Survey Methodology, December 1998
Vol. 24, No. 2, pp. 109-122
Statistics Canada

Data and Modelling Strategies in Estimating Labour Force Gross Flows Affected by Classification Errors

FRANCESCA BASSI, NICOLA TORELLI and UGO TRIVELLATO[1]

ABSTRACT

Gross flows among labour force states are of great importance in understanding labour market dynamics. Observed flows are typically subject to classification errors, which may induce serious bias. In this paper, some of the most common strategies, used to collect longitudinal information about labour force condition are reviewed, jointly with the modelling approaches developed to correct gross flows, when affected by classification errors. A general framework for estimating gross flows is outlined. Examples are given of different model specifications, applied to data collected with different strategies. Specifically, two cases are considered, *i.e.*, gross flows from (i) the U.S. Survey of Income and Program Participation and (ii) the French Labour Force Survey, a yearly survey collecting retrospective monthly information.

KEY WORDS: Correlated classification errors; Latent class models; Longitudinal data; Recall errors; Seam effect.

1. INTRODUCTION

Gross flows among labour force states, are a powerful tool to analyse labour market dynamics. Gross flows regard changes at individual level, and therefore their estimation rests on the availability of longitudinal data.

The effects of erroneous classification of units with respect to their position in the labour market, can cause spurious transitions. Even if one might assume that these errors cancel out when estimating net flows, they cannot be ignored when estimating gross flows.

Various strategies can be adopted, in order to correct gross flows for classification errors. Basically, they depend on:

(a) assumptions about the classification error mechanism, following from
  (a1) the survey design (panel surveys – possibly with a rotating scheme, retrospective surveys, some mixture of retrospective and panel surveys, *etc.*), and/or;
  (a2) the content and structure of the questionnaire (availa-bility of one or more indicators of the variable of interest, format of the questions – episode based or event based, *etc.*);
(b) assumptions about the generating process of the transitions among labour force states.

In this paper, some of the most common strategies used to collect longitudinal information about labour force condition are reviewed, jointly with modelling approaches developed to correct gross flows when affected by classification errors. It is shown that most of the usual specifications proposed in the literature, can be seen as special cases of a general formulation, which allows to elucidate advantages and disadvantages of each specification, and makes it possible to consider a common estimation strategy.

The focus of the paper is on sound applications of this general modelling approach, for estimating gross flows from survey data collected with different strategies. Two cases are considered: (i) the U.S. Survey of Income and Program Participation and (ii) the French Labour Force Survey, a yearly rotating panel survey with retrospective monthly information.

The organization of the paper is as follows. Section 2 briefly discusses various strategies for collecting longitudinal data on labour force participation, and their likely implications for classification errors, as they emerge from the survey methodology literature. In section 3, a fairly general approach for modelling gross flows affected by classification errors, *i.e.*, for jointly estimating true gross flows and conditional response probabilities, is outlined. Examples are also given on how some well known models for correcting observed gross flows, can be specified as special cases of this approach (section 3.1). Attention is then devoted to a convenient framework for formulating the above models, provided by latent class models and, more specifically, by the so-called "modified LISREL model" proposed by Hagenaars (1990), a general tool to describe causal relationships among observed and unobserved categorical variables (section 3.2).

The final, and main part of the paper (section 4), is devoted to a detailed presentation of the two case-studies. The modelling approach is common: *a priori* information on the measurement characteristics of the survey (and possibly on the true process), is combined with specification searches, in order to obtain parsimonious and (hopefully) sensible models. As already noted, the two case-studies are reasonably different, chiefly in terms of the design of the surveys: this diversification turns out to be useful for illustrating different model specifications, and various strategies for reaching/testing the final formulation.

[1] Francesca Bassi, Nicola Torelli and Ugo Trivellato, Dipartimento di Scienze Statistiche, Via San Francesco, 33, 35121, Padova, Italy.

From the two case-studies, the following overall evidence can be drawn:

(a) the modified LISREL model has proved to be a set-up, flexible enough for modelling the error mechanism in longitudinal data collected with different survey designs, as well as the generating process of true labour force transitions;

(b) specifically, in the measurement part of the model, we were able to incorporate the pattern and the effects of correlated classification errors, which are particularly important in surveys with retrospective features;

(c) observed transitions are corrected towards the direction expected, on the basis of theoretical and empirical evidence on measurement errors effects, (not mechanically towards mobility, as strategies based on the assumption of independent classification errors do).

2. THE ROLE OF DATA COLLECTION STRATEGIES

Information for labour gross flows estimation comes from longitudinal data, *i.e.*, observations on the same units pertaining to different time points. Recently, there have been increasing efforts in collecting longitudinal data. This is true also for surveys, whose main goal is to measure the labour force condition of individuals in a given population. On the other side, this focus on collecting, and using longitudinal data, raised new questions about the origin and pattern of measurement (≡ classification) errors, as well as their possible effects on estimates of the quantities of interest. General references about sources of classification errors for longitudinal data, collected by surveys across time, are Duncan and Kalton (1987) and Kalton and Citro (1993). In this section, some main implications of classification errors on modelling strategies, to correct gross flows are briefly discussed.

A typical argument about the effect of measurement error in estimating gross flows, is that it leads to over-estimation of changes. This is true when one assumes that measurement errors are not correlated over time. This assumption is not realistic in many cases (see Skinner and Torelli 1993; Singh and Rao 1995; van de Pol and Langeheine 1997), and should be reconsidered taking carefully into account, the data collection strategy actually adopted. Broadly speaking, if longitudinal data are (at least partly) collected by retrospective interrogation, one can argue that memory inaccuracy leads to correlated errors.

Specific assumptions about classification errors can be successfully introduced in appropriate statistical models, only if additional information is available in the form of plausible *a priori* knowledge about the error generating mechanism and/or supplementary data about the labour force state.

Modelling strategies to correct gross flows for classification errors, should then take into account the measurement process actually used, in the sense that the amount of classification errors and the direction of possible bias, are related to the strategy adopted to collect longitudinal data.

As it is well known, longitudinal data can be obtained by different survey strategies. It is convenient to distinguish at least between (i) panel surveys and (ii) retrospective surveys. In addition, the availability of multiple indicators deserves specific attention.

Panel surveys are the most natural ways of collecting longitudinal information. Among these, rotating panel surveys play a prominent role. In fact, this is the scheme adopted in most national Labour Force Surveys (LFSs), whose primary goal is estimation of labour force stocks. For LFSs with a rotating sampling design, longitudinal information on the (usually short) sequence of states, can be easily obtained by matching data on individuals participating in two or more successive surveys. In LFSs, the reference period, concepts and definitions for classifying people, are typically consistent with the International Labour Office (ILO) recommendations (Hussmanns, Mehran and Verma 1990): this makes measures of labour force conditions reasonably accurate and comparable over space and time. Data on labour force participation are collected also through general purpose household surveys. In this case, attention to labour force condition is less prominent than in the preceding type of surveys, and reference periods, concepts and definitions, might be less consistent with ILO recommendations.

Alternatively, longitudinal information can be collected by retrospective surveys. Cross-sectional surveys can include retrospective questions, to get information on the sequence of labour force states experienced by sampled individuals. In this case, the interrogation strategy is crucial to reduce errors due to memory (recall errors, telescoping, *etc.*). Procedures to improve accurate reporting in retrospective surveys, rely upon contributions from cognitive psychology and survey methodology (for a review, see O'Muircheartaigh 1996). Besides, evidence on the amount and the direction of bias due to memory inaccuracy, is found in many empirical studies. It is worth adding, that in retrospective surveys, factors related to length of recall period, salience of events considered, and/or difficulty in retrieving data on past events, usually lead to a simplified format of questions, not consistent with ILO conventions on labour force condition.

