Latent Class Markov Models (Baum, 1966): such models are characterized by an unobservable discrete Markov process and by a sequence of observable responses, which could be defined on a discrete or continuous support. The unobservable process is called “hidden” or “latent” since we only observe the output responses, and each of this responses is a deterministic or stochastic function of the current latent state. Therefore, Latent Class Markov Models may be seen as a bivariate stochastic process where two basic assumptions hold:

$$P(X_t \mid X_{t-1}, X_{t-2} \ldots X_1, Y_t) = P(X_t \mid X_{t-1}) \qquad (2)$$

$$P(Y_t \mid X_t, X_{t-1} \dots X_1, Y_{t-1}, Y_{t-2} \dots Y_1) = P(Y_t \mid X_t) \qquad (3)$$

the latent variable *X* at the current moment, given the one at the immediately previous moment, is independent of the all previous states, and the observed (manifest) variable *Y* at the current moment, given the corresponding latent state, is independent of all previous observations. An Hidden Markov Model can be seen as shown in Figure 1 below.

Figure 1. Graphic representation of an Hidden Markov Model



In our study, latent variables are considered as the real individual's states, i.e. the latent meaning of variable under investigation, whilst output variables are considered as responses affected by errors, because they are collected by a survey.

Latent Class Markov Models are divided in discrete and continuous, according to their support. In discrete Latent Class Markov Models, two fundamental matrices, along with an initial vector, need to be estimate: a matrix for transition probabilities between latent states, a matrix for response probabilities (i.e. probabilities of observing responses, given the latent state) and a vector for initial probabilities of being in each latent class. Rabiner and Juang (1986) give an exhaustive overview of methods and problems related to these models, providing applications on speech recognition. In continuous Latent Class Markov Models, the matrix of conditional response probabilities is replaced by a matrix containing the response distribution parameters, conditionally on each latent state.

In the paper we focus our attention on application of Latent Class Markov Models for panel data (Maruotti, 2011; Bartolucci *et al.*, 2012), i.e. an extension of the hidden Markov structure for univariate time series towards models describing more individual processes. This feature marks the difference in terminology between Hidden Markov Models and Latent Class Markov Models. The first is typically referred to data structures with long time series and short number of individuals, whereas the second is typically referred to data structures with a large number of individuals and few instants of time[1].

Moreover, we will handle the effects of some covariates on the output responses, by using these models in a regression contest. Generalized Linear Models (McCullagh and Nelder, 1989) are a generalization of ordinary linear regression, which allows for non-normally distributed response variables: regression can be fitted for response variables belonging to exponential family, by means of some link functions transforming the expected value in a linear predictor. We will extend these procedures to Latent Class Markov Models for multiple output sequences by supposing that the observed responses of Markov process are affected by a set of known variables: finally, an expected value of output variable given the

[1] Thus, in our context Latent class Markov Models is the most appropriate term.

latent state and the value of covariates will be obtained. Above models will be applied to longitudinal data from a panel survey, i.e. a survey repeated on the same sample over time, suitable whenever the goal is to study the dynamic population processes along with estimates of particular parameters such as net and gross changes into population. Some devices as rotating scheme may be used in a panel in order to find a payoff between estimation accuracy and representativeness of the sample.

We will use the Expectation-Maximization (EM) algorithm in order to estimate model parameters. It is a two-steps algorithm in which at first a conditional expectation of model log-likelihood is computed, according to the current values of parameters, and then a new estimation of parameters is obtained by maximizing this function: the two steps are iteratively repeated till convergence.

3. Fuzzy measures of monetary and non-monetary deprivation

In the traditional approach, poverty is characterized by a simple dichotomization of the population into poor and non-poor defined in relation to some chosen poverty line. This poverty line may represent a certain percentage of the mean or the median of the income distribution. This approach presents two main limitations: firstly, it is unidimensional, since it refers to only one proxy of poverty and secondly, it divides the population into a simple dichotomy. However, poverty is a complex phenomenon that cannot be reduced solely to the monetary dimension but must also take account of non–monetary indicators of living conditions. In fact, many recent papers have considered dimensions beyond income (or consumption) in identifying and measuring poverty: see, for example, Atkinson (2003), Bourguignon and Chakravarty (2003), Alkire and Santos (2010), Alkire and Foster (2011).

