Title*: Memory bias in retrospectively collected employment careers: a model based approach to correct for measurement error*

Word count: *8813*

Authors: *Anna Manzoni, CIQLE, Yale University**

Jeroen K. Vermunt, Ruud Luijkx, Ruud Muffels
Tilburg University, The Netherlands

* This research was funded as part of the NWO project 400-04-552 during the author's PhD studentship at Tilburg University

Memory bias in retrospectively collected employment careers: a model based approach to correct for measurement error

Abstract

Event history data constitute a valuable source to analyze life-courses, although their reliance on autobiographical memory raises many concerns over their reliability. In this article, using Swedish survey data, we investigate bias in retrospective reports of employment biographies applying a novel model-based latent Markov method.

A descriptive comparison of the biographies as reconstructed by the same respondents at two interviews carried out about 10 years apart reveals that careers appear simpler, less heterogeneous, with fewer elements and episodes when reported longer after their occurrence, with a particularly high underreport of unemployment. Using matching techniques, the dissimilarity between the two reconstructions turns out not to be affected by respondents' socio-demographic characteristics, but particularly by the occurrence of unemployment spells and the career complexity.

We use latent Markov models assuming correlated errors across occasions to determine the measurement error and to obtain a more reliable estimate of the (true) latent state occupied at a particular time point. The results confirm that (correlated) measurement errors lead to simplification and conventionalism. Career complexity makes recall particularly problematic at longer recall distances, whereas unemployment underreporting also happens very close to the interview. However, only a small portion of the sample makes consistent errors over time, while the great majority makes no errors at all.

1. INTRODUCTION

The study of the patterns and determinants of employment careers, their heterogeneity across socio-demographic groups, their changes over time and across cohorts, as well as the impact of labor market and welfare contexts, requires the use of longitudinal data sets (Mayer and Huinink, 1990; Solga, 2001). One possible way to collect employment career data is through panel studies, which are prospective studies in which respondents are followed during a certain period of time and in which the same set of questions about the respondent's current employment status is asked at each of the panel waves An alternative is to use a retrospective study design, in which respondents report their complete previous (employment) in a single survey (Taris, 2000). Retrospective data collection designs are very valuable for studying change over time in that they allowing tracing, in a relatively fast and cheap way, the entire individual career for respondents belonging to different cohorts and covering a long period in historical time. Such retrospective surveys usually collect information in the form of event histories, where respondents are asked to report the full sequence of episodes that they experienced up to the moment of interview (Scott and Alwin, 1998).

The quality of life-history data obtained via retrospective surveys depends on whether respondents are able and willing to recall and date each of the relevant episodes (or, equivalently, all relevant transitions). That is, retrospective surveys rely on the respondents' autobiographical remembering (Rubin and Baddeley, 1989), in the course of which information may be voluntarily or involuntarily distorted. Completeness and accuracy might be severely at risk when using memory-based reports, especially when the recall covers a longer period (Reimer, 2004). As a consequence, reported life-histories may contain serious errors, which may invalidate the substantive findings obtained when analyzing such data. Insight into the quality of reconstructed histories is therefore of great value for researchers using such data sets, since it will prevent them from drawing wrong conclusions. Moreover, once the type and amount of measurement error is known, it may be possible to correct for it during the actual data analysis.

In this article, we examine the quality of reconstructed employment biographies and how that might depend on factors such as the time lag between the events of interest and the interview, using a unique Swedish data set containing two reports of the same period. The two reports are analyzed using descriptive methods, using matching techniques for quantifying similarities between event sequences (Brzinsky-Fay and Kohler, 2001, Elzinga, 2007) as well as using latent Markov models (Magidson, Vermunt and Tran 2009, Van de Pol and Langeheine, 1990) The main contribution of this paper is that is demonstrates how to use the latter model-based approach to determine measurement error in life-history data, which as will be shown provides an improvement over purely descriptive approaches and matching techniques. The latent Markov model we use expands on the models proposed among others by Van de Pol and Langeheine (1990) in that it relaxes the assumption that errors are independent across time points. The remainder of this article is organized as follows. First, we explain the data used for our analyses. Subsequently, possible sources of errors in the survey and in retrospective employment history data in general are discussed. Then, we report on the data analyses and we round off with discussion and conclusions.

