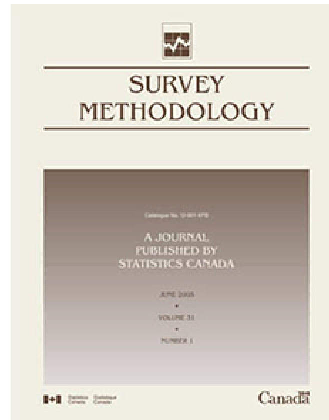
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

A mixed latent class Markov approach for estimating labour market mobility with multiple indicators and retrospective interrogation



by Francesca Bassi, Marcel Croon and Davide Vidotto

Release date: June 22, 2017

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

email at STATCAN.infostats-infostats.STATCAN@canada.ca

telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2017

All rights reserved. Use of this publication is governed by the Statistics Canada Open Licence Agreement.

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2017
Vol. 43, No. 1, pp. 107-124
Statistics Canada, Catalogue No. 12-001-X

A mixed latent class Markov approach for estimating labour market mobility with multiple indicators and retrospective interrogation

Francesca Bassi, Marcel Croon and Davide Vidotto[1]

Abstract

Measurement errors can induce bias in the estimation of transitions, leading to erroneous conclusions about labour market dynamics. Traditional literature on gross flows estimation is based on the assumption that measurement errors are uncorrelated over time. This assumption is not realistic in many contexts, because of survey design and data collection strategies. In this work, we use a model-based approach to correct observed gross flows from classification errors with latent class Markov models. We refer to data collected with the Italian Continuous Labour Force Survey, which is cross-sectional, quarterly, with a 2-2-2 rotating design. The questionnaire allows us to use multiple indicators of labour force conditions for each quarter: two collected in the first interview, and a third one collected one year later. Our approach provides a method to estimate labour market mobility, taking into account correlated errors and the rotating design of the survey. The best-fitting model is a mixed latent class Markov model with covariates affecting latent transitions and correlated errors among indicators; the mixture components are of mover-stayer type. The better fit of the mixture specification is due to more accurately estimated latent transitions.

Key Words: Gross flows; Labour market; Mixture models; Latent class models.

1 Introduction

Analysts can exploit panel data to estimate labour force gross flows - i.e., transitions in time between different states. Net flows measure variations in time in various market states, whereas gross flows provide information on the dynamics of the labour market.

A large body of literature on gross flows estimation is based on the assumption that errors are uncorrelated over time, i.e., they are Independent Classification Errors (ICE). The ICE assumption implies that: (i) classification errors referring to two different occasions are independent of each other conditionally on the true states, and (ii) errors only depend on the present true state. Thus, classification errors produce spurious transitions and consequently induce overestimation of changes.

However, in many contexts, the ICE assumption turns out not to be realistic, because of the survey design and data collection strategies. In these circumstances, classification errors may be correlated: observed states may also depend on true states at other times or on true transitions, or direct effects may exist between observed states (Bound, Brown and Mathiowetz 2001).

In this paper, we use a model-based approach to adjusting observed gross flows for classification errors. It combines a structural sub-model for unobserved true transition rates and a measurement sub-model relating true states to observed ones. A convenient framework for formulating our model is provided by latent class (LC) analysis.

1. Francesca Bassi, Department of Statistical Sciences, University of Padova, Italy, Via C. Battisti 241, 35121, Padova, Italy. E-mail: francesca.bassi@unipd.it; Marcel Croon and Davide Vidotto, Department of Methodology and Statistics, University of Tilburg, The Netherlands.

We apply our approach to observed gross flows among the three labour force states - Employed (E), Unemployed (U) and Not in the labour force (N) - taken from the Italian Continuous Labour Force Survey (CLFS), a quarterly survey with a 2-2-2 rotating design which yields two-wave panels one quarter, three quarters and one year apart. We consider data collected from 2005 to 2009.

The questionnaire allows us to use multiple indicators of labour force conditions for each quarter: (i) all respondents are classified as Employed, Unemployed, or Not in the labour force, according to the definition of the International Labour Office (ILO) on the basis of answers given to a group of questions; (ii) respondents are asked to classify themselves as employed, unemployed, or not in the labour force, the self-perceived condition; (iii) a retrospective question asks about respondents' state in the labour market one year before the interview. This approach provides a way of estimating labour market mobility by taking into account correlated measurement errors and the rotating design of the survey.

In detail, the best-fit model is a mixed latent class Markov (LCM) model with covariates affecting latent transitions and correlated errors among indicators. The mixture is obtained by assuming the existence of two unobservable sub-populations, movers, i.e., respondents who change their state in the labour market during the observation period, and stayers. A secondary result of our research is that the mover-stayer model and the LCM estimate the same amount of measurement error in the data. The better fit of the mixture specification is due to more accurately estimated latent transitions. Magidson, Vermunt and Tran (2007) also found that the mixed LC Markov model has a better fit to the data than the traditional one. However, in that case, the difference in fit was due to the fact that, as heterogeneity was not taken into account, the result was overestimation of measurement error.

Our paper follows recent contributions to the scientific literature on the topic of gross flows estimation with hidden Markov chain and multiple indicators. An accurate description of the model may be found in Langeheine (1994). The method was not only applied to estimation of labour market gross flows but also to many other contexts, longitudinal data being available. Paas, Vermunt and Bijmolt (2007), for example, estimated an LCM model to study acquisitional patterns in the financial product market; multiple indicators of ownership of financial products were used to identify not directly observable market segments among which customers could move on consecutive measurement occasions. Bartolucci, Lupparelli and Montanari (2009) estimated the same model in following changes in health status in a sample of patients over time. Manzoni, Vermunt, Luijkx and Muffels (2010) applied an LCM model to estimate gross flows in the Swedish labour market. In a more recent work, Pavlopoulos and Vermunt (2015) used a hidden Markov model to estimate the amount of measurement error in information from the Dutch Labour Force Survey and the Dutch Institute for Employee Insurance on the type of job (permanent or temporary).

The contribution of this paper to the scientific literature on the topic of gross flows estimation is that we have three indicators, one of them collected retrospectively, on labour force state and we can also take into account the rotating design of the survey. The paper also contributes to the literature on the quality of data from the CLFS (Bassi, Padoan and Trivellato 2012).

The paper is organised as follows. Section 2 introduces the traditional (or standard) and the mixed LCM model. Section 3 describes the survey and its data. Section 4 compares the performances of the traditional versus mixed LCM models. Section 5 provides results, referring to the best fitting model to correct gross flows in the labour market from measurement errors. Section 6 concludes.

2 The latent class Markov model

Latent class analysis has been applied in a number of studies on panel data to separate true changes from observed ones affected by unreliable measurements. Relatively recent contributions include Bassi, Torelli and Trivellato (1998), Biemer and Bushery (2000), Bassi, Croon, Hagenaars and Vermunt (2000), Bassi and Trivellato (2009).

The true labour force state is treated as a latent variable and the observed one as its indicator. The model consists of two parts:

a) structural, describing true dynamics among latent variables;
b) measurement, linking each latent variable to its indicator(s).