Moreover, poverty it is not an attribute that characterises an individual in terms of presence or absence, but is rather a state that manifests itself in different shades and degrees. The fuzzy approach considers poverty as a matter of degree rather than an attribute of the individual. An early attempt to incorporate the concept of poverty as a matter of degree at methodological level was made by Cerioli and Zani (1990) who drew inspiration from the theory of fuzzy sets. Subsequently, Cheli and Lemmi (1995) proposed the so called Totally Fuzzy and Relative (TFR) approach in which the membership function - quantitative specification of individuals' or households' degrees of poverty and deprivation - is defined as the distribution function of income, normalised (linearly transformed) so as to equal 1 for the poorest and 0 for the richest person in the population. Betti and Verma (2008) defined the membership function of monetary and non-monetary deprivation for any individual *i* as:

$$\mu_{i,K} = \left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma \mid X_\gamma > X_i}{\sum_{\gamma=2}^{n} w_\gamma \mid X_\gamma > X_1} \right)^{\alpha_K - 1} \left(\frac{\sum_{\gamma=i+1}^{n} w_\gamma X_\gamma \mid X_\gamma > X_i}{\sum_{\gamma=2}^{n} w_\gamma X_\gamma \mid X_\gamma > X_1} \right) \tag{4}$$

where X is the equivalised income in the monetary deprivation index, or the overall score in the non-monetary deprivation index; w_γ is the sample weight of individual of rank γ ($\gamma = 1,...,n$) in the ascending distribution, and α_K (K = 1,2) are two parameters, one for monetary and the other for non-monetary indicator. Each of the two parameters α_K is estimated so that the mean of the corresponding membership function is equal to the at-risk-of-poverty rate computed on the basis the official Eurostat poverty line (60% of the median income). We refer to the monetary based indicator as Fuzzy Monetary (FM), and to the non-monetary indicator as Fuzzy Supplementary (FS).

4. Empirical analysis

In the present work we have used data from the EU Statistics on Income and Living Conditions (EU-SILC) survey, distributed by Eurostat. EU-SILC is designed to collect detailed information on the income of each household member, on various aspects of the material and demographic situation of the household, and for producing structural indicators on social cohesion. EU-SILC surveys are conducted annually in each country, based on a rotational panel design. The national sample designs and sizes have been determined primarily for the purpose of estimation and reporting of indicators at the national level, with limited breakdown by major socio-demographic subgroups of the population. It provides two types of annual data: i) cross-sectional data; and ii) longitudinal data pertaining to individual-level changes over time, observed periodically over a four year period for each person.

In the present work we have used the longitudinal 2009 data set covering years from 2006 to 2009. For each wave, we have calculated the FM and the FS measures according to equation (1). Following Betti *et al.* (2012), we have grouped the total number of deprivation items adopted for constructing the overall non-monetary measures (FS) into meaningful dimensions. Exploratory and confirmatory factor analyses, as proposed in Whelan *et al.* (2001), allow us to achieve this objective. We have applied such procedures to three EU countries - France, Italy and United Kingdom – to identify six dimensions of non-monetary deprivation:

FS1 = Basic lifestyle; FS2 = Consumer durables; FS3 = Housing amenities; FS4 = Financial situation; FS5 = Work & Education; FS6 = Health related.

Table 1 reports average poverty and deprivation measures for the three countries under study. Column *H_cs* shows the percentage of poor individuals (the at-risk-of-poverty rate) based on the cross-sectional EU-SILC samples and published by Eurostat. All the remaining columns are based on the balanced 4-year panels[1]: those averages are weighted by EU-SILC target variable RB064 (Longitudinal weight, four-year duration), which lets the balanced panels to be representative of the population present continuously in the country over the period. This variable is constructed via a calibration procedure based on a set of post-stratification variables (see Verma *et al.*, 2007 for details). It is possible to observe that, in the case of France and UK, FM (and FS) differs significantly from *H_cs*.

[1] These balanced panels are the base for the analysis via Latent class Markov Models.

There are at least two explanations: i) “longitudinal” population may differ from cross-sectional population; ii) longitudinal weights RB064 have been calibrated using variables which are not very strongly correlated with monetary poverty. This leads to a new source of measurement errors, which could be corrected by the latent dimension intrinsic in the Markov models.