2. THE SWEDISH LEVEL OF LIVING SURVEY

A major source for studying labor market dynamics in Sweden is the Swedish Level of Living Survey (LNU), in which retrospective information on work histories is collected. The LNU, conducted by researchers at the Swedish Institute for Social Research (SOFI) at Stockholm University, is one of the longest running longitudinal social science surveys. It was conducted for the first time in 1968 and then repeated in 1974, 1981, 1991 and 2000, using a panel approach. These surveys were carried out as face-to-face interviews with national probability samples of individuals between the ages of 18 and 75 (15 to 75 until 1981). The methodology of the survey has gradually developed and in the 1991 wave a retrospective section was added to the questionnaire, recording information on cohabitation, children, education and economic activities with a precision of one month duration (Jonsson and Mills, 2001). The 2000 wave contained the same retrospective section.

In the working life biography section, data were collected in the form of event histories, where respondents had to recall and date one by one the full sequence of subsequent episodes in their life-course, in a temporal order, starting with the first job that lasted at least six months. For each episode, respondents had to provide information on the labor force status and end date of the episode,[1] with monthly precision in dates.[2]

---

[1] The design assumes no gaps, since no information on start was collected, neither are parallel activities registered.

[2] In 1991 this information was collected only for respondents born between 1925 and 1965. Furthermore those who claimed to have had 15 jobs or more were diverted to a shorter section on occupations. We wanted to focus on employment (and not job) histories, and we consequently decided not to consider those who

Consequently, a respondent's employment history consists of a continuous sequence of employment episodes interrupted by episodes of unemployment and labor market inactivity; that is, respondents move chronologically through mutually exclusive states. For the employment spells, information was gathered on occupation, industry, firm size, and sector (both in 1991 and 2000), as well as on part-time versus full-time work, permanent versus temporary employment, and whether the temporary employment forms part of a labor market program (in 2000 only).

New respondents in 2000 as well as those who participated in 1991 but did not answer the retrospective part had to report the whole employment history (from the first job lasting at least 6 months), whereas for those who participated in the 1991 survey, the data collection in 2000 starts asking the activity in January 1990. This allows us to reconstruct the whole career history of each respondent, until the last interview, offering a unique source of information to analyze the entire career patterns of several birth cohorts over a long time period.

The Swedish Level of Living Survey provides a unique opportunity to study memory bias in retrospective reports. As previously mentioned, respondents who participated in the 1991 survey are re-interviewed in 2000 (when possible) and asked to report their employment biography starting from the activity they held in January 1990. As a side effect, respondents provided a second report on the period from January 1990 to the

---

completed only the occupational history in our analysis, which means that we excluded respondents who declared to have had more than 15 jobs.

date of the interview in 1991.[3] More specifically, for respondents who participated in both waves, we have the information from the 1991 survey starting in January 1990 up to the interview date in 1991 and the information from the 2000 survey starting in January 1990 up to the interview date in 2000. This yields overlapping information for a certain number of months during the years 1990 and 1991. It should be noted that for the 1991 interview the overlap concerns a recall on what happened in the work life in the past year whereas in the 2000 interview it concern a reconstruction of about 10 years ago. Figure 1 provides a graphical illustration of the information collected in the two interviews and how these overlap.

[Figure 1 about here]

Having employment biographies from two retrospective interviews with an overlapping period may generate inconsistencies between reports. Such inconsistencies could be attributed to inconsistent autobiographical reconstructions and to the greater recall

---

[3] Although the period reconstructed is relatively short, the levels of mobility on the Swedish labor market allow us to obtain a sufficiently large number of transitions for which memory error can be observed (Blomskog, 1997). The labor market is characterized by intermediate levels of job mobility, predominantly upward in direction (DiPrete, 2002), weak occupational boundaries and organization-based labor market boundaries causing jobs to be less stable compared to other countries, and firms to resort more quickly to layoffs—even of experienced workers—as their way of adjustment (Björklund and Holmlund 1987; Büchtemann 1993; Grubb and Wells 1993; OECD 1994).

difficulties in the second interview since about 10 years have elapsed in 2000 for the recall of events in 1990/1991. This might, on the one hand, be somewhat problematic because merging the histories to obtain the complete history should deal in some way with the inconsistencies between the two sources. On the other hand, it also offers a unique opportunity to study the quality of retrospective reports, allowing us to investigate whether and how recall patterns differ with different length of the recall period.