Let us consider the simplest formulation of latent class Markov (LCM) models (Wiggins 1973), which assumes that true unobservable transitions follow a first-order Markov chain. As in all standard LCM specifications, local independence among indicators is assumed, i.e., indicators are independent conditionally on latent variables In the LCM model with one indicator per latent variable, the assumption of local independence coincides with the Independent Classification Errors condition.

Let X_{it} denote the true labour force condition at time t for a generic sample individual $i, i = 1, \ldots, n$; Y_{it} is the corresponding observed condition; $P(X_{i1} = l_1)$ is the probability of the initial state of the latent Markov chain, and $P(X_{it+1} = l_{t+1} \mid X_{it} = l_t)$ is the transition probability between state l_t and state l_{t+1} from time t to $t+1,$ with $t = 1, \ldots, T-1,$ where T represents the total number of consecutive, equally spaced time-points over which an individual is observed. In addition, $P(Y_{it} = j_t \mid X_{it} = l_t)$ is the probability of observing state j at time $t,$ given that individual i at time t is in the true state l_t: this is also called the model measurement component.

It follows that $P(Y(1), \ldots, Y(T))$ is the proportion of units observed in a generic cell of the $T-$way contingency table. For a generic sample individual $i,$ a LCM model is defined as:

$$P(\mathbf{Y}_i = \mathbf{y}) = \sum_{l_1}^{K} \ldots \sum_{l_T}^{K} P(X_{i1} = l_1) \prod_{t=2}^{T} P(X_{it} = l_t \mid X_{it-1} = l_{t-1}) \prod_{t=1}^{T} P(Y_{it} = j_t \mid X_{it} = l_t) \tag{2.1}$$

where $\mathbf{y}$ is the vector containing observed values for individual i, l_t and j_t vary over K classes (in our application, three labour force conditions). Equation (2.1) specifies the proportion of units in the generic cell of a $T-$way contingency table as a product of marginal and conditional probabilities.

In an LCM model with concomitant variables, latent class membership and latent transitions are expressed as functions of covariates with known distributions (Dayton and McReady 1988). $P(X_{i1} = l_1 | \mathbf{Z}_{i1} = \mathbf{z}_1)$, where $\mathbf{z}_1$ is a vector containing the values of covariates for respondent i at time 1, estimates covariate effects on the initial state, and $P(X_{it} = l_t | X_{it-1}, \mathbf{Z}_{it} = \mathbf{z}_t)$, where $\mathbf{z}_t$ is a vector containing the values of covariates for respondent i at time t, estimates covariate effects on latent transitions.

On the basis of the above components, the complete model for individual i is given by:

$$\begin{aligned} P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) = & \sum_{l_1}^{K} \ldots \sum_{l_T}^{K} P(X_{i1} = l_1 | \mathbf{Z}_1 = \mathbf{z}_1) \\ & \prod_{t=2}^{T} P(X_{it} = l_t | X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \\ & \prod_{t=1}^{T} P(Y_{it} = j_t | X_{it} = l_t) \end{aligned} \tag{2.2}$$

When more than one (M) indicators per latent variable are observed, the model formulation becomes the following (Vermunt 2010):

$$\begin{aligned} P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) = & \sum_{l_1}^{K} \ldots \sum_{l_T}^{K} P(X_{i1} = l_1 | \mathbf{Z}_1 = \mathbf{z}_1) \\ & \prod_{t=2}^{T} P(X_{it} = l_t | X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \\ & \prod_{m=1}^{M} \prod_{t=1}^{T} P(Y_{mit} = j_t | X_{it} = l_t) \end{aligned} \tag{2.3}$$

In our application, the M indicators are given by the three pieces of information collected for all respondents on their labour market condition.

Typically, conditional probabilities are parameterised and restricted by logistic regression models. The parameters are estimated via maximum likelihood (Vermunt and Magidson 2013). Identification is a well-known problem in models with latent variables and, although the number of independent parameters must not exceed the number of observed frequencies, this is not a sufficient condition. According to Goodman (1974), a sufficient condition for local identifiability is that the information matrix is positive definite. Latent Gold software (Vermunt and Magidson 2008), provides information on parameter identification. Another problem linked to estimation is that of local maxima, to deal with which we estimated our models several times with different sets of starting values.

A mixed LCM model assumes the existence in the population of not directly observable groups moving across time, following latent chains with different initial state probabilities and different transition probabilities; the groups may also be assumed to have different response probabilities (van de Pol and Langeheine 1990). Such a model can be extended to include time-varying and time-constant covariates

(Vermunt, Tran and Magidson 2008). A special case of a two-class mixed LCM model is the mover-stayer model: the group of movers has positive probabilities of transferring from one state to another over time, and the group of stayers do not change. For the latter, transition probabilities between different states are imposed as zero. A two-class mixed LCM model with concomitant variables has the following form:

$$P\left(\mathbf{Y}_i = \mathbf{y} \mid \mathbf{Z}_i = \mathbf{z}\right) = \sum_{w=1}^{2}\sum_{l_1}^{K}\ldots\sum_{l_T}^{K} P\left(W = w\right) P\left(X_{i1} = l_1 \middle| \mathbf{Z}_1 = \mathbf{z}_1, W = w\right)$$
$$\prod_{t=2}^{T} P\left(X_{it} = l_t \middle| X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t, W = w\right)$$
$$\prod_{j_t=1}^{K}\prod_{t=1}^{T} P\left(Y_{it} = j_t \middle| X_{it} = l_t, W = w\right) \tag{2.4}$$

where W is a binary latent variable. The mover-stayer model is obtained assuming, for $l_t \neq l_{t-1}$, $P\left(X_{it} = l_t \middle| X_{it-1} = l_{t-1}, W = 2\right) = 0$ and, consequently, for $l_t = l_{t-1}$ $P\left(X_{it} = l_t \middle| X_{it-1} = l_{t-1}, W = 2\right) = 1$.

The likelihood function of an LC model can also be estimated if information is missing in the response variables. We exploit this opportunity to take into account the response patterns generated by the survey rotation design. Sampled households are interviewed for two consecutive quarters, do not participate in the survey for the subsequent two quarters, and are then re-interviewed on two other occasions (see Table 3.1). We assumed that missing information due to survey design is missing at random. In this case, each unit only contributes to the likelihood function with the information available (Vermunt 1997).

3 The data

The Continuous Labour Force Survey (CLFS), conducted by ISTAT (Italian Institute of Statistics), is the main and official source of statistical documentation on the Italian labour market. The CLFS has been conducted since 1969 and has been modified many times. In 2004, major updating was carried out, mainly dictated by the requirement to adapt the survey to new EU (European union) standards. The principal changes involved interviews distributed throughout the years of the study, new criteria to classify respondents' status in the labour market, computer-assisted data collection techniques, and dependent interviewing. Every year the survey collects information on about 280,000 households, for a total of about 700,000 individuals. The reference population consists of all household members officially resident in Italy.