Interesting opportunities for estimating gross labour flows in the presence of classification errors, come from the widespread practise of using a mixture of the panel and the retrospective strategies. Panel surveys use retrospective questions, at least on a limited number of topics, to cover the period between two successive waves (this is the case of the Survey of Income and Program Participation, as will be seen in section 4.2). The main characteristics of the measurement process when such a mixed strategy is used, have to be carefully considered, as they might have a considerable impact in formulating reasonable models for classification errors. More specific traits of the measurement process emerge also from consideration of the peculiarities of the survey design.

From a different perspective, an important opportunity for modelling classification errors is given by the availability of multiple measurements of labour force state, *i.e.*, data on the labour market condition of an individual at a given time, provided by two or more different sources. This information is of great importance in general, and particularly when fairly complicated patterns of correlated classification errors are to be considered. Multiple indicators on labour force state can be collected (i) in the same interview or (ii) in different interviews (*e.g.*, in different waves of a panel survey).

The first case is not very common, but sometimes questions regarding labour force condition are asked in different contexts, and in different ways. For instance, first, a self-classification of the individual with respect to labour force condition is asked; then, in a different section of the questionnaire, a sequence of questions are put forward that allow to classify the respondent according to standard labour force definitions. (For a different example, see the case of the Survey of Income and Program Participation in section 4.2.)

The second case covers several situations. At least two of them are worth considering:

(a) data from reinterview studies, often collected specifically to get information on classification errors probabilities (in such a case, the common practice is to assimilate reinterview data to validation data: for classical procedures to correct gross flows based on reinterview data, see Abowd and Zellner 1985, Poterba and Summers 1986, and Chua and Fuller 1987);

(b) data collected retrospectively in panel surveys, but referring to a time point already covered by the preceding interview, or collected in a supplementary survey carried out occasionally and covering the reference period(s) of the current panel survey. It is obvious that, in this case different measures of the same variable(s) of interest can be polluted by classification errors with largely different characteristics.

Many of the points raised here will be clarified in the case-studies presented in section 4, where the joint presence of panel and retrospective information and of multiple indicators of the same latent variable is exploited in order to get parsimonious models.

3. ESTIMATING GROSS FLOWS AFFECTED BY CLASSIFICATION ERRORS

3.1 A General Framework

Specification of statistical models to adjust labour force gross flows for classification errors, should allow one to take into account, the nature of available data (as reviewed in the previous section), and substantial assumptions on the generating process of (i) transitions among labour force states (*e.g.*, Markov chain structures) and (ii) measurement errors (*e.g.*, uncorrelated *vs.* correlated measurement errors).

In the simplest case, we consider panel data, where at each time period $t = 1, ..., T$, a discrete variable Y_t is observed for a generic unit, in a random sample of size n. In our case-studies, the units will be individuals, and the time periods, months or quarters. Y_t takes one among r possible distinct values or states. Y_t is an imperfect measure of y_t, which denotes the true state of a generic unit at time t. In general, it is not necessary to assume, that y_t varies over the same set of states $1, 2, ..., r$, but for simplicity, and without loss of generality, we will consider here the same set of states as for Y_t.

Strategies for estimating gross flows, rely upon an appropriate specification of the joint probability of the true and the observed process $P(Y_1, ..., Y_T, y_1, ..., y_T)$. Statistical analysis is then based on marginalization with respect to unobserved quantities:

$$P(Y_1, ..., Y_T) = \sum_{y_1=1}^{r} ... \sum_{y_T=1}^{r} P(Y_1, ..., Y_T, y_1, ... y_T). \quad (3.1)$$

Models are based on parsimonious specifications of the joint probability function $P(Y_1, ..., Y_T, y_1, ... y_T)$. Essentially this can be obtained by decomposing it into a product of conditional probabilities, following from an appropriate set of assumptions about the dependence structure among the components $Y_1, ..., Y_T, y_1, ..., y_T$.

For our purposes, a convenient starting point for model specification, comes from assumptions (i) about the structure of the generating process of the true transitions among labour force states and (ii) about the measurement process (exploiting, for instance, substantial knowledge or empirical evidence from the data collection strategy adopted).

In a model aimed at distinguishing between true and observed turnover in the labour market, a typical example that exploits this idea, is provided by Latent Class Markov (LCM) models (van de Pol and Langeheine 1990). For a generic unit, the following probabilities are specified:

$$q_t^{l_t j_t} = P(Y_t = l_t \mid y_t = j_t) \qquad t = 1, ..., T \quad (3.2)$$

$$\pi_t^{j_t j_{t-1}} = P(y_t = j_t \mid y_{t-1} = j_{t-1}) \qquad t = 2, \ldots, T \qquad (3.3)$$

$$\pi_1^{j_1} = P(y_1 = j_1) \qquad (3.4)$$

Conditional probabilities (3.2) represent the relationship between true and observed states, *i.e.*, the probability of reporting at time t, state l_t, while the true state is j_t. Clearly, this specification implies the local independence assumption, *i.e.*, $Y_1, \ldots, Y_T$ are independent, given $y_1, \ldots, y_T$. Conditional probabilities (3.3) describe the dynamics in the labour market, *i.e.*, the probability that a transition from j_{t-1} to j_t occurs, when moving from time $t-1$ to t: according to (3.3), the true transition process evolves following a first order Markov chain. Finally, probabilities (3.4) describe the initial condition for the Markov process.

The marginal probability for the observed sequence (3.1) is then given by:

$$P(Y_1 = l_1, \ldots Y_T = l_T) = \sum_{j_t=1}^{r} \ldots \sum_{j_r=1}^{r} \pi_1^{j_1} \prod_{t=2}^{T} q_t^{l_t j_t} \pi_t^{j_t j_{t-1}} \qquad (3.5)$$

For four measurement points, model (3.5) is equivalently represented by the path diagram in Figure 1, where arrows indicate direct effects between variables.

$$\begin{array}{ccccccc} y_1 & \rightarrow & y_2 & \rightarrow & y_3 & \rightarrow & y_4 \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ Y_1 & & Y_2 & & Y_3 & & Y_4 \end{array}$$

Figure 1. Path Diagram of a LCM Model for Four Measurement Points

It is worth observing, that the assumption of local independence is equivalent to the Independent Classification Errors (ICE) assumption. As noted in the previous section, the ICE assumption has been severely criticised, and seems definitely unreasonable when longitudinal data are collected by retrospective questions.

As another example, for $T = 2$, classical strategies to correct gross flows based on reinterview studies, can be represented within the framework outlined above. In this case, additional information is used, in the sense that the q_t parameters are exogenously estimated from the reinterview study, and are plugged in (3.5) in order to obtain directly $P(y_1, y_2)$.

The same framework can be used, to encompass more general assumptions on both the latent and measurement processes, up to include serially correlated classification errors. As an interesting case, we consider the model by Pfeffermann, Skinner and Humphreys (1998). Ignoring here initial conditions, they reformulate conditional response probabilities as follows:

$$q_t^{l_t j_t} = P(Y_t = l_t \mid y_t = j_t, Y_{t-1} = l_{t-1}) \qquad t = 2, \ldots, T, \qquad (3.6)$$

thus overcoming the ICE assumption.