3. POSSIBLE SOURCES OF MEASUREMENT ERROR IN RETROSPECTIVE REPORTS

In line with the definition of the concept of reliability itself, it seems reasonable to assume that the more similar two reports (for the same respondents and period) are, the less measurement error they contain (De Graaf, 1989). When the reports differ, at least one of the two contains a mistake in the reconstruction of the episode concerned. Measurement errors in respondents' careers may originate from poor autobiographical reconstruction (Reimer, 2004). Recall bias might endanger the reliability of retrospective data, due to involuntary memory lapses but also to a voluntary modification of events (Powers et al, 1978). Respondents' lack of memory accuracy may consequently produce an artificial sequence of episodes, which is different from the real life-course sequence. Furthermore, respondents may report their employment biographies differently when asked at different moments in their life and at different distance from the occurrence of the events (Babbie, 1973; Sudman and Bradburn, 1973).

During the recall process in the data collection, respondents have to reconstruct the full sequence of subsequent episodes in their life course in chronological order, labeling episodes, describing their main activity, and defining transitions between episodes by providing the month and year in which each episode ends. This occurs through the subjective reconstruction by the respondent, which might be affected by (and adapted to) normative considerations about the logic of a life-course and the individual's self perception of that at the time of recall. Errors may arise when respondents recall their status at a given point in the past and when they reconstruct a transition. They may completely forget episodes but also erroneously add some, artificially increasing the number of transitions. Memory's tendency towards reduction in complexity is usually expected to lead to transitions being more often forgotten than added. Since in our data the reference period is limited in time (i.e. it may run from January 1990 until the interview month in 1991), misdating the end of an episode could mean that the whole episode (and the transition between episodes) is moved to a date before or after the reference period. Incorrectly dating the occurrence of an event may lead it to being reported as having occurred before or after our observation window (i.e. before January 1990 or after the interview in 1991), which would lead to the loss of episodes within our reference period.

A variety of reasons might lead respondents to intentionally or unintentionally give an inaccurate account of their work histories, and such reasons might act in opposite directions (i.e. overreport or underreport of certain states) and follow different mechanisms at different recall distances. In the case of unemployment, for example, although it is likely to mark a profound change in a person's life,,findings from survey research show that episodes have a particularly high risk of being forgotten (Elias, 1997; Paull, 1997; Dex, 1991; Reimer, 2004). Different arguments run for or against it: while shame could lead someone not to admit being unemployed, adverse events which happened in the past, and from which the respondent recovered,might be reported more easily. Also, suspicions about the privacy of the information and fear to lose benefits for example might make respondent reluctant to report unemployment episodes. On the other hand, it might be more difficult to lie about the current situation. Overall, we could think of different reasons leading to misreports in both the directions at both occasions, and in no systematic way. For this reasons, we find it reasonable to assume that intentional misreport does not lead to errors in one specific direction.

If the same respondents are re-interviewed and no changes occurred in the survey format, which is the case in our data, the sample as well as questionnaire format cannot be an explanation for the differences encountered at the two interviews. What is, however, different in the two sources is the length of the recall period (about 1 year against about 10 years). Given the assumption that memory tends to decay over time, it can be assumed that recall difficulties were much larger at the second interview, in which respondents reported their employment histories which occurred about 10 years before. Therefore, it seems reasonable to be more confident about the 1991 report.[4] However, differences between the

[4] It should also be taken into consideration that 83% of the months reported in 1991 are part of the current spell at the interview date. This makes it more difficult to skip reporting that very spell and may also be seen as a reason for more accuracy in the 1991 report. Looking at the posterior estimates we indeed find that months which are not part of the current spell at the interview date in 1991 contain much more error than months which are part of the spell hold at the interview date (23% against 6% of error at the person-month level).

two interviews cannot be fully explained by the larger memory bias in the 2000 report. For example, the conditions under which the interview took place may differ and, for some respondents, be less favorable in the 1991 interview, and also in the 1991 report there may be mistakes in the data coding. In other words, both interviews may contain errors, and it is thus not correct to treat the 1991 report as the gold standard without errors.