The Italian CLFS sampling design has two stages: 1) municipalities were denominated as primary sampling units (PSUs) with stratification, and households as final sampling units (FSUs) with rotation. PSUs were stratified according to demographic size. Large municipalities, with population over a given threshold (also called self-representative municipalities), were always included in the sample; smaller municipalities (not self-representative) were grouped in strata, so that one municipality in each stratum was selected with probability proportional to its population; 2) households were randomly selected from the population registers in all municipalities drawn at stage 1.

The survey was quarterly with a 2-2-2 rotating design. Householders were interviewed in two consecutive quarters. After a two-quarter break, they were interviewed again, twice in the corresponding two quarters of the following year. As a result, each household was included in four waves of the survey over a period of 15 months. This rotation system meant that half of the sample remained unchanged in two consecutive quarters and in quarters one year apart, and 25% of the sample remained unchanged over three quarters.

All the following statistical analyses are made on the so-called longitudinal population. The CLFS is not designed as a proper panel: the initial population changes during the observation period due to demographic events and migrations. Although ISTAT has proposed a procedure to calculate longitudinal weights (Boschetto, Discenza, Lucarelli, Rosati and Fiori 2009), they are not available to researchers, so that we could not take into account the complex sample design. However, we consider that it was reasonable to assume that respondents belonging to the same households were independent.

Information on labour force condition in one reference quarter was collected three times: (i) each respondent was classified as employed, unemployed or not in the labour force according to the definition of the ILO on the basis of answers given to a selected group of questions; (ii) in a subsequent section of the questionnaire, all respondents were asked to classify themselves in the labour market, in order to collect the “self-perceived” condition; (iii) after one year, a retrospective question asked about respondents’ state in the labour market one year before the first interview.

According to the ILO definition, respondents were classified as employed in the reference quarter if, aged 15 years or over, during the reference week they performed some kind of work, for at least one hour, for pay, profit or family gain, or were not at work but had a job or business from which they were temporarily absent because of illness, holidays, industrial dispute, or education and training. Respondents were classified as unemployed if, aged from 15 to 74, they were: (a) without work during the reference week; (b) currently available for work in the two weeks following the reference week; (c) actively seeking work, i.e., had taken specific steps, in the four-week period ending with the reference week, to seek work or who did not seek work but who had found a job to be started later, within a period of up to three months (International Labour Organization (ILO) 2008).

Current self-perception and the retrospective question classified respondents in eight categories: employed; unemployed looking for new employment; unemployed looking for first employment; fulfilling domestic tasks; student; retired; disabled for work; other.

Table 3.1 shows the rotating design of the survey for two consecutive calendar years. Letters identify rotation groups: four rotation groups were interviewed in each quarter. With reference to one calendar year, information on labour market condition came from nine rotation groups. However, the rotation design generates a specific pattern of missing data. For example, for units of rotation group A who are interviewed for the fourth time in the first quarter of year 1, only the ILO (I) indicator and self-perception (S) of labour market condition in the first quarter of year 1 are available. For units in rotation group F, who were first interviewed in the first quarter of year 1, we only have information on labour force state based on the ILO definition, self-perception and the retrospective question (R) for the first and second quarters of year 1.

Table 3.1
CLFS rotation design

Rotation Group	Year 1				Year 2			
	I quarter	II quarter	III quarter	IV quarter	I quarter	II quarter	III quarter	IV quarter
A	I-S							
B	I-S	I-S						
C		I-S	I-S					
D			I-S	I-S				
E	I-S-R			I-S	I-S			
F	I-S-R	I-S-R			I-S	I-S		
G		I-S-R	I-S-R			I-S	I-S	
H			I-S-R	I-S-R			I-S	I-S
I				I-S-R	I-S-R			I-S
L					I-S-R	I-S-R		
M						I-S-R	I-S-R	
N							I-S-R	I-S-R
O								I-S-R

I = ILO indicator, S = self-perception of labour market condition, R = retrospective indicator.

We examined data collected from 2005 to 2010. (Excluded from these analyses are data collected in 2004, the first year of implementation of the new labour force survey, because the data may not be totally reliable; with reference to 2010, here we use only information collected with the retrospective question and referring to labour condition in 2009.) Table 3.2 lists labour market composition in the first quarter from pooled data over the five-year period. The ILO indicator clearly counts a lower percentage of unemployed and a higher percentage of persons not in the labour force than the other two indicators. The two measures based on self-perception give a higher unemployment rate because ILO applies a very strict definition of unemployment. To be classified as unemployed, respondents between the ages of 15 and 74 must not be in employment at the moment of the interview but would accept suitable jobs in the next two weeks if the opportunity arose, and had actively looked for ways of obtaining jobs in the preceding two weeks. ILO provides these guidelines in order to facilitate comparisons of labour market performance over time and across countries (ILO 2008). However, this framework was set up when the prevailing type of employment was full-time and under permanent contract; since then, the employment situation has changed to one of more flexibility, with more part-time and fixed-term types of work, especially for those about to enter the labour market.

Table 3.2
Labour market composition 2005 – 2009 I quarter, % - pooled data

	E	U	N
ILO	43.07	3.60	53.33
S	41.73	6.73	51.54
R	41.55	6.49	51.96

E = Employed, U = Unemployed, N = Not in the Labour Force.

Other studies in the literature show that the distinction between labour market states is not always clear-cut: people may not know official definitions or perceive their labour condition as different from that arising from standard criteria (see, for example, Clark and Summer 1979; Flinn and Heckman 1983; Gonul 1992). In most cases, it is difficult to distinguish between unemployment and not in the labour force: the most critical condition seems to be that of actively seeking a job, since respondents may perceive themselves as unemployed even when they are not actively looking for a job. Inconsistencies may consequently arise between information collected in surveys and effective behaviour. Another explanation of the differences between the ILO and the self-perceived classifications is that respondents with temporary jobs in terms of hours of work per week may not classify themselves as employed.

Table 3.3 lists inconsistencies, i.e., different labour conditions observed for the same respondent with two indicators, among the three indicators for the period in question. Data over quarters and years were pooled for reasons of space. The number of inconsistencies is clearly higher for the state of unemployment than for the other two states, and most of the misclassifications tend to refer to people out of the labour force rather than in employment, as many previous studies show (see, for example, Poterba and Summers 1986). Comparing the labour condition according to the ILO definition with that reported according to answers to the retrospective question generated the highest number of inconsistencies. Examining consistencies over quarters and years for couples of the three indicators (not reported here for reasons of space) we note that consistency tends to increase slightly over time, perhaps because all the actors involved in the survey process - interviewers, respondents, etc. - learn how to collect and supply good-quality information while participating in the survey. Although we did not observe seasonal effects in the number of inconsistencies, the number of inconsistencies indicated non-negligible measurement error in the data, which means that one of the two indicators, or both, were reported incorrectly.