A similar formulation, aimed at introducing, at least partially, dependence between the observed state at time t and the sequence of true states at times t and $t-1$, has been suggested by van de Pol and Langeheine (1992), who extend the model to allow also for a second order Markov chain, for the true transition process.

The modelling strategy for estimating true flows can be further extended in various directions, namely:

(a) It is straightforward to extend the model, to exploit the availability of multiple indicators of the same unobserved true state. This implies that response probabilities, as those in (3.2), are defined for one or more additional observed variables, treated as imperfect measures of the same latent state y_t. As an example, a LCM model for two indicators per latent variable, and four points in time, is represented in Figure 2. In this model, each couple of indicators referring to a given point in time, is assumed to be independent, conditionally on the corresponding latent variable, in the sense that the correlation between them, is completely explained by their relation with y_t.

(b) Observed heterogeneity at the individual level, in the transition and/or the measurement processes, can be introduced by conditioning on a set of covariates X_t. An example is given in Pfeffermann *et al.* (1998). They use covariate information at the unit level and model their impact on labour market condition by multinomial logit .

(c) Unobserved heterogeneity can also be considered, which leads to mixed latent class models (van de Pol and Langeheine 1990). A simple case is the movers/stayers model, where a different behaviour, at the latent level, is assumed for groups of units, while the group membership of the units cannot be directly observed.

$$\begin{array}{ccccccc} W_1 & & W_2 & & W_3 & & W_4 \\ \uparrow & & \uparrow & & \uparrow & & \uparrow \\ y_1 & \rightarrow & y_2 & \rightarrow & y_3 & \rightarrow & y_4 \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ Y_1 & & Y_2 & & Y_3 & & Y_4 \end{array}$$

Figure 2. Path Diagram of a LCM Model for Four Measurement Points and Two Indicators for Each Latent Variable

3.2 Latent Class and Related Models as a Tool for Estimating Gross Flows With Measurement Errors

A special case of the general model formulation outlined in the above section, are latent class models, where the true state in the labour market plays the role of the latent variable, and the observed state acts as its indicator. Some of the specifications outlined in the previous section, include dependence among classification errors. A general and convenient approach for handling it, which includes standard latent class models with correlated classification errors, is the so called modified LISREL model proposed by Hagenaars (1990).

The modified LISREL approach consists of an extension of Goodman's (1973) path analysis, which is a tool to describe causal relationships among observed categorical variables, through a system of logit equations. Basically, the extension incorporates latent variables. Thus, a modified LISREL model combines a measurement sub-model, which specifies the dependence of the indicators on latent variables, and a structural sub-model, which specifies ordered relations among latent and possible external variables. As the name itself suggests, it can also be viewed as the analogue for discrete variables, of the well known LISREL model for continuous variables (Joreskög and Sörbom 1988).

Modified LISREL models, allow to introduce serially correlated classification errors, by inserting direct effects between the indicators (Hagenaars 1988). The presence of direct effects implies, that the association among observed variables, is not completely explained by the effects of the latent variables on their indicators, but that there exists a source of additional association among the indicators, over and above the part that is explained by their relation with the latent variables.

Once a reasonable model has been specified, identification should be ascertained. The model involves many unobservables, and identification of all parameters is not automatically assured.

Reasonable opportunities to achieve identification, rest on two strategies, possibly used in combination: (i) imposition of plausible equality restrictions among the set of parameters and (ii) availability of multiple indicators of the unobserved true state. The latent class Markov model represented in Figure 1, for example, is not identified without extra restrictions on its parameters. If the latent chain is assumed to be time homogeneous, or response probabilities are restricted to be equal across time, the model can be shown to be identified (Lazarsfeld and Henry 1968). Availability of multiple indicators for the unobserved true state, can also help identification of complex measurement models. Identification criteria for some very special specifications, have been proven (for example, the model in Figure 2 can be shown to be identified), but no general rules have been provided yet to ascertain global identification. It is advisable to check at least local identification, *i.e.*, identifiability of the unknown parameters in a neighbourhood of the maximum likelihood solution. Goodman (1974) stated that a sufficient condition for local identifiability of a latent class model, is that the Information matrix be full of rank. Goodman's condition may be computationally difficult to check. Moreover, with some data sets, it may happen that the Information matrix is not of full rank, simply because some estimates are very close to the boundaries of the parameter space. An alternative, empirical way to check identifiability, is to estimate the model using different sets of starting values. If different sets of starting values result in the same value for the log-likelihood function but in different parameter estimates, then the model is not identifiable.

As for estimation, modified LISREL models may be treated as directed loglinear models with latent variables (Hagenaars 1997). A directed loglinear model results in a sequence of parsimonious multinomial logit models, possibly with latent variables, which are estimated stepwise. At each step, one dependent variable is considered, and a multinomial logit model is estimated on a contingency table, which has been collapsed over the variables, that do not directly influence the dependent variable in the causal order. Estimates obtained at each step are, at the end, combined in order to obtain estimated parameters for the full model. Directed loglinear modelling yields exactly the same parameter estimates, standard errors and test statistics as the Goodman standard procedure, but using simpler marginal tables. If the causal model contains one or more latent variables, an appropriate estimation technique must be used, *e.g.*, an implementation of the EM algorithm (Meng and Rubin 1993).

The empirical validity of the complete causal model may be tested, comparing the estimated expected frequencies with the observed ones in the complete table, by means of the likelihood ratio L^2 and the Pearson X^2 statistics. However, the structure of the observed data on labour market transitions, is such that many cells show very low observed frequencies. For this reason, the usual X^2 and L^2 criteria must be used only as a general indication of fit, since their asymptotic χ^2 distribution is no longer guaranteed, due to the sparse and unbalanced pattern of the contingency table.

Various strategies can be adopted to extend and improve model evaluation, and three of them are worth mentioning in this context:

(i) A restricted model nested within a larger one, can be tested with the conditional test, *i.e.*, considering the difference in the L^2 values of the two models, which is asymptotically distributed as χ^2 under weaker conditions (Goodman 1981, and Haberman 1978).

(ii) In general, using multiple criteria can be a sensible strategy. Indices based on the information criterion, such as AIC or BIC, can be useful to compare alternative non-nested models. Another advantage of AIC and BIC is that, in the selection procedure, they weight the goodness of fit of a model against its parsimony, considering the model degrees of freedom and the sample size. (AIC = L^2 - 2 x degrees of freedom. BIC = L^2 -1n(N+1) x degrees of freedom.) The model that is preferred, in this context, is the one with the lowest value of AIC or BIC.

(iii) Monte Carlo resampling techniques can be implemented to simulate the asymptotic distribution of X^2 and L^2 (Langeheine, Pannekoek and van de Pol 1995).

4. TWO CASE-STUDIES

4.1 The General Set Up

In this section we present two applications of the modified LISREL approach to correct observed gross flows in the labour market. Data come from surveys with partly different designs:

(1) the U.S. Survey of Income and Program Participation (SIPP), a multi panel household survey, which collects retrospective information on the between waves working history;
(2) the French Labour Force Survey (FLFS), a yearly retrospective survey, with one month overlapping reference periods.

For each case-study, a model is specified on the basis of *a priori* information on both the true transition process and the error generating mechanism. *A priori* information is crucial for model specification, in order to obtain parsimonious and plausible models.

All the models are written in the form of a modified LISREL model, and estimated by the EM algorithm. Actually, we used the *l*EM program (Vermunt 1993) and checked all the models for local maxima.