4. ANALYSIS OF THE LEVEL OF LIVING DATA SET

All our analyses are based on the overlapping period of retrospective work histories described in section 2.[5] The analyses are based on the reports of 1973 respondents, 47% of which are men and 53% women. For these respondents, the 1991 and 2000 measurements 2000 contained between 3 and 22 months of overlap, on average 15 months, during the period starting in January 1990 (the earliest possible date) and November 1991 (the latest recorded interview date in 1991).[6] The average length of the recall distance is 8 months for the 1991 survey and almost 10 years for the 2000 survey. All our analyses are based on a distinction of four employment states: employed (E), self-employed (S), not employed (N) or unemployed (U).

[5] We checked whether attrition affects the representativeness of our sample, but this does not seem to be the case: our sample composition does not differ from the original sample in terms of age, gender, education, and career complexity.

[6] A few respondents reported their activity only from a date later than January 1990. Consequently, for these respondents we have information on less than 13 months.

In section 4.1, we provide a descriptive analysis of the two event sequences. In section 4.2,.we quantify the difference between the reported individual career sequences at the two interviews using matching techniques,[7] and link the dissimilarity measures to the career and personal characteristics. Next, in section 4.3, we move to a model-based approach using latent Markov (LM) models.

4.1 Description and comparison of the 1991 and 2000 sequences

Table 1 provides information on the number of reported episodes per state for the two interviews. It can be seen that the number of episodes reported is much smaller in the 2000 than in the 1991 interview. The difference is largest for unemployment episodes (43%). However, rather than looking at episodes, it is more informative to look at descriptive characteristics of the full career trajectories, that is at the full sequence of episodes in the biography of each respondent during the period of interest.

[Table 1 and 2 about here]

In Table 2 we provide such a descriptive comparison of the sequences in the two reports. Employment careers appear less heterogeneous according to the report in the second interview (2000) than in the first (1991): fewer different elements (i.e. states) and fewer

[7] Huber and Schmucker (2009) recently adopted the same approach to identify types and determinants of inconsistencies between administrative and survey data on employment cycles.

episodes seem to be reported on average in a career sequence when the reconstruction took place longer after the occurrence of the events (in 2000).

[Table 3 about here]

Table 3 presents a two-way cross-tabulation of the type of sequence reported at the two occasions. The least complex sequences consisting of one episode appear the ones most likely to be in agreement across the two interviews, showing also the similarity to be largest for employment (E). Given that the sequences are ordered from the least to the most complex,[8] the larger frequencies under the main diagonal indicate that the reported sequences in the 2000 interview tend to be simpler than those in the 1991 interview. The 2000 reports contain fewer changes and simpler careers.

These findings are particularly relevant given the current debate about the increasing (in)stability on the labor market. Our results suggest that the earlier part of the career patterns tend to be simplified (reporting fewer transitions in earlier life than there were in reality) in retrospective reports, which confirms the contended trend, but simply due to measurement error.[9] To further address the issue of career instability, we use a recently developed measure of (career) 'complexity' (Elzinga, 2009), also referred to as

[8] The order in which sequences are presented is based on the number of different episodes in the trajectories and their frequency.

[9] If increasing length of the recall period leads to more simplified careers, career volatility is underestimated especially for the very past and less for the time closer to the interview, which would overestimate the trend towards increasing instability over time.

'turbulence' (Elzinga, 2007; Elzinga and Liefbroer, 2007), which quantifies the number of employment changes, the predictability of the order of jobs, and the variability of the durations spent in different employment states. Elzinga's measure takes into account the order of career states as well as their duration variation in different career states.[10]

[Table 4 about here]

Table 4 reports the mean turbulence for the two reports,[11] as well as their bootstrap confidence intervals (Efron and Tibshirani, 1993). As it can be seen, the average turbulence is larger for the 1991 than for the 2000 report, and the difference between these two means is significant since the bootstrap confidence intervals show no overlap. This indicates again that the 2000 report may underestimate the actual career complexity. This finding is relevant because it shows that the length of the recall period may affect the observed turbulence yielding biased results on changes in employment statuses over the life course.