Table 3.3
Inconsistencies 2005 – 2009, % - pooled data

	EU	EN	UE	UN	NE	NU
ILO – Self-perception	0.97	1.72	0.44	13.02	0.17	5.80
ILO – Retrospective	1.14	2.06	5.22	16.76	1.00	5.76
Self-perception – Retrospective	0.92	1.62	6.03	8.73	1.00	0.89

EU = Classified as Employed with first indicator but Unemployed with second indicator.
EN = Classified as Employed with first indicator but Not in the Labour Force with second indicator.
UE = Classified as Unemployed with first indicator but as Employed with second indicator.
UN = Classified as Unemployed with first indicator but Not in the Labour Force with second indicator.
NE = Classified as Not in the Labour Force with first indicator but Employed with second indicator.
NU = Classified as Not in the Labour Force with first indicator but Unemployed with second indicator.

However, the inconsistencies emerging from Tables 3.2 and 3.3 may also occur because all three indicators are exposed to measurement error. Previous studies have investigated the causes of labour condition misperception, finding that it is influenced by social, demographic, economic and institutional factors (e.g., Richiardi 2002). Inconsistencies between the two self-perceptions (actual and retrospective) may mainly be due to memory decay (Bound, Brown and Mathiowetz 2001). Lastly, the higher consistency between the self-perception indicators suggests the possibility of correlated measurement errors.

Table 3.4 lists observed quarterly transition probabilities among the three labour force conditions from the first to the second quarter of the years from 2005 to 2009 with the three indicators. The ILO indicator describes a much more dynamic labour market, especially for unemployed respondents, than that described by the self-perceived and retrospective indicators. This difference is another piece of evidence revealing measurement error in the data. From the existing literature, we know that even small degrees of classification error may lead to severe bias in the estimation of transition probabilities (Hagenaars 1994; Pavlopoulos, Muffles and Vermunt 2012). If errors are uncorrelated over time, we can expect to observe a more dynamic labour market than the true one, and the opposite if error correlation over time also exists.

Table 3.5 compares observed gross flows, as an example from the first to the second quarter of 2005, by gender and age. The three age intervals were obtained by dividing the samples into three groups, with equal dimensions (i.e., 33rd and 66th percentiles). In detail, for the year 2005, in age 1 we find respondents aged between 16 and 36; in age 2 they are between 36 and 55, and in age 3 between 56 and 75. The evidence is that women are more dynamic, especially with regard to unemployment, than men. When leaving unemployment, women tend to leave the labour market more often than to become employed. There are also some important differences in observed gross flows across ages. The older respondents were more stable when out of the labour market and had higher probabilities of moving out of the labour market than of becoming unemployed after being employed. Younger respondents have lower probabilities than those in the second age-group of leaving unemployment and the condition of not being in the labour market by finding jobs. This evidence suggests that gender and age should be included as covariates in our model, to estimate corrected gross flows in the labour market.

Table 3.4
Observed gross flows I quarter to II quarter 2005 - 2009, %, International Labour Office (ILO), Self-perceived (S) and Retrospective (R) indicators

		EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	ILO	96.49	0.87	2.63	18.97	50.50	30.53	1.49	1.99	96.52
	S	96.99	1.33	1.69	15.32	69.85	14.83	1.29	1.50	97.21
	R	95.32	2.10	2.58	20.96	59.56	19.48	1.96	2.22	95.81
2006	ILO	96.13	0.78	3.09	20.40	45.21	34.39	2.42	1.74	95.84
	S	96.11	1.74	2.16	19.84	63.66	16.50	1.88	1.75	96.37
	R	95.55	1.72	2.73	17.93	66.57	15.50	2.00	1.75	96.25
2007	ILO	96.22	0.68	3.10	21.45	40.41	38.14	2.21	1.78	96.02
	S	96.08	1.74	2.16	19.84	63.66	16.50	1.88	1.75	96.37
	R	95.66	1.78	2.56	19.95	60.67	19.38	2.26	1.93	95.80
2008	ILO	97.05	0.80	2.16	19.82	48.50	31.68	1.87	1.87	96.26
	S	96.92	1.54	1.53	15.25	70.84	13.92	1.56	1.69	96.75
	R	95.76	2.13	2.11	19.04	62.60	18.36	2.02	2.26	95.72
2009	ILO	96.58	0.88	2.54	18.41	48.10	33.49	2.08	1.83	96.09
	S	96.14	1.76	2.10	15.17	70.09	14.75	1.59	1.61	96.80
	R	95.45	1.88	2.66	16.88	67.15	15.97	1.78	1.89	96.33

EE = Employed in both quarters.
EU = Employed in first quarter and Unemployed in second one.
EN = Employed in first quarter and Not in the Labour Force in second one.
UE = Unemployed in first quarter and Employed in second one.
UU = Unemployed in both quarters.
UN = Unemployed in first quarter and Not in the Labour Force in second one.
NE = Not in the Labour Force in first quarter and Employed in second one.
NU = Not in the Labour Force in first quarter and Unemployed in second one.
NN = Not in the Labour Force in both quarters.

Table 3.5
Observed gross flows I quarter to II quarter 2005, by gender and age, %, International Labour Office (ILO), Self-perceived (S) and Retrospective (R) indicators

		EE	EU	EN	UE	UU	UN	NE	NU	NN
Males	ILO	97.20	0.78	2.02	22.73	51.60	25.68	1.93	2.07	96.00
	S	97.63	1.08	1.29	18.97	73.80	7.23	1.36	1.10	97.53
	R	96.13	1.84	2.03	26.14	65.27	8.60	2.13	1.50	96.37
Females	ILO	95.43	1.01	3.57	15.70	49.31	34.99	1.23	1.98	96.79
	S	96.00	1.69	2.31	11.93	65.73	22.34	1.26	1.81	96.93
	R	94.14	2.46	3.40	16.24	53.56	30.19	1.86	2.71	95.43
Age 1	ILO	88.27	0.46	11.27	21.16	27.50	51.35	0.26	0.06	99.67
	S	89.66	0.56	9.78	10.20	60.09	29.71	0.31	0.10	99.60
	R	83.36	0.45	16.19	20.78	42.54	36.68	0.51	0.13	99.36
Age 2	ILO	97.65	0.55	1.80	21.62	43.01	35.37	2.72	2.95	94.33
	S	97.87	0.92	1.20	16.83	64.65	18.52	2.52	2.61	94.87
	R	97.04	1.23	1.74	24.60	53.42	21.98	4.05	4.24	91.70
Age 3	ILO	96.18	1.32	2.50	17.54	51.14	31.32	3.81	6.75	89.44
	S	96.83	1.89	1.28	14.77	71.97	13.27	3.17	4.82	92.01
	R	94.82	3.29	1.89	19.62	63.60	16.78	4.52	6.68	88.80

EE = Employed in both quarters.
EU = Employed in first quarter and Unemployed in second one.
EN = Employed in first quarter and Not in the Labour Force in second one.
UE = Unemployed in first quarter and Employed in second one.
UU = Unemployed in both quarters.
UN = Unemployed in first quarter and Not in the Labour Force in second one.
NE = Not in the Labour Force in first quarter and Employed in second one.
NU = Not in the Labour Force in first quarter and Unemployed in second one.
NN = Not in the Labour Force in both quarters.