The two final models turn out to be rather complex, since they incorporate correlation among classification errors, and specific assumptions on respondent's behaviour. This fact, together with the sparse and unbalanced pattern of the observed contingency table, typical of labour force transitions, demands for goodness of fit evaluation criteria, other than L^2 and X^2. In the first case-study, alternative models have been judged by means of the BIC index, and on the basis on substantive knowledge on the labour market in the U.S.. In the second case, alternative models have been compared by means of the conditional test.

In the following sections, models are presented in a logical and verbal form, while the mathematical formulation for the final model is given in the relevant Appendix.

4.2 The SIPP Data

SIPP is a multi panel household survey conducted by the U.S. Bureau of the Census, in order to collect information on topics such as employment, income, participation in social programs, *etc.* The reference population is the U.S. noninstitutionalized individuals over 14.

The survey started in 1984, and is a continuing one: as a general pattern, each year a new sample of households, called "panel", has been selected for the survey and followed for two and half years (for a detailed description of SIPP, see U.S. Department of Commerce 1991, and Citro and Kalton (1993)).

Each panel is randomly divided into four "rotation groups" and interviewed at 4-months intervals for eight times. For practical reasons, each rotation group is interviewed in each of four consecutive months, and retrospective questions collect information with reference to the 4-months period elapsing between subsequent interviews. Each set of interviews with the full sample is termed a "wave".

We will refer to the 1986 panel, which started in February 1986 and ended in August 1988. We will consider the intermediate period from January 1986 to January 1987, over which we have information from all four rotation groups. Figure 3 represents the survey design with regard to our sample.

Information on labour force participation, is collected mainly in the "Labour Force and Recipiency" section of the questionnaire (for an additional piece of information, collected in another section of the questionnaire, see below), where each respondent is asked to report on a weekly basis his/her labour market history in the preceding four months (18 weeks), by going through a series of filtered questions. The respondent is first asked whether he/she had a job or a business, at any point in time during the reference period. If the respondent gives a negative answer, he/she is asked whether he/she spent any time looking for work, or was in layoff, and, if so, in exactly which weeks. On the other hand, if the answer to the

Interview Month	Rot. Group	Wave	Reference months									
February	2	1	Oct	Nov	Dec	Jan						
March	3	1		Nov	Dec	Jan	Feb					
April	4	1			Dec	Jan	Feb	Mar				
May	1	1				Jan	Feb	Mar	Apr			
June	2	2					Feb	Mar	Apr	May		
July	3	2						Mar	Apr	May	Jun	
August	4	2							Apr	May	Jun	Jul
September	1	2								May	Jun	Jul	Aug

Figure 3. Rotation Plan for the 1986 SIPP Panel (First 2 Waves)

starting question is positive (*i.e.*, he/she worked some time), and the respondent declared a job or a business with continuity during the reference period, he/she will move to the following section of the questionnaire. The respondent not declaring a stable situation in the labour market, is asked a long series of questions in order to establish the labour force state occupied, in each single week of the reference period.

The weekly based information is usually recorded, to obtain a monthly classification based on the usual three categories: Employed (E), Unemployed (U) and Not in the labour force (N). For individuals covering different positions during one month, the monthly labour force state is the one identified by the "modal" category with regard to the weeks of that month (Martini 1989).

Observed gross flows between two generic calendar months are then obtained as follows:

(a) For individuals belonging to three rotation groups, on the basis of retrospective data collected in the same interview. These observed flows will be called "within wave" (WW) transitions.
(b) For individuals in the fourth rotation group, by combining information collected in two different interviews, four months apart. These observed flows are termed "between waves" (BW) transitions.

When estimating monthly changes, a peculiar problem with SIPP data, is the so called "seam effect" (Young 1989): more changes are observed when data for two adjacent months are collected in two different waves – the transition covers the seam of the waves – than when they come from the same interview. The seam effect is pervasive in the survey: evidence of it for several variables of interest, is reported in Martini (1988), Marquis and Moore (1989), Kalton and Miller (1991).

Table 1 illustrates this phenomenon for our 1986 SIPP panel sample. Row 4→1 contains average BW transition rates; rows 1→2, 2→3 and 3→4 contain average WW transition rates, pertaining to the position of the two relevant reference months in each wave (for example, row 1→2 contains transition rates between the first two reference months in each wave). From Table 1, there is clear evidence that observed WW transitions describe a more stable labour market than BW ones. Moreover, WW stability increases, moving backwards in the wave (from 3→4 to 1→2).

One reasonable explanation for the seam effect, and for the systematic pattern of observed transitions throughout a wave, is the different role of measurement errors, for data obtained under the BW and WW strategies respectively. Specifically, it is likely that classification errors have a different degree of correlation for WW and BW observed flows: the higher stability documented by WW transitions may be induced by highly correlated classification errors. Indeed, if errors were uncorrelated, specifically for WW transitions, no evidence of seam effect would be expected.

A variety of plausible causes of correlated errors, is suggested by the cognitive psychology and the survey methodology literature on memory effect and recall errors (see, Bernard, Killworth, Kronenfeld and Sailer 1984, and O'Muircheartaigh 1996), among which a "conditioning" effect: respondents tend to give the same answer going backwards within the wave, and in extreme cases, they mechanically repeat the same answer for all four months.

Table 1
Observed Monthly Transition Rates (×100) for the 1986 SIPP Panel, January 1986 to January 1987

	Type	EE	EU	EN	UE	UU	UN	NE	NU	NN
1→2	WW	98.27	1.04	0.69	15.46	79.63	4.91	1.15	1.42	97.43
2→3	WW	97.91	1.13	0.96	17.34	75.96	6.70	1.38	1.71	96.91
3→4	WW	97.85	1.20	0.95	19.23	73.25	7.52	1.28	1.69	97.03
4→1	BW	94.03	2.10	3.87	26.81	42.20	30.99	5.65	3.77	90.58

Abundant empirical literature shows, that this sort of conditioning effect is the main source of classification errors in SIPP data. Other potential sources of error, typical of panel surveys, do not affect SIPP data dramatically. Administrative record check studies find little, if any, evidence of time-in-sample effect (Chakrabartry and Williams 1989; McCormick, Butler and Singh 1992). As a general consideration, we may say that in SIPP data, the seam effect dominates over other sources of error, that potentially bias gross flows estimates.

Summing up, a model-based approach to obtain unbiased gross flows from SIPP data, is justified by two arguments:

(a) the patent presence of correlated classification errors;
(b) *a priori* information on the data generating mechanism, drawn from two sources:
    (b1) specific evidences emerging from SIPP observed gross flows, such as the seam effect, and the increase in stability going backwards within the wave, just documented;
    (b2) general hints provided by the social survey literature on respondent behaviour.