Table 1: Mean poverty and deprivation measures in cross-sectional and balanced panel samples

	H_cs	***FM***	***FS***	***FS1***	***FS2***	***FS3***	***FS4***	***FS5***	***FS6***
FR_06	13.2	12.7	13.4	10.1	6.1	5.2	29.8	11.2	17.0
FR_07	13.1	12.4	13.1	9.8	5.6	5.2	30.1	10.9	16.5
FR_08	12.7	11.9	12.8	9.7	5.3	6.0	7.0	10.5	16.2
FR_09	12.9	11.9	12.7	9.8	5.1	5.9	8.1	7.0	17.4
IT_06	19.6	19.8	18.5	15.7	6.9	5.8	24.3	14.4	21.5
IT_07	19.8	19.8	17.6	14.6	6.2	5.6	26.9	14.8	22.8
IT_08	18.7	18.6	17.0	15.0	6.0	5.4	26.3	13.7	21.6
IT_09	18.4	18.4	16.8	13.9	5.3	6.0	9.4	7.3	21.1
UK_06	19.0	17.9	18.0	12.8	6.5	7.1	27.3	13.3	17.4
UK_07	18.6	16.9	16.7	13.1	5.9	6.1	22.8	9.4	16.2
UK_08	18.7	17.7	17.9	13.3	6.1	6.0	26.4	13.5	14.8
UK_09	17.3	17.0	17.7	13.3	5.8	5.8	27.4	13.8	15.0

Generalized Linear Models have been used to fit the data: we have started by modelling FM and FS singularly as Gamma-distributed response variables with inverse canonical link function, and then have put them together in a bivariate model, i.e. a model with a shared hidden Markov chain and two response variables. We also have tried to insert FS1-FS6 into the model, but their distributions are not suitable to be fitted as a Gamma variable. They need some appropriate transformation, which we hope to implement in future research.

Latent class Markov models need to be initialised with starting values for parameters. As initial values we have used parameters obtained from a similar analysis carried out according to the traditional approach. Covariates introduced into the models were: sex, age, education level and marital status; macro-region have also been introduced for Italy. To begin with, we used SAS to prepare and process the data. Subsequently we used R to carry out the analysis and obtain the results.

For each state, we have tested all combinations of covariates in order to find the best model by minimising Akaike Information Criterion (AIC). The influence of covariates is measured on the bivariate FM-FS response variable. In Table 2 the best set of covariates is described. The final model for France includes sex, education level, age and marital status. The lowest education level and marital status in the set “widowed, separated or divorced” have a strong negative effect on both FM and FS poverty status. Being female has similarly a negative effect on both FM and FS, though effect is less marked. By contrast age between 35 and 64 has a positive effect on FM but a negative one on FS poverty status. The final model for Italy includes only sex and education level. Female gender and lowest education level have a negative effect on both FM and FS poverty status. Models with macro-region covariate were also tried in the case of Italy, but they have a higher AIC. The final model for UK includes the education level again, along with the age. Education level has the same effect described before, while age has two different effects, one for 35-64 years old and another for 65+ years old persons. Both age groups have a negative effect on

FM poverty status compared to the 0-35 baseline category. On FS poverty status the effect of 35-64 category is still negative, but the effect of 65+ category is positive. Hidden process has two states, defined as poor and non-poor: this latent condition only affects the magnitude of coefficients, but it does not affect their sign.

Table 2: Best set of covariates

Country	***Sex***	***Age***	***Education level***	***Marital status***	***Macro-region***
FR	✓	✓	✓	✓	-
IT	✓		✓		
UK		✓	✓		-

Table 3: Initial and transition probabilities, Latent class Markov models

Country	***Latent states***	***Initial probabilities***	***Transition probabilities***	
			poor	non-poor
FR	poor	0.143	0.871	0.129
	non-poor	0.857	0.020	0.980
IT	poor	0.203	0.878	0.122
	non-poor	0.797	0.020	0.980
UK	poor	0.182	0.856	0.144
	non-poor	0.818	0.029	0.971

Table 3 reports initial and transition probabilities estimated via Latent class Markov models. Initial non-poverty probability is higher in France

(85.7%) than in UK and Italy (81.8% and 79.7% respectively). Transition matrices show high probabilities of remaining in the non-poor state (around 98%) and low probabilities of moving from non-poor to poor state in all countries. Transition probabilities from poor to non-poor state are higher, with a small variation from 12.2% for Italy to 14.4% for UK.

5. Concluding remarks and future research

The above empirical study shows the benefit of the use of latent class Markov models in longitudinal poverty analysis, since the attained results were not clear from simply observing cross-sectional analysis along time.

There are several potential extensions of the work presented here. From a theoretical point of view, one may try to identify a larger set of link functions better fitting the manifest variable distributions. From an empirical point of view, we would like to see a full comparative analysis among the 28 EU countries, and including in the models the non-monetary dimensions FS1-FS6.

References

Alkire, S., Foster, J.E. (2011), Counting and Multidimensional Poverty Measurement, *Journal of Public Economics*, **95**, pp. 476–487.

Alkire, S., Santos, M.E. (2010), Acute Multidimensional Poverty: A New Index for Developing Countries, OPHI Working Paper No 38.