4 Results: Comparisons of mixed and standard LCM models

We estimate various specifications of the standard and mixed LCM models. The standard model consists of two parts: structural, describing true dynamics among latent variables (true states) by a first-order Markov chain; and measurement, which links each latent variable to its indicators (observed conditions in the labour market). Some restrictions incorporating a priori information and/or assumptions are imposed on the parameters of the measurement part, based on evidence from observed data (inconsistencies and transitions) and on findings from the survey methodology and cognitive psychology literature on the error - generating mechanism. Only four of the nine rotation groups supplying information referring to one calendar year were interviewed in every quarter, and only for two of these groups do we have all three indicators of labour market conditions (see Table 3.1). For the other two groups, we do not have the information collected with the retrospective question. The pattern of missing information due to the rotation design of the survey is included in the estimated LCM models as data missing at random.

All estimated models share the following characteristics: true transitions follow a first-order Markov chain; (Due to the survey design, there were no individuals observed for three consecutive waves, i.e., a second-order Markov chain cannot be estimated, since the relative sufficient statistics are missing. However, although the labour market condition in one quarter may very plausibly affect the condition in the subsequent quarter, that it may do so in a significant manner after two quarters is far less plausible.) classification errors are assumed constant over time for each indicator; the ICE assumption is included.

Model fit is evaluated by the BIC (Bayesian Information Criterion) index because of the large sample size (average 250,000 units per year; see Table 4.1).

The specification of a mixed LCM model is also recommended by the fact that the sample may contain various groups of respondents with different behaviour in the labour market. As already noted, the recent literature shows that not taking unobserved heterogeneity in transitions into account when estimating LCM models may result in biased estimates of measurement error (Magidson et al. 2007). In addition, a mixed LCM model may give the data a better fit.

We estimate a mover-stayer LCM model with the assumption of constant measurement errors across the two latent groups. It should be noted that all estimated models were identified and that, in order to reduce the risk of detecting local maxima, estimation was performed several times with different sets of starting values. Latent Gold 4.5 software was implemented (Vermunt and Magidson 2008).

Table 4.1 compares the mixed and standard LCM models fitted to our five data samples, referring to the years from 2005 and 2009 and using the BIC index. The mixed model shows a better fit for all samples. Table 4.2 lists the percentages of movers and stayers in the first quarter of 2005, and the distribution of the two unobserved groups in the first quarter of each year. Clearly, unobserved heterogeneity is highly correlated with the initial state and, as expected, stayers are either employed or not in the labour market, i.e., only a very small percentage is unemployed.

Table 4.1
Comparison of standard and mixed LCM models: BIC index

Year	n	Standard	Mixture
2005	220,051	650,241	649,401
2006	206,037	587,794	587,058
2007	274,484	748,788	748,654
2008	277,363	667,399	666,335
2009	274,723	747,997	746,991

Table 4.2
Mixed LCM model: proportion of movers and stayers and distribution in initial state 2005, I quarter, %

	Proportion	E	U	N
Movers	10.23	39.85	39.09	21.06
Stayers	81.79	41.79	3.36	54.85

E = Employed, U = Unemployed, N = Not in the Labour Force.

As the data in Tables 4.3-4.5 show, (Labour market composition, estimated transitions and estimated measurement errors show the same pattern in the other three quarters of each year.) the better fit to the data of the mixed model is all due to the different estimated transition rates; labour market composition and estimated measurement errors are the same in both models. This result is the opposite of that obtained by Magidson et al. (2007), who compared the mover-stayer and standard LCM models applied to labour market transitions from the Current Population Survey. The above authors found that the mixed LCM model provides a better fit to the data than the standard LCM model and that the latter, not taking unobserved heterogeneity into account, overestimates the degree of measurement error with respect to the mover-stayer model. In detail, the above authors used simulated results to estimate a violation of homogeneous transition probabilities, so that heterogeneity correlated with the initial state produces inflated estimates of measurement errors in a standard LCM model.

Table 4.3
Comparison of standard and mixed LCM models: labour market composition I quarter 2005, %

		E	U	N
2005	Standard	41.67	7.00	51.33
	Mixture	41.59	7.02	51.39

E = Employed, U = Unemployed, N = Not in the Labour Force.

Table 4.4
Comparison of standard and mixed LCM models: estimated transitions I quarter to II quarter 2005 – 2009, %

		EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	Standard	97.36	1.32	1.32	15.59	76.18	8.23	0.57	0.74	98.69
	Mixture	96.46	1.68	1.86	19.61	69.65	10.74	0.91	1.09	98.00
2006	Standard	96.75	1.68	1.56	19.52	71.27	9.21	1.01	0.99	90.00
	Mixture	96.22	1.92	1.87	22.11	66.96	10.93	1.25	1.22	97.54
2007	Standard	96.69	1.67	1.64	18.84	70.56	10.60	1.01	0.99	98.00
	Mixture	96.42	1.80	1.78	20.22	67.80	11.98	1.10	1.45	95.45
2008	Standard	97.56	1.41	1.03	15.86	79.73	4.42	0.53	0.62	98.85
	Mixture	96.45	1.89	1.66	19.56	73.25	7.19	0.83	0.89	98.28
2009	Standard	96.85	1.71	1.44	14.04	75.33	9.63	1.04	1.01	97.95
	Mixture	96.27	1.95	1.78	17.09	71.16	11.75	1.30	1.22	97.48

EE = Employed in both quarters.
EU = Employed in first quarter and Unemployed in second one.
EN = Employed in first quarter and Not in the Labour Force in second one.
UE = Unemployed in first quarter and Employed in second one.
UU = Unemployed in both quarters.
UN = Unemployed in first quarter and Not in the Labour Force in second one.
NE = Not in the Labour Force in first quarter and Employed in second one.
NU = Not in the Labour Force in first quarter and Unemployed in second one.
NN = Not in the Labour Force in both quarters.

Table 4.5
Comparison of Standard and mixed LCM models: estimated measurement errors I quarter 2005 – 2009, %, ILO indicator

		EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	Standard	99.82	0.01	0.17	6.17	45.04	48.80	0.89	0.50	98.61
	Mixture	99.82	0.01	0.17	6.16	45.06	48.78	0.90	0.51	98.59
2006	Standard	99.83	0.01	0.16	6.50	41.92	51.58	0.75	0.45	98.80
	Mixture	99.87	0.01	0.13	5.17	37.28	57.55	0.68	0.40	98.92
2007	Standard	99.75	0.01	0.24	6.84	39.83	53.34	0.75	0.47	98.79
	Mixture	99.75	0.01	0.24	6.77	39.92	53.31	0.77	0.47	98.76
2008	Standard	99.83	0.01	0.17	3.81	42.45	53.74	0.61	0.38	99.02
	Mixture	99.83	0.01	0.17	3.82	42.41	53.76	0.62	0.38	99.00
2009	Standard	95.34	0.98	3.68	18.30	41.17	40.53	2.06	1.61	96.33
	Mixture	95.22	2.34	2.44	15.60	68.02	16.37	1.74	2.14	96.13

EE = Truly Employed and classified as Employed by ILO indicator.
EU = Truly Employed but classified as Unemployed by ILO indicator.
EN = Truly Employed but classified as Not in the Labour Force by ILO indicator.
UE = Truly Unemployed but classified as Employed by ILO indicator.
UU = Truly Unemployed and classified as Unemployed by ILO indicator.
UN = Truly Unemployed but classified as Not in the Labour Force by ILO indicator.
NE = Truly Not in the Labour Force but classified as Employed by ILO indicator.
NU = Truly Not in the Labour Force but classified as Unemployed by ILO indicator.
NN = Truly Not in the Labour Force and classified as Not in the Labour Force by ILO indicator.