[10] As a measure of turbulence of sequence x of duration t, Elzinga (2007: 33) proposed using $0 \le T(x,t) = \log_2\left(\phi(x)\frac{s_{t,\max}^2+1}{s_t^2+1}\right)$, with $\phi(x)$ denoting the number of distinct subsequences, s_t^2 denoting the variance of state durations and $s_{t,\max}^2$ the maximum of that variance given the total duration of the sequence. *T(x, t)* is therefore a sequence property quantified such that the sequencing of the states and the variance of their durations is taken into account; it is not sensitive to the specific time scale used and it increases with decreasing variance of the durations of the states (Elzinga and Liefbroer, 2007).

[11] Whereas in Table 4 we present the results based on this turbulence measure, we also computed it without accounting for duration (Elzinga, 2007: 33). Although the absolute values of the measure differ, the substantive results on differences in the two reports remain the same.

4.2 Quantifying sequence similarities using optimal matching techniques

So far, we looked at aggregated sequence patterns and aggregated differences between the 1991 and 2000 sequences. In the next step, we want to quantify the amount of measurement error in the sequences at the level of the individual. For the moment, we assume that the dissimilarity between the two sequences is a good measure for the amount of measurement error they contain. One can quantify differences between pairs of sequences using specific algorithmic approaches. The obtained dissimilarity measure can subsequently be used as dependent variable in a regression analysis to investigate whether differences are associated with respondent and career characteristics.

Various methods have been developed for quantifying the dissimilarity between pairs of sequences, of which the most widely used and best known in the social sciences is optimal matching (OM) (Abbott, 1995; Abbott and Forrest, 1986). This method compares sequences using the Levenshtein distance, a measure based on the minimal cost (i.e., the most efficient set of operations) to transform one string into another using a series of "elementary operations" of insertion, deletion, and substitution. The Needleman-Wunsch algorithm uses this distance measure to calculate the dissimilarity or alignment between the pairs (for a more detailed explanation of OM see Abbott, 1995; Abbott and Forrest, 1986; Abbott and Tsay, 2000).

In OM, the costs associated with each of the transformation operations have to be defined, implying assumptions about the relative cost of substituting, inserting and deleting. Especially the specification of the substitution costs between pairs of elements (which in our application are the four labor market states) may be complicated because it involves quantifying the dissimilarity between the elements (states).[12] We used the default settings of the Stata OM routine that we applied in our analysis, which implied using a substitution cost matrix treating all the states as equally dissimilar and assuming substitution costs to be twice as large as the cost associated with inserting or deleting elements.

[Table 5 about here]

We first employed optimal matching to quantify the dissimilarity between the pair of sequences for each respondent and subsequently regressed the dissimilarity on respondent and career characteristics. The first column of Table 5 contains the results of the regression analysis. As it can be seen, gender and age[13] are not significantly related to sequence dissimilarity. Career complexity and the occurrence of episodes of

---

[12] This step in OM is always arbitrary to a certain extent and in the social sciences a theory is not always sufficiently precise to inform us about appropriate cost specification, which may cast worries that there is no theoretical basis to determine the substitution costs matrix and therefore no way to justify the analysis. Theoretical considerations about the similarity between elements may guide the definition and differentiation of substitution costs. Alternatively, it has been proposed that they are fully derived from the data themselves based on the frequency of transitions from one state to another, with lower costs for high transition frequency, based on the intuition that two statuses are less different when there are more transitions between them. We do not use such a specification of the substitution cost matrix because we consider, for example, employment more similar to self-employment than to not employment, although we observe fewer transitions in the first case.

[13] Being the time of the interview fixed, birth cohort also indicates the age at interview, where later cohorts - i.e. increasing birth year- are younger at the time of interview.

unemployment,[14] instead, have strong positive effects on sequence dissimilarity, indicating that recall accuracy is lower for respondents with more transitions and with episodes of unemployment. The significance of both the terms indicates that each has an effect controlling for the other, which means that episodes of unemployment decrease recall accuracy independent of the effect of unemployment on raising the career complexity. Trusting the first report (1991) more, these results also suggest that in such circumstances recall is more problematic when the recall period is longer.[15]

We acknowledge that OM has been subject to criticism (Wu, 2000), especially for the way it (fails to) handle duration and because it is not able to account for the order of events and the direction of change over time. Some aspects in OM might be particularly problematic. The substitution cost matrix in OM has to be symmetric which, in our case, means that reporting being employed when actually being unemployed is implicitly considered to be equally wrong to reporting of being unemployed when actually being employed. Moreover, OM does not handle right censoring, neglects nonlinear time dependencies and lacks concern with covariates (Halpin, 2003).