Atkinson, A.B. (2003), Multidimensional deprivation: contrasting social welfare and counting approaches, *Journal of Economic Inequality*, **1**, pp. 51-65.

Bartolucci, F., Farcomeni, A., Pennoni, F. (2012), *Latent Markov Models for Longitudinal Data*, Chapman and Hall/CRC press.

Baum, L.E., Petrie, T. (1966), Statistical inference for probabilistic functions of finite state Markov chain, *Annals of Mathematical Statistics*, **37**, pp. 1554-1563.

Betti, G. (1996), A longitudinal approach to poverty analysis: the latent class Markov model, *Statistica,* **56**, pp. 345-359.

Betti, G., Gagliardi, F., Lemmi, A., Verma, V. (2012), Sub-national indicators of poverty and deprivation in Europe: methodology and applications, *Cambridge Journal of Regions, Economy and Society*, **5**, pp. 149-162.

Betti, G., Verma, V. (2008), Fuzzy measures of the incidence of relative poverty and deprivation: a multi-dimensional perspective, *Statistical Methods and Applications*, **17**, pp. 225-250.

Bossert, W., Chakravarty, S.R., D'Ambrosio, C. (2012), Poverty and Time, *Journal of Economic Inequality*, **10**, pp. 145-162.

Bourguignon, F., Chakravarty, S. (2003), The measurement of multidimensional poverty, *Journal of Economic Inequality*, **1**, pp. 25-19.

Calvo, C., Dercon, S. (2009), Chronic Poverty and All That: The measurement of Poverty Over Time. In: Addison, T., Hulme, D., Kanbur, R. (eds.), *Poverty Dynamics: Interdisciplinary Perspectives*, pp. 29-58. Oxford University Press.

Cerioli, A., Zani, S. (1990), A fuzzy approach to the measurement of poverty. In: Dagum, C., Zenga, M. (eds.) *Income and wealth distribution, inequality and poverty*, pp. 272-284. Springer Verlag, Berlin.

Cheli, B., Lemmi, A. (1995), A Totally Fuzzy and Relative Approach to the Multidimensional Analysis of Poverty, *Economic Notes*, **24**, pp. 115-134.

Duclos, J., Araar, A., Giles, J. (2010), Chronic and Transient Poverty: Measurement and Estimation, with Evidence from China, *Journal of Development Economics*, **91**, pp. 266-277.

Foster, J.E., Santos, M.E. (2013), Measuring chronic poverty. In: Betti, G., Lemmi, A. (eds.), *Poverty and Social Exclusion: New Methods of Analysis*, pp. 143-165, Abington: Routledge.

Gradin, C., Del Rio, C., Cantó, O. (2012), Measuring Poverty Accounting for Time, *The Review of Income and Wealth*, **58**, pp. 330-354.

Hoy, M., Zheng, B. (2011), Measuring Lifetime Poverty, *Journal of Economic Theory*, **146**, pp. 2544-2562.

Hulme, D., Shepherd, A. (2003), Conceptualizing Chronic Poverty, *World Development*, **31**, pp. 403-423.

Maruotti, A. (2011), Mixed Hidden Markov Models for longitudinal data: an overview, *International Statistical Review*, **79**, pp. 427-454.

McCullagh, P., Nelder, J.A. (1989), *Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC, Boca Raton, Florida.

Mendola, D., Busetta, A. (2012), The Importance of Consecutive Spells of Poverty: A Path Dependent Index of Longitudinal Poverty, *The Review of Income and Wealth*, **58**, pp. 355-374.

Porter, C., Quinn, N.N. (2013), Measuring Intertemporal Poverty: Policy Options for the Poverty Analyst. In Betti, G., Lemmi, A. (eds.), *Poverty and Social Exclusion: New Methods of Analysis*, pp. 166-193, Abington: Routledge.

Rabiner, L.R., Juang, B.H. (1986), An Introduction to Hidden Markov Models, *IEEE ASSP Magazine*, **3**, pp. 4-16

Verma, V., Betti, G., Ghellini, G. (2007), Cross-sectional and longitudinal weighting in a rotational household panel: applications to EU-SILC, *Statistics in Transition*, **8**, pp. 5-50.

Verma, V., Gagliardi, F. (2013), On assessing the time-dimension of poverty. In: Betti, G., Lemmi, A. (eds.), *Poverty and Social Exclusion: New Methods of Analysis*, pp. 109-127, Abington: Routledge.

Whelan, C.T., Layte, R., Maitre, B., Nolan, B. (2001), Income, deprivation and economic strain: an analysis of the European Community Household Panel, *European Sociological Review*, **17**, pp. 357-372.