In order to correct SIPP observed labour force gross flows from classification errors, a model has been built, based on the following assumptions/information:

(a) the true transition process follows a first order Markov chain;
(b) WW data transitions are affected by correlated classification errors, according to a pattern that will be specified in the sequel;
(c) for BW, the standard ICE assumption holds;
(d) rotation groups are equivalent samples also for modelling purposes, *i.e.*, respondents behave in the same way in all four rotation groups;
(e) SIPP data provide two indications on the monthly labour force state of each individual: the detailed information collected in the "Labour Force and Recipiency" section

of the questionnaire, just presented, and the additional information collected in the "Earnings and Employment" section, where the respondent is asked if he/she did/did not have a job in the reference period, on a weekly basis.

$$
\begin{tikzcd}
W_1 & W_2 \arrow[l] & W_3 \arrow[l] & W_4 \arrow[l] \\
y_1 \arrow[u] \arrow[r] \arrow[d] & y_2 \arrow[u] \arrow[r] \arrow[d] & y_3 \arrow[u] \arrow[r] \arrow[d] & y_4 \arrow[u] \arrow[d] \\
Y_1 & Y_2 \arrow[l] & Y_3 \arrow[l] & Y_4 \arrow[l]
\end{tikzcd}
$$

Figure 4. Path Diagram of a Modified Lisrel Model for Four Measurement Points and Two Indicators for Each Latent Variable (for the Meaning of Symbols, See Main Text)

Figure 4 contains the path diagram of a simplified version of the model (*i.e.*, a version that does not aim at representing in detail, the pattern of correlated classification errors, nor at taking into account the fact that we are dealing with four rotation groups) for four points in time, *i.e.*, for four consecutive calendar months. Here y_t ($t = 1, 2, 3, 4$) represents latent variables; Y_t and W_t represent indicators; arrows indicate direct effects between pairs of variables. Indicator Y_t refers to the reported labour force state, described by the usual three categories (E, U and N), while W_t refers to the binary variable Job/No Job. Since information is collected in two different sections of the questionnaire, and with different interviewing procedures, Y_t and W_t can be assumed to be independent given y_t. On the other hand, direct effects between the indicators, account for correlated classification errors over time: the response given for time $t + 1$ affects that given for time t. Note also, that an additional variable G with four categories should be added to the diagram, to account for rotation group membership. All indicators depend on G, since units in different groups are interviewed in different calendar months.

The basic equation of the model, decomposes the proportion in the generic cell of the 9-way contingency table, in the product of the conditional probabilities reported in Appendix A, equations (A1) to (A7). A preliminary version of the model has been proposed in Bassi, Croon, Hagenaars and Vermunt (1995).

Equation (A1) defines the probability of belonging to one of the four rotation groups. Equations (A2) and (A3) define the initial condition, and the transition probabilities, of the latent first order Markov chain respectively. Equations (A4) and (A5) define the response probabilities for indicator Y_t; equations (A6) and (A7) the analogous probabilities for the dichotomous indicator W_t. The response probabilities are defined in such a way that the answer given for a certain month, depends jointly on the current true state (y_t) and on the "past" true and "past" reported states (y_{t+1} and Y_{t+1}). The term "past" refers to the way respondents think, while answering retrospective questions: they start recalling from the moment of time nearest to the interview, and go backwards up to the end of the reference period.

A complex set of constraints has been imposed on response probabilities of (A4), (A5), (A6) and (A7), to account for (i) the conditioning effect, and (ii) the fact that the four rotation groups are equivalent samples in terms of the error generating mechanism.

These constraints are formulated in detail in Appendix A. Basically, they incorporate *a priori* knowledge on respondent's behaviour, and allow us to specify a parsimonious model. Specifically, equations (A8) to (A14) correspond to the following statements:

(a) With regard to WW classification errors, following Hubble and Judkins (1989), it is assumed that:
   (a1) a respondent who reports wrongly his/her labour force state for a certain month, continues to repeat this same answer also for the adjacent month, going backwards within the wave (A8);
   (a2) if, however, the status at time $t + 1$ is correctly reported, the response probability for the adjacent month depends only on the current true state (A9);
   (a3) the same error generating mechanism operates for both indicators. For W_t, we state that a correct answer is given when the true state is E and 'Job' is reported and when the true state is U or N and 'No Job' is reported, (A10) and (A11).

(b) Response probabilities are set equal across rotation groups, (A12) to (A15). As an example, equalities in (A12) mean that response probabilities for individuals in rotation group 1 for the month of April, are equal to response probabilities for individuals in group 4 for the month of March, to those for individuals in group 3 for the month of February, and to those for individuals in group 2 for the month of January. (They are set to be equal, since they all refer to the answer given for the last month of the wave.)

The model has been estimated to correct observed monthly gross flows for the quarter January to April 1986 (Table 2). The comparison between observed and estimated flows, highlights that the model reduces the seam effect: WW transitions are corrected towards a more dynamic labour market; BW transitions are corrected in the opposite direction. It is worth noting, that effects of model correction are more evident for flows from unemployment, which are characterised by higher mobility.

The goodness of fit of the model has been judged by multiple criteria such as the BIC index and the conditional test for nested models, together with estimate interpretability and consistency, with substantive knowledge of the dynamics of the U.S. labour market in '80s.

4.3 The French Labour Force Survey Data

The second case-study refers to the flows in the labour market, observed with the French Labour Force Survey (FLFS) conducted yearly by INSEE in France.

Table 2
SIPP Observed and Estimated Monthly Transition Rates (×100), January to April 1986

		EE	EU	EN	UE	UU	UN	NE	NU	NN
J-F	WW	98.11	1.17	0.72	14.53	80.16	5.31	0.90	1.57	97.53
	BW	94.08	2.17	3.75	23.58	44.30	32.12	5.62	3.45	90.93
	Estimated	97.25	1.47	1.28	16.08	77.16	6.76	1.59	1.32	97.09
F-M	WW	98.66	0.92	0.42	16.06	78.67	5.27	0.64	1.65	97.71
	BW	94.88	1.91	3.21	21.90	48.54	29.56	4.99	4.11	90.90
	Estimated	97.83	1.20	0.97	19.40	74.01	6.59	1.21	1.50	97.29
M-A	WW	98.71	0.64	0.65	20.76	71.74	7.50	1.47	1.05	97.48
	BW	95.59	1.52	2.89	30.48	34.92	34.60	6.34	3.78	89.88
	Estimated	98.11	0.95	0.94	26.42	65.75	7.83	2.17	0.71	97.12

The reference population of the FLFS are all members of French households, who are above 15 in the year in which the interview is planned. The survey has a rotating design: each year, one third of the sample is renewed.

Information on labour force participation is collected with retrospective questions, having as a reference period the 13 months preceding the interview. Each respondent is asked to recall his/her position in the labour market on a monthly basis, by filling in a grid in which he/she can classify himself/herself, for each month, over eight categories: self-employed, employed on a fixed term basis, permanently employed, unemployed, on training, student, serving in the Army, other (retired, housewife, *etc.*).

For our analysis, we aggregated the eight categories in the usual three states E, U and N. We consider 'Employed' respondents who classify themselves in the first three categories, 'Unemployed' those who classify themselves in the fourth category and 'Not in the labour force' the remaining ones.

We analyze the information collected in the two consecutive waves of March 1991 and March 1992, on a subsample of individuals: those who answered to three consecutive interviews (January 1990, March 1991 and March 1992) and who were 18 to 29 years old in 1992, for a total of 5,427 individuals. The reference periods of the two waves considered, overlap in March 1991. We have, then, two pieces of information on the labour force state for this month: one collected in March 1991, and the other one collected with a retrospective question 12 months afterwards.

The pattern of observed monthly transitions in our FLFS sample shows some interesting evidence, largely dictated by the characteristics of the subsample – young people.

Transitions exhibit a moderate degree of seasonal variation, related to the school calendar. From June to July, for example, we observe a proportion of people who enter the labour market as employed, greater than the average; on the contrary, from August to September, a proportion greater than the average leaves employment (presumably to education).

The marginal distribution of the three states from March 1990 to March 1992, shows that the individuals in our sample progressively enter the labour market: in March 1990, 44% are observed to be Employed or Unemployed, whereas by March 1992, this proportion has risen to 54%.