---

[14] Career complexity was operationalized as the number of episodes according to the 1991 report and standardized to the number of months that are observed calculating the degree to which a respondent's number of episodes deviates from the average number of episodes in his/her 'length of observation group', where three groups were separated according to the time for which the respondent was observed.

[15] We also included education in the model (the results of which are available from the first author upon request), but its effect was found insignificant. Due to a slight loss of power resulting from a few cases with missing information on education, we preferred presenting the result of the model without education. However, the effects of the included variables remained very similar.

Several alternatives[16] to the standard OM have been proposed. Yet, our results turned out to be robust not only to different specifications of the substitution cost matrix[17] but also to the specific technique used to compare the two sets of sequences. To demonstrate this robustness, the second column of Table 5 provides the results of a regression analysis using a recently developed distance measure by Elzinga (2007), which is based on a non-alignment technique that accounts for the order of events within sequences and which avoids arbitrary decisions on transformation costs since it does not use transformation operations.[18]

The algorithmic matching approaches differ from formal statistical models in that they do not require specifying the process that generated the data. As a result, they can only be used for description and exploration, and thus not for prediction and explanation. Though stochastic variants for testing hypotheses may be developed in the future, there is still a long way to get there (Biemann et al., 2009). In contrast, most sequences of interest

[16] Some alternatives for standard OM are: *OMAV*, an optimal matching technique adjusting for duration; *Degenne*, a method focusing on comparing vectors of cumulated duration at each time point; *Hamming* distance, which compares sequences element by element such that the inter-sequence distance is the sum of the element-wise distances; *Elzinga*'s technique, which calculates the similarity between the sequences based on identical subsequences.

[17] The results of such analyses are not shown here but are available from the first author upon request.

[18] This technique is based on a different metric and handles durations in a more consistent way: it calculates the similarity between the sequences based on identical subsequences. Distance is defined as the number of non-common subsequences, weighted by their occurrence frequency in either sequence and their duration. It has advantages concerning duration and the importance of the order of non-adjacent states. The algorithm consists in an efficient way of counting the number of common n-tuples in pairs of sequences.

develop stochastically over time, where, for example, earlier events affect the probability of occurrence of later events (path dependency). That is, for most sequences we will be more interested in the processes that created them than their actual occurrence over time (Halpin, 2003: 9). Next, we describe a model-based approach that allows specifying the underlying process that generated the observed sequence data.

4.3 Modeling sequences using latent Markov models

To overcome several of the shortcomings of the algorithmic approach presented above, we propose investigating the recall bias problem using a model-based approach. The two key requirements of the stochastic model for modeling the two sets of reports are that it should be suited (1) for modeling event sequences or discrete-time longitudinal data on transitions across a finite number of states and (2) for dealing with measurement error in the recorded states, which in our application also involves combining the possibly inconsistent information in the 1991 and the 2000 reports of the month-specific labor market status.

A suitable stochastic model is the Markov model, which yields estimates for the probability of making a transition between each pair of states of the outcome variable of interest. While a simple Markov model solves the problem associated with the OM approach that adjacent time points are treated as somewhat disconnected, it does not tackle the measurement error issue discussed above. However, there exists a more extended variant of the Markov model called the latent (or hidden) Markov model that can be used for this purpose. Here, "latent" or "hidden" refers to the fact that a person's true state at time point t is unknown, and it should thus be treated as a latent variable. In a latent Markov model, the true states are the categories (classes) of a dynamic latent (unobservable) variable. The latent states are connected to the observed states using a model similar to a latent class model, which defines the measurement error component of the latent Markov model.