The mover-stayer model describes a more dynamic labour market, especially for unemployed respondents: the probability of remaining unemployed over the quarter is lower than that estimated by the standard model.

5 Results: Mixed LCM model with covariates and correlated measurement errors

The results shown in the previous section showed that a mixed LCM model gives a better fit to our data. Like the standard LCM model, it takes into account misclassification and the pattern of missing data assuming the latter at random, and also includes unobserved heterogeneity. Assuming that data are missing at random is explained by the fact that each rotation group is observed in two quarters, but not in the two subsequent quarters, and also that these data are missing by design and do not depend on respondents' true or reported status or other unobserved variables. In estimating our models, we simultaneously used information from all rotation groups, i.e., a Full Information Maximum Likelihood approach. Evidence from the observed gross flows, especially the fact that observed mobility is quite different between men and women and across ages (Table 3.5) indicated estimating a mixed LCM model with these two covariates affecting latent transitions.

Various models were estimated with the common following characteristic: mover-stayer and latent transitions follow a first-order Markov chain. In order to specify the measurement model, the following considerations were made: (i) the answer to the question on self-perceived condition in the labour market is given in the same interview after respondents answer the questions on which the ILO indicator is based; (ii) however, the ILO indicator is determined by ISTAT according to answers given to a series of questions following ILO guidelines, whereas S represents respondents' self-perceptions: it is plausible that respondents are not aware of the ISTAT classification; (iii) indicator S and the indicator resulting from retrospective interrogation describe a more stable labour market than that of ILO and show the highest level of consistency: respondents may be influenced by the answers they gave the previous quarter; (iv) information for R is collected one year after answers to ILO and S; (v) for individuals who are in a steady state, reporting labour force condition correctly is an easier cognitive task than for those who experience at least one change, and may consequently show higher probabilities of giving incorrect answers.

Among the various possible specifications, the best-fitting model, for all analysed years, was to assume that stayers report their labour market condition correctly and that, for movers, measurement errors are constant over time and that the two indicators based on self-perception, S and R, are correlated, i.e., a direct effect between these two indicators is inserted in the model specification. (All estimated models were identified and, in order to avoid local maxima, estimation was performed several times with different sets of starting values; to estimate more parsimonious models, all three variable interactions were set at 0.) As an example, Tables 5.1 to 5.3 list some of the estimation results: labour market composition and estimated flows for the overall population, movers and stayers together, (The complete set of estimation results is available from the authors.) and estimated measurement errors. On average, over the five years, the percentage of movers was 17.69.

Table 5.1
Estimated labour market composition I quarter 2005 – 2009, %

	2005	2006	2007	2008	2009
E	42.01	42.36	40.72	40.92	40.00
U	5.93	5.64	5.75	5.27	6.46
N	52.07	52.00	53.53	53.81	53.53

E = Employed, U = Unemployed, N = Not in the Labour Force.

Table 5.2
Estimated gross flows I quarter to II quarter 2005 - 2009, %, standard errors in brackets

	EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	96.70	1.60	1.61	17.41	71.80	10.78	0.97	0.70	98.29
	(0.0017)	(0.0012)	(0.0012)	(0.0133)	(0.0142)	(0.0079)	(0.0013)	(0.0011)	(0.0017)
2006	96.10	1.93	1.93	19.16	67.04	13.80	1.71	0.89	97.41
	(0.0027)	(0.0020)	(0.0020)	(0.0112)	(0.0150)	(0.0136)	(0.0011)	(0.0015)	(0.0018)
2007	96.30	1.79	1.89	18.11	67.95	13.94	1.42	1.24	97.34
	(0.0023)	(0.0016)	(0.0017)	(0.0145)	(0.0158)	(0.0094)	(0.0018)	(0.0018)	(0.0025)
2008	96.88	1.77	1.35	18.00	74.57	7.43	1.61	1.03	97.37
	(0.0037)	(0.0027)	(0.0028)	(0.0118)	(0.0157)	(0.0138)	(0.0013)	(0.0017)	(0.0020)
2009	96.50	1.83	1.62	15.04	71.62	13.35	1.55	1.10	97.35
	(0.0024)	(0.0019)	(0.0016)	(0.0153)	(0.0168)	(0.0092)	(0.0019)	(0.0014)	(0.0024)

EE = Employed in both quarters.
EU = Employed in first quarter and Unemployed in second one.
EN = Employed in first quarter and Not in the Labour Force in second one.
UE = Unemployed in first quarter and Employed in second one.
UU = Unemployed in both quarters.
UN = Unemployed in first quarter and Not in the Labour Force in second one.
NE = Not in the Labour Force in first quarter and Employed in second one.
NU = Not in the Labour Force in first quarter and Unemployed in second one.
NN = Not in the Labour Force in both quarters.

Table 5.3a
Estimated measurement errors 2005 – 2009 ILO indicator, %, standard errors in brackets

	EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	99.75	0.02	0.23	0.93	89.72	9.36	0.97	1.04	98.00
	(0.0002)	(0.0001)	(0.0001)	(0.0028)	(0.0050)	(0.0051)	(0.0004)	(0.0003)	(0.0005)
2006	99.75	0.01	0.24	1.17	89.39	9.44	0.55	0.99	98.46
	(0.0007)	(0.0004)	(0.0005)	(0.0025)	(0.0042)	(0.0035)	(0.0003)	(0.0002)	(0.0004)
2007	99.82	0.01	0.24	0.84	88.28	10.88	0.58	0.87	98.55
	(0.0002)	(0.0001)	(0.0002)	(0.0028)	(0.0050)	(0.0051)	(0.0004)	(0.0003)	(0.0005)
2008	99.44	0.10	0.46	1.16	89.36	9.48	0.57	1.38	90.05
	(0.0007)	(0.0004)	(0.0005)	(0.0025)	(0.0042)	(0.0035)	(0.0003)	(0.0002)	(0.0004)
2009	99.77	0.01	0.22	0.43	88.98	10.57	0.33	0.86	98.79
	(0.0001)	(0.0000)	(0.0001)	(0.0025)	(0.0038)	(0.0039)	(0.0003)	(0.0002)	(0.0003)

EE = Truly Employed and classified as Employed by ILO indicator.
EU = Truly Employed but classified as Unemployed by ILO indicator.
EN = Truly Employed but classified as Not in the Labour Force by ILO indicator.
UE = Truly Unemployed but classified as Employed by ILO indicator.
UU = Truly Unemployed and classified as Unemployed by ILO indicator.
UN = Truly Unemployed but classified as Not in the Labour Force by ILO indicator.
NE = Truly Not in the Labour force but classified as Employed by ILO indicator.
NU = Truly Not in the Labour Force but classified as Unemployed by ILO indicator.
NN = Truly Not in the Labour Force and classified as Not in the Labour Force by ILO indicator.