The double information for March 1991, provides some crude evidence on response error in the data: 8% of respondents declare a different state in the two interviews. For the period from February to April 1991, two types of flows may be observed: a within wave (WW) one, *i.e.*, information about the labour force state is collected in the same interview, and a between waves (BW) one, *i.e.*, information is collected in two different interviews (Table 3).

Table 3
FLFS Observed Monthly Transition Rates (×100) from February to April 1991

		EE	EU	EN	UE	UU	UN	NE	NU	NN
F-M	WW	98.19	1.67	0.14	9.11	90.65	0.24	0.28	0.11	99.61
	BW	93.17	3.58	3.25	25.18	65.23	9.59	3.75	1.96	94.29
M-A	WW	98.60	1.04	0.36	8.89	90.37	0.74	0.24	0.29	99.47
	BW	93.24	3.33	3.43	25.90	63.79	10.31	3.79	2.07	94.14

As expected, WW transitions describe a more stable labour market than the BW ones. This can be considered as an indication of correlated classification errors in the data. Patterns and causes of errors correlation in retrospective surveys, have been extensively discussed in the two previous sections, and the above considerations can largely be extended to the FLFS data.

In general, we expect that, in a retrospective survey with such a long recall period, lack of memory results in the major cause of classification errors. We also expect that the probability of answering incorrectly, increases as the distance between the reference month and the interview month gets longer. This may be considered as the major source of correlation among classification errors, together with telescoping and conditioning effects, which possibly affect FLFS data as well (see Magnac and Visser 1995).

The overall effect of correlated classification errors, reasonably results in an underestimation of mobility in the French labour market.

Moving from these considerations, we specified a model to correct observed quarterly gross flows, from measurement error (Table 4). The last column of Table 4 contains the percentage of individuals who are observed to change state, between the two months considered (OM = observed mobility). On the average over the five WW transitions, 6.122% of mobility between two consecutive months is observed.

As in the previous case-studies, let us denote with $y_t (t = 1, 2, 3, 4, 5, 6)$ true labour force states, and with upper case letters their indicators: $Y_t (t = 2, 3, 4, 5, 6)$ represents labour force states observed in March 1992 (referring to March, June, September, December 1991 and March 1992); $W_t (t = 1, 2)$ represents labour force states observed in March 1991 (referring to December and March 1991). As usual, y_t, Y_t and W_t distribute over the three categories of E, U and N.

The model is specified by decomposing the proportion in the generic cell of the 7-way contingency table as in Appendix B, equations (B1) to (B6).

Since we observe two indicators only for one month, a model which assumes direct effects between the indicators, would be under identified. Thus, we can not explicitly model dependencies between observed states. The only way to account for correlated classification errors in FLFS data, is to let observed states depend on latent transitions. By the way, this seems to be a sensible assumption in retrospective surveys. Indeed, flows between two different states may easily undergo wrong placements in time, because in some situations, events might truly be difficult to place exactly. As an example, employees who loose their job or retire (flows EU and EN), will generally use the holidays they are entitled to, and may not clearly know when they exactly left employment. The moment people entered the labour force, may also be hard to recall, especially when they left school (flows NU and NE) (van de Pol and Langeheine 1997).

The modified LISREL model, formulated in mathematical terms in Appendix B, is based on the following substantive assumptions.

At the latent level, transitions follow a first order non stationary Markov chain (equations (B1) and (B2)). Indeed, the evidence on seasonality in observed transitions, suggest avoiding the imposition of stationarity of any order, on the latent Markov chain.

Response probabilities for data collected in both waves, depend on the latent transition occurring between t and $t + 1$ (equations (B3) and (B4) refer to data collected in March 1992, equations (B5) and (B6) to data collected in March 1991).

In order to describe the error generating mechanism in detail, and specify a more parsimonious model, the following constraints have been imposed on response probabilities:

(a) response probabilities referring to the same month of subsequent years (December and March) are set equal;
(b) response probabilities at time t, given that the true state has not changed between time t and time $t + 1$, are set constant over time;
(c) response probabilities are set equal for June and September 1991;
(d) in general, respondents who move between month t and $t + 1$ (transitions EU, EN, UE, UN and NU), at time t, report either the true state occupied at time t, or the true state occupied at time $t + 1$, *i.e.*, they, do not report a state they have not been moved from/to;
(e) if however, the latent transition occurs between states N and E, we admit all three answers at time t, *i.e.*, we consider that people who find a job may confuse their previous position (at time t), and be uncertain between U and N.

Constraint (c) is imposed mainly for reasons of model parsimony. It captures the notion that response probabilities for months that are placed more or less in the central part of the reference period, do not vary too much.

Constraints (b) and (d) reflect the fact that response probabilities depend on latent transitions. We expect that these probabilities do not vary too much over time when there is no latent change (constraint (b)), whereas we expect that the probability of misplacing change, especially in ambiguous situations, increases with the length of the recall

Table 4
FLFS Observed Quarterly Transition Rates (×100), December 1990 to March 1992
(OM = Observed Mobility)

		EE	EU	EN	UE	UU	UN	NE	NU	NN	OM
D90-M91	WW	94.77	4.25	0.98	24.53	72.40	3.07	0.98	0.66	98.36	5.08
	BW	91.50	4.86	3.64	31.60	56.84	11.56	4.40	2.10	93.50	10.16
M91-J91	WW	96.03	3.02	0.95	23.21	74.32	2.47	1.28	0.68	98.04	4.54
	BW	91.48	4.63	3.89	35.01	54.20	10.79	4.84	2.14	93.02	12.04
J91-S91	WW	94.29	3.94	1.77	20.93	78.29	0.78	4.71	2.95	92.34	7.85
S91-D91	WW	93.73	4.48	1.79	23.63	74.89	1.48	3.22	1.65	95.13	7.23
D91-M92	WW	93.90	4.80	1.30	21.67	76.74	1.59	1.70	0.59	97.71	5.91

period. Constraints under (d) aim at catching the telescoping effect.

Figure 5 gives the path diagram of the estimated model.

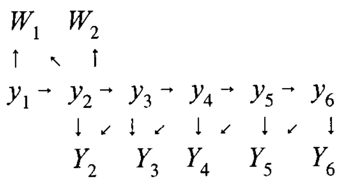

Figure 5. Path Diagram of a Modified Lisrel Model for Six Measurement Points and Two Indicators for One Latent Variable

Table B.1 in Appendix B reports the pattern of restrictions on response probabilities, (a) to (e); it shows which parameters are set equal, and which are fixed to 0, in order to introduce into the basic model, as defined by equations (B1) to (B6), the above constraints.

The final model has been selected after comparing a sequence of models, as can be seen from Table 5.

Table 5
Model Selection (EM = Estimated Mobility)

MODEL	L^2	df	ΔL^2	p-value cond. test	EM
A	2509.5759	2124			5.424
A1	3450.1716	2154	940.5957	0	4.918
A2	3849.9470	2178	399.7754	0	5.798
B	816.1620	2076			5.888
B1	855.2282	2094	39.0662	0.01	5.818
B2	864.9657	2106	9.7375	0.40	5.906
B3	879.5996	2121	14.6339	0.10	6.252

We started the analysis by estimating a model based on the ICE assumption (model A in the table), which, as expected, shows a bad fit.

The following models (A1 and A2) are based on the work by Magnac and Visser (1995). These authors consider monthly transitions over a period longer than ours (from January 1989 to March 1992), but on the same sample of individuals. They assume that the labour force state in the interview month is correctly reported, while the probability of making mistakes increases with the distance between the reference month and the time of interview, according to a deterministic function of time. Response probabilities are assumed to be constant over the survey waves, and true transitions are assumed to follow a first order stationary Markov chain. Our model A1 is a less restricted version of Visser and Magnac's model – no stationarity assumption is made, applied to quarterly transitions from December 1990 to March 1992. Our model A2 adds to model A1, the hypothesis of first order stationarity at the latent level. Both models perform quite badly, and (from column EM), we see that, on average, they correct the observed labour market towards stability: a result which contradicts the evidence on the effects of classification errors in retrospective surveys.

Model B introduces correlation among classification errors, by letting each indicator to depend on the true transition that occurred between times t and $t+1$; moreover, it encompasses constraint (a). The fit increases dramatically (see L^2). All subsequent models are nested in model B, and additional restrictions may be evaluated by a conditional test. Model B1 introduces constraints under (b); model B2 the additional constraints under (c); and model B3 is our final model.

Table 6 presents estimated transition rates with our best fitting model. The French labour market is corrected towards a greater mobility. The average estimated mobility amounts to 6.252%. Moreover, estimated response probabilities show a pattern consistent with the notion, that the probability of making mistakes gets bigger, the longer the recall period.

ACKNOWLEDGEMENTS

Research for this paper was supported by grants from CNR n. 94.02242.CT10 and MURST n. 02.09.02.110 and n. 02.09.02.124. We are grateful to Michael Visser and Thierry Magnac for providing us with anonimyzed individual data from the French Labour Force Survey. A preliminary version of this paper was presented at the IASS/IAOS Satellite Meeting on Longitudinal Studies, Jerusalem, August 27-31, 1997, where we benefited from comments and discussion. Detailed criticisms and suggestions from two referees are especially acknowledged.

Table 6
FLFS Estimated Quarterly Transition Rates (×100), December 1990 to March 1992
(EM = Estimated Mobility)

	EE	EU	EN	UE	UU	UN	NE	NU	NN	EM
D90-M91	94.85	4.48	0.67	12.70	66.28	21.02	1.09	1.55	97.36	6.27
M91-J91	95.65	1.37	2.98	28.43	62.35	9.22	3.61	1.48	94.91	7.49
J91-S91	93.71	4.25	2.04	14.88	82.50	2.62	4.11	3.49	92.40	7.70
S91-D91	98.32	1.67	0.01	15.42	83.75	0.83	3.80	0.47	95.73	4.24
D91-M92	93.23	5.02	1.75	9.99	88.65	1.36	2.07	1.28	96.65	5.56

APPENDIX A

Final Model Specification for the SIPP Data, in Terms of Conditional Probabilities

(1) Basic model decomposition

$$z_g = P(G = g) \quad \text{(A1)}$$

$$\pi_1^{j_1} = P(y_1 = j_1) \quad \text{(A2)}$$

$$\pi_t^{j_t j_{t-1}} = P(y_t = j_t \mid y_{t-1} = j_{t-1}) \qquad t = 2, 3, 4 \quad \text{(A3)}$$

$$q_{yt}^{l_t j_t l_{t+1} j_{t+1} g}$$
$$= P(Y_t = l_t \mid y_t = j_t, Y_{t+1} = l_{t+1}, y_{t+1} = j_{t+1}, G = g)$$
$$t = 1, 2, 3 \quad \text{(A4)}$$

$$q_{y4}^{l_4 j_4 g} = P(Y_4 = l_4 \mid y_4 = j_4, G = g) \quad \text{(A5)}$$

$$q_{wt}^{m_t j_t m_{t+1} j_{t+1} g}$$
$$= P(W_t = m_t \mid y_t = j_t, W_{t+1} = m_{t+1}, y_{t+1} = j_{t+1}, G = g)$$
$$t = 1, 2, 3 \quad \text{(A6)}$$

$$q_{w4}^{m_4 j_4 g} = P(W_4 = m_4 \mid y_4 = j_4, G = g) \quad \text{(A7)}$$

g varies over 1, 2, 3 and 4; l_t and j_t, $t = 1, 2, 3, 4$, vary over the categories E, U and N, m_t, $t = 1, 2, 3, 4$, vary over the categories 'Job' and 'No Job'.

(2) Constraints on conditional probabilities

$$q_{yt}^{l_{t+1} j_t l_{t+1} j_{t+1} g}$$
$$= P(Y_t = l_{t+1} \mid y_t = j_t, Y_{t+1} = l_{t+1}, y_{t+1} = j_{t+1}, G = g) = 1$$
$$\text{for } l_{t+1} \neq j_{t+1}\ t = 1, 2, 3 \quad \text{(A8)}$$

$$q_{yt}^{l_t j_t l_{t+1} j_{t+1} g} = P(Y_t = l_t \mid y_t = j_t, G = g)$$
$$\text{for } l_{t+1} = j_{t+1} \text{ and } t = 1, 2, 3 \quad \text{(A9)}$$

$$q_{wt}^{m_{t+1} j_t m_{t+1} j_{t+1} g}$$
$$= P(W_t = m_{t+1} \mid y_t = j_t, W_{t+1} = m_{t+1}, y_{t+1} = j_{t+1}, G = g) = 1$$
$$\text{for } m_{t+1} \neq j_{t+1} \text{ and } t = 1, 2, 3 \quad \text{(A10)}$$

$$q_{wt}^{m_t j_t m_{t+1} j_{t+1} g} = P(W_t = m_t \mid y_t = j_t, G = g)$$
$$\text{for } m_{t+1} = j_{t+1} \text{ and } t = 1, 2, 3 \quad \text{(A11)}$$

$$q_{t=4}^{g=1} = q_{t=3}^{g=4} = q_{t=2}^{g=3} = q_{t=1}^{g=2} \quad \text{(A12)}$$

$$q_{t=4}^{g=2} = q_{t=3}^{g=1} = q_{t=2}^{g=4} = q_{t=1}^{g=3} \quad \text{(A13)}$$

$$q_{t=4}^{g=3} = q_{t=3}^{g=2} = q_{t=2}^{g=1} = q_{t=1}^{g=4} \quad \text{(A14)}$$

$$q_{t=4}^{g=4} = q_{t=3}^{g=3} = q_{t=2}^{g=2} = q_{t=1}^{g=1} \quad \text{(A15)}$$

APPENDIX B

Final Model Specification for the FLFS Data, in Terms of Basic Model Decomposition and Pattern of Restrictions on Parameters

(1) Basic model decomposition

$$\pi_l^{j_1} = P(y_1 = j_1) \quad \text{(B1)}$$

$$\pi_t^{j_t j_{t-1}} = P(y_t = j_t \mid y_{t-1} = j_{t-1})$$
$$t = 2, 3, 4, 5 \quad \text{(B2)}$$

$$q_{yt}^{l_t j_t j_{t+1}} = P(Y_t = l_t \mid y_t = j_t, y_{t+1} = j_{t+1})$$
$$t = 2, 3, 4, 5 \quad \text{(B3)}$$

$$q_{y6}^{l_6 j_6} = P(Y_6 = l_6 \mid y_6 = j_6) \quad \text{(B4)}$$

$$q_{wt}^{m_1 j_1 j_2} = P(W_1 = m_1 \mid y_1 = j_1, y_2 = j_2) \quad \text{(B5)}$$

$$q_{w2}^{m_2 j_2} = P(W_2 = m_2 \mid y_2 = j_2) \quad \text{(B6)}$$

j_t, l_t and m_t vary over E, U and N.