Table 5.3b
Estimated measurement errors 2005 – 2009 S and R indicators, %, standard errors in brackets

	True state					SR				
		EE	EU	EN	UE	UU	UN	NE	NU	NN
2005	E	94.83	1.17	2.28	0.22	0.18	0.11	0.44	0.07	0.70
		(0.0008)	(0.0006)	(0.0005)	(0.0002)	(0.0001)	(0.0002)	(0.0003)	(0.0004)	(0.0003)
	U	0.01	0.00	0.00	0.97	97.16	1.11	0.09	0.31	0.35
		(0.0001)	(0.0001)		(0.0006)	(0.0008)	(0.0004)	(0.0009)	(0.0004)	(0.0003)
	N	0.00	0.00	0.01	0.12	0.70	0.70	0.78	0.98	96.72
				(0.0001)	(0.0005)	(0.0009)	(0.0008)	(0.0004)	(0.0006)	(0.0008)
2006	E	94.86	0.96	2.21	0.16	0.11	0.10	0.45	0.06	1.06
		(0.0052)	(0.0006)	(0.0005)	(0.0001)	(0.0002)	(0.0009)	(0.0001)	(0.0004)	(0.0003)
	U	0.00	0.01	0.00	0.86	97.98	0.50	0.11	0.32	0.22
			(0.0001)		(0.0001)	(0.0006)	(0.0001)	(0.0002)	(0.0003)	(0.0003)
	N	0.01	0.00	0.01	0.13	0.82	0.74	0.71	0.74	96.83
		(0.0001)		(0.0001)	(0.0006)	(0.0005)	(0.0004)	(0.0004)	(0.0001)	(0.0005)
2007	E	95.17	1.06	1.06	0.16	0.11	0.10	0.45	0.06	0.82
		(0.0009)	(0.0003)	(0.0005)	(0.0002)	(0.0004)	(0.0005)	(0.0006)	(0.0004)	(0.0004)
	U	0.00	0.01	0.00	0.90	97.74	0.73	0.09	0.31	0.21
			(0.0001)		(0.0005)	(0.0009)	(0.0003)	(0.0005)	(0.0004)	(0.0002)
	N	0.01	0.01	0.01	0.15	0.59	0.66	1.10	0.89	96.59
		(0.0001)	(0.0001)	(0.0001)	(0.0005)	(0.0006)	(0.0008)	(0.0004)	(0.0004)	(0.0020)
2008	E	94.65	1.48	1.83	0.16	0.02	0.14	0.72	0.04	0.96
		(0.0006)	(0.0009)	(0.0005)	(0.0003)	(0.0006)	(0.0004)	(0.0003)	(0.0004)	(0.0002)
	U	0.00	0.03	0.00	1.32	97.39	0.82	0.05	0.33	0.05
			(0.0001)		(0.0002)	(0.0010)	(0.0009)	(0.0005)	(0.0004)	(0.0004)
	N	0.01	0.02	0.01	0.17	0.45	1.34	1.05	1.50	95.45
		(0.0001)	(0.0001)	(0.0001)	(0.0009)	(0.0005)	(0.0003)	(0.0006)	(0.0004)	(0.0003)
2009	E	96.11	0.65	1.21	0.12	0.24	0.10	0.42	0.10	1.04
		(0.0004)	(0.0002)	(0.0001)	(0.0002)	(0.0003)	(0.0008)	(0.0009)	(0.0008)	(0.0009)
	U	0.01	0.01	0.00	0.59	98.23	0.55	0.08	0.26	0.25
		(0.0001)	(0.0001)		(0.0004)	(0.0004)	(0.0002)	(0.0005)	(0.0006)	(0.0006)
	N	0.01	0.00	0.01	0.08	0.76	0.52	0.74	0.78	97.08
		(0.0001)		(0.0001)	(0.0004)	(0.0002)	(0.0002)	(0.0004)	(0.0003)	(0.0008)

E = Employed, U = Unemployed, N = Not in the Labour Force.
EE = Classified as Employed by Self-perceived and Retrospective indicators.
EU = Classified as Employed by Self-perceived indicator and Unemployed by Retrospective indicator.
EN = Classified as Employed by Self-perceived indicator and Not in the Labour Force by Retrospective indicator.
UE = Classified as Unemployed by Self-perceived indicator and Employed by Retrospective indicator.
UU = Classified as Unemployed by Self-perceived and Retrospective indicators.
UN = Classified as Unemployed by Self-perceived indicator and Non in the Labour Force by Retrospective indicator.
NE = Classified as Not in the Labour Force by Self-perceived indicator and Employed by Retrospective indicator.
NU = Classified as Not in the Labour Force by Self-perceived indicator and Unemployed by Retrospective indicator.
NN = Classified as Not in the Labour Force by Self-perceived and Retrospective indicators.

The estimated labour market composition in the first quarter, compared with the observed one (Table 3.2), shows a percentage of unemployment slightly lower than that obtained with the two self-perception indicators and higher than that with the ILO indicator.

Estimated transitions describe a more stable labour market than that observed with all three indicators, with the only exception of two transitions (see Table 3.4). Estimated gross flows are much more similar to those observed with self-perception and retrospective questions than those observed with the ILO indicator. This evidence also appears from the estimated measurement error (Table 5.3). An immediate objection to this result would be that we used two very similar indicators (the two self-perceptions) and a third one which was quite different (ILO). In fact, a similar result - lower measurement errors for self-perception than for the ILO indicator - was obtained by estimating an LCM model with only two indicators per latent variable: ILO and self-perception.

6 Concluding remarks

This paper presents a latent class approach to correct gross flows from correlated errors. The emphasis is on the capacity to account for correlated classification errors across panel data, due to the rotating design of the survey which generates patterns of missing data and of unobserved heterogeneity.

The latent class approach was applied to transitions in the Italian labour market among the three usual conditions of employed, unemployed and not in the labour force. The data refer to the years from 2005 to 2009 and were collected by the Continuous Italian Labour Force Survey on a sample of Italian households with a 2-2-2 rotating design over quarters. Information on labour force condition in one reference quarter was collected three times: (i) respondents were classified as employed, unemployed or not in the labour force according to the definition of the International Labour Office on the basis of answers to a selected group of questions; (ii) respondents were asked to classify themselves as employed, unemployed or not in the labour force (i.e., the self-perceived condition); (iii) a retrospective question asked about state in the labour market one year previously. This means that three indicators of labour condition were available. The three indicators gave quite different descriptions of the Italian labour market, revealing a significant degree of inconsistency. This evidence indicates measurement error in the data.