(2) Pattern of restrictions on response probabilities

Table B.1
Month of Observation

Probability of observing a state given a latent transition	March 91	June 91 & Sept. 91	Dec. 90 & Dec. 91
E|ee	1	1	1
U|ee	2	2	2
N|ee	3	3	3
E|eu	F	F	F
U|eu	F	F	F
N|eu	*	*	*
E|en	F	F	F
U|en	*	*	*
N|en	F	F	F
E|ue	F	F	F
U|ue	F	F	F
N|ue	*	*	*
E|uu	4	4	4
U|uu	5	5	5
N|uu	6	6	6
E|un	*	*	*
U|un	F	F	F
N|un	F	F	F
E|ne	F	F	F
U|ne	F	F	F
N|ne	F	F	F
E|nu	*	*	*
U|nu	F	F	F
N|nu	F	F	F
E|nn	7	7	7
U|nn	8	8	8
N|nn	9	9	9

Equal numbers indicate response probabilities fixed to be equal.
* indicates a probability fixed to 0.
F indicates a free parameter.

REFERENCES

ABOWD, J.M., and ZELLNER A. (1985). Estimating gross labour force flows. *Journal of Business and Economics Statistics*, 3, 254-283.

BASSI, F., CROON, M., HAGENAARS, J., and VERMUNT, J. (1995). Estimating Latent Turnover Tables When Data are Affected by Correlated and Uncorrelated Classification Errors. WORC PAPER 95.12.25/7, Tilburg University.

BERNARD, H.R., KILLWORTH, P., KRONENFELD, D., and SAILER, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual Review of Anthropology*, 13, 495-517.

CHAKRABARTRY, R.P., and WILLIAMS, T.R. (1989). Time-in-sample biases in the Survey of Income and Program Participation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 309-314.

CHUA, T.C., and FULLER, W.A. (1987). A model for multinomial response error applied to labor flows. *Journal of the American Statistical Association*, 82, 46-51.

CITRO, C.F., and KALTON, G. (1993). *The Future of the SIPP*. Washington, D.C.: National Academy Press.

DUNCAN, G., and KALTON, G. (1987). Issues of design and analysis of surveys across time. *International Statistical Review*, 55, 97-117.

GOODMAN, L.A. (1973). The analysis of a multidimensional contingency table when some variables are posterior to the others. *Biometrika*, 60, 179-192.

GOODMAN, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 79, 1178-1259.

GOODMAN, L.A. (1981). Three elementary views of log-linear models for the analysis of cross-classifications having ordered categories. In *Sociological Methodology*, (Ed. S.Leinhardt), 193-293. San Francisco: Jossey Bass.

HABERMAN, S.J. (1978). *Analysis of Qualitative Data. Vol.1. Introductory Topics*. New York: Academic Press.

HAGENAARS, J.A. (1988). Latent structure models with direct effects between the indicators, local dependence models. *Sociological Methods and Research*, 16, 379-405.

HAGENAARS, J.A. (1990). *Categorical Longitudinal Data: Log-linear, Panel, Trend and Cohort Analysis*. Newbury Park: Sage.

HAGENAARS, J.A. (1997). Categorical Causal Modeling: Directed Loglinear Models With Latent Variables. WORC PAPER 97.04.002/7, Tilburg University.

HUBBLE, D.L., and JUDKINS, D.R. (1989). Measuring the Bias in Gross Flows in the Presence of Autocorrelated Response Errors. SIPP Working Paper No. 8712, U.S. Bureau of the Census.

HUSSMAN, R., MEHRAN, F., and VERMA, M. (1990). *Surveys of Economically Active Population, Employment and Underemployment: an ILO Manual on Concepts and Definitions*. Geneva: ILO.

JORESKÖG, K.G., and SÖRBOM, D. (1988). *Lisrel 7: A Guide to the Program and Applications*. Chicago: SPSS INC.

KALTON, G., and CITRO, C.F. (1993). Panel surveys: adding the fourth dimension. *Survey Methodology*, 19, 205-215.

KALTON, G., and MILLER, M.W. (1991). The seam effect with social security income in the SIPP. *Journal of Official Statistics*, 7, 235-245.

LANGEHEINE, R., PANNEKOEK, J., and van de POL, F. (1995). Bootstrapping goodness-of-fit measures in categorical data analysis. *Sociological Methods and Research*, 24, 492-516.

LAZARSFELD, P.F., and HENRY, N.W. (1968). *Latent Structure Analysis*. New York: Houghton Mufflin.

MAGNAC, T., and VISSER, M. (1995). Transition Models With Measurement Errors. Working Paper, Institut National de la Recherche Agronomique (INRA), Paris.

MARQUIS, K.H., and MOORE, J.C. (1989). Some response errors in SIPP with thoughts about their effects and remedies. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 381-386.

MARTINI, A. (1988). Retrospective versus panel data in estimating labour force gross flows: comparing SIPP and CPS. *Proceedings of the Social Science Section, American Statistical Association*, 109-114.

MARTINI, A. (1989). Seam effect, recall bias, and the estimation of labour force transition rates from SIPP. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 387-392.

MENG, X.L., and RUBIN, D.B. (1993). Maximum likelihood estimation via ECM algorithm: a general framework. *Biometrika*, 80, 267-278.

McCORMICK, M., BUTLER, D., and SINGH, R. (1992). Investigating time-in-sample effects for the Survey of Income and Program Participation. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 554-559.

O'MUIRCHEARTAIGH, C. (1996). Measurement errors in panel surveys: implications for survey design and for survey instruments. *Proceedings of the Scientific Reunion of the Italian Statistical Society*, 1, 207-218. Rimini: Maggioli.

PFEFFERMAN, D., SKINNER, C.J., and HUMPHREYS, K. (1998). The estimation of gross flows in the presence of measurement error using auxiliary variables. *Journal of the Royal Statistical Society*, Series A, 161, 13-32.

POTERBA, J.M., and SUMMERS, L.H. (1986). Reporting errors and labor market dynamics. *Econometrica*, 54, 1319-1338.

SINGH, A.C., and RAO, J.N.K. (1995). On the adjustment of gross flows estimates for classification errors with application to data from the Canadian Labor Force Survey. *Journal of the American Statistical Association*, 90, 1-11.

SKINNER, C.J., and TORELLI, N. (1993). Measurement error and the estimation of gross flows from longitudinal economic data. *Statistica*, 3, 391-405.

U.S. DEPARTMENT OF COMMERCE (1991). *SIPP User's Guide*. Washington D.C.

van de POL, F., and LANGEHEINE, R. (1990). Mixed Markov latent class models. In *Sociological Methodology*, (Ed. C.Clogg), 213-247. Oxford: Blackwell.

van de POL, F., and LANGEHEINE, R. (1992). Analysing Measurement Error in Quasi-experimental Data: An Application of Latent Class Models to Labour Market Data. Working Paper of the European Scientific Network on household Panel Studies, 57, Colchester, University of Essex.

van de POL, F., and LANGEHEINE, R. (1997). Separating change and measurement error in panel surveys with an application to labour market data. In *Survey Measurement and Process Quality*, (Ed. L.Lyberg *et al.*), 671-688. New York: Wiley.

VERMUNT, J.K. (1993). Log-linear and event history analysis with missing data using the EM algorithm. WORC PAPER 93.09.015/7, Tilburg University.

YOUNG, N. (1989). Wave seam effects in the SIPP. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 393-398.