The best-fitting model was a mover-stayer LCM, in which latent transitions in the labour market follow a first-order Markov chain, stayers always report their market condition correctly; for movers, measurement errors were constant over time and correlated to the two self-perception indicators; the gender and age of respondents were included as covariates; the rotating design of the survey was treated as information missing at random. The model corrects observed gross flows towards a more stable labour market and estimates that the indicator of labour market condition based on the ILO definition is affected by the greatest degree of measurement error.

A second result found here is that, when unobserved heterogeneity occurs, a mixed LCM model fits the data better than the standard LCM model. This finding is consistent with other reports (e.g., Magidson et al. 2007). However, in our case, the two models estimate the same quantity of measurement error, the difference in fit being due to estimated flows. Instead, the above authors found an overestimation of measurement error when unobserved heterogeneity was not taken into account.

A final consideration regards the sample design of the survey, which is two-stage, as described in Section 3. In our analyses, we did not take into account the complex sample design, but estimated gross flows on the longitudinal population provided by the Italian Institute of Statistics. In subsequent research, it will be of interest to compare how results may be affected by incorporating methods for surveys on complex samples with our estimation strategy, an interesting reference to which was made by Lu and Lohr (2010).

References

Bartolucci, F., Lupparelli, M. and Montanari, A. (2009). Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *Annals of Applied Statistics,* 3, 611-636.

Bassi, F., and Trivellato, U. (2009). A latent class approach for estimating gross flows in the presence of correlated classification errors. In *Methodology of Longitudinal Surveys*, (Ed., P. Lynn), Chichester: Wiley, 367-380.

Bassi, F., Padoan, A. and Trivellato, U. (2012). Inconsistencies in reported characteristics among employed stayers. *Statistica*, 1, 93-109.

Bassi, F., Torelli, N. and Trivellato, U. (1998). Data and modelling strategies in estimating labour force gross flows affected by classification errors. *Survey Methodology*, 24, 2, 109-122. Paper available at http://www.statcan.gc.ca/pub/12-001-x/1998002/article/4348-eng.pdf.

Bassi, F., Croon, M., Hagenaars, J.A. and Vermunt, J.K. (2000). Estimating true changes when categorical panel data are affected by correlated and uncorrelated classification errors. An application to unemployment data. *Sociological Methods and Research*, 29, 230-268.

Biemer, P.P., and Bushery, J.M. (2000). On the validity of Markov latent class analysis for estimating classification error in labor force data. *Survey Methodology*, 26, 2, 139-152. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2000002/article/5534-eng.pdf.

Boschetto, B., Discenza, A.R., Lucarelli, C., Rosati, S. and Fiori, F. (2009). Longitudinal data for the analysis of Italian labor market flows. *Italian Journal of Applied Statistics*, 22, 129-150.

Bound, M., Brown, C. and Mathiowetz, N.A. (2001). Measurement error in survey data. In *Handbook of Econometrics*, (Eds., J.J. Heckman and E. Leamer), Amsterdam: Elsevier, 3705-3843.

Clark, K., and Summers, L.H. (1979). Labour market dynamics and unemployment: A reconsideration. *Brooking Papers on Economic Activity*, 1, 13-69.

Dayton, C.M., and McReady, G.B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173-178.

Flinn, C.J., and Heckman, J.J. (1983). Are unemployment and out of the labour force behaviourally distinct market states? *Journal of Labour Economics*, 1, 28-42.

Gonul, F. (1992). New evidence on whether unemployment and out of the labour force are two distinct states. *Journal of Human Resources*, 27, 329-361.

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

Hagenaars, J.A. (1994). Latent variables in log-linear models of repeated observations. In *Latent Variable Analysis. Applications for Developmental Research*, (Eds., A. von Eye and C. Clogg), Thousand Oaks (CA): Sage, 329-352.

International Labour Organization (ILO) (2008). Resolution concerning updating the International Standard Classification of Occupation, Geneva. Paper available at http://www.ilo.org/public/english/bureau/stat/isco/docs/resol08.pdf.

Langeheine, R. (1994). Latent variable Markov models. In *Latent Variable Analysis. Applications for Developmental Research*, (Eds., A. von Eye and C. Clogg), Thousand Oaks (CA): Sage, 373-395.

Lu, Y., and Lohr, S. (2010). Gross flow estimation in dual frame surveys. *Survey Methodology*, 36, 1, 13-22. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2010001/article/11248-eng.pdf.

Magidson, J., Vermunt, J.K. and Tran B. (2007). Using a mixture of latent Markov model to analyze longitudinal U.S. employment data involving measurement error. In *New Trends in Psychometrics*, (Eds., K. Shigemasu, A. Okada, T. Imaizumi and T. Hoshino), Tokyo: Universal Academy Press, 235-242.

Manzoni, A., Vermunt, J.K., Luijkx, R. and Muffels, R. (2010). Memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement errors. *Sociological Methodology*, 40, 39-73.

Paas, L.J., Vermunt, J.K. and Bijmolt, T.H. (2007). Discrete-time discrete-state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Series A*, 170, 955-974.

Pavlopoulos, D., and Vermunt, J.K. (2015). Measuring temporary employment. Do survey or register data tell the truth? *Survey Methodology*, 41, 1, 197-214. Paper available at http://www.statcan.gc.ca/pub/12-001-x/2015001/article/14151-eng.pdf.

Pavlopoulos, D., Muffles, R. and Vermunt, J.K. (2012). How real is mobility between low pay, high pay and non-employment? *Journal of the Royal Statistical Society, Series A*, 170, 749-773.

Poterba, J.M., and Summers, L.S. (1986). Reporting errors and labour market dynamics. *Econometrica*, 54, 1319-1338.

Richiardi, M. (2002). What does the ECHP tell us about labour status misperception? A journey in less known regions of labour discomfort. *LABORatorio Revelli*, Working paper No. 69.

van de Pol, F., and Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 33, 231-247.

Vermunt, J.K. (1997). *Log-Linear Models for Event History*. Thousand Oaks (CA): Sage.

Vermunt, J.K. (2010). Longitudinal research using mixture models. In *Longitudinal Research with Latent Variables*, (Eds., K. van Montfort, J.H.L. Oud and A. Satorra), Heidelberg: Springer, 119-152.

Vermunt, J.K., and Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., and Magidson, J. (2013). *Technical Guide for Latent Gold 5.0. Basic, Advanced, and Syntax.* Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., Tran, B. and Magidson, J. (2008). Latent class models in longitudinal research. In *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, (Ed., S. Menard), Burlington, MA: Elsevier, 375-385.

Wiggins, L.M. (1973). *Panel Analysis: Latent Probability for Attitude and Behavior Processes*. New York: Elsevier Scientific.