There exist various examples of applications of latent Markov models for the analysis of discrete-time longitudinal data subject to measurement error (Collins and Wugalter, 1992; Hagenaars, 1990; Poulsen, 1982; Van de Pol and Langeheine, 1990; Breen and Moisio, 2004; Vermunt, Tran and Magidson, 2008; Magidson, Vermunt and Tran, 2009). In our application, we use the latent Markov model to estimate the error in the recorded employment status at the two interviews, which we believe will yield more interesting information on the pattern and amount of recall bias in the two retrospective reports than was obtained with the descriptive and algorithmic approaches. It should be noted that the latent Markov model not only yields estimates of transition and measurement error probabilities, but that it can also be used to obtain (posterior mode) estimates of the true state occupied at each time point. Hence, whereas in OM we could only compare the distance of the 2000 report to the 1991 report, which was in fact used as a kind of perfect measure, the latent Markov approach allows us to compare each of these two reports with the estimated true sequences, and thus to relax the assumption that the information from the 1991 interview is free from errors. Typically, latent or hidden Markov models are applied with a single report per respondent per time point. However, our data consist of two (imperfect) reports of a respondent’s employment status per time point (per person-month). As it is explained in more detail below, this makes it possible to relax certain key assumptions of the basic latent Markov model which are unrealistic in our application.

We denote the 1991 and 2000 report of the employment status of person i at month t by y_{it}^{1991} and y_{it}^{2000}, respectively. Note that y_{it}^{1991} and y_{it}^{2000} are categorical variables with M=4 categories (1=employed (E), 2=self-employed (S), 3=not employed (N), 4=unemployed (U)); that is, $1\leq y_{it} \leq M$. The total number of time points (months) for which person i provides information is $T_i + 1$, with $0 \leq t \leq T_i$.[19] The observed data for a particular respondent consist of two response vectors of length $T_i + 1$ denoted by $\mathbf{y}_i^{1991}$ and $\mathbf{y}_i^{2000}$, respectively. The probability of having the observed set of responses is denoted by $P(\mathbf{y}_i^{1991}, \mathbf{y}_i^{2000})$. Note that $P(\mathbf{y}_i^{1991}, \mathbf{y}_i^{2000})$ is the probability for which we are going to define a statistical model.

Besides the observed responses, the latent Markov model for $P(\mathbf{y}_i^{1991}, \mathbf{y}_i^{2000})$ contains $T_i + 1$ discrete latent variables, each having K categories or latent states. We denote the true (latent) state at time point t by x_t, where $1 \leq x_t \leq K$. In the current study, K=M=4; that is, the number of latent labor force states is assumed to be equal to the number of observed labor force states. The latent Markov model has the following form:

$$P\left(\mathbf{y}_i^{1991}, \mathbf{y}_i^{2000}\right) = \sum_{x_0=1}^{K}\sum_{x_1=1}^{K}\ldots\sum_{x_T=1}^{K} P(x_0)\left[\prod_{t=1}^{T_i} P\left(x_t \mid x_{t-1}\right)\right]\left[\prod_{t=1}^{T_i} P\left(y_{it}^{1991} \mid x_t\right)\prod_{t=1}^{T_i} P\left(y_{it}^{2000} \mid x_t\right)\right]. \quad (1)$$

The unknown model probabilities to be estimated are the initial latent state probabilities $P(x_0=s)$, the latent transition probabilities $P(x_t = r \mid x_{t-1}=s)$, and the classification error

[19] The number of months for which we have employment information varies across respondents, which is why we use the index i.

probabilities $P(y_{it}^{1991}=\ell \mid x_t=s)$ and $P(y_{it}^{2000}=\ell \mid x_t=s)$, were s and r refer to a particular latent state and ℓ to a particular observed state. In the Latent GOLD software (Vermunt and Magidson, 2008) that we used to estimate our models, the model probabilities are parameterized using multinomial logistic equations; i.e.

$$\mathrm{P}\left(x_0 = s\right) = \frac{\exp\left(\alpha_s\right)}{\sum_{k=1}^{K}\exp\left(\alpha_k\right)} \tag{2}$$

$$\mathrm{P}\left(x_t = r \mid x_{t-1} = s\right) = \frac{\exp\left(\gamma_{rs}\right)}{\sum_{k=1}^{K}\exp\left(\gamma_{ks}\right)} \tag{3}$$

$$\mathrm{P}\left(y_{it}^{1991} = \ell \mid x_t = s\right) = \frac{\exp\left(\beta_{\ell s}^{1991}\right)}{\sum_{m=1}^{M}\exp\left(\beta_{ms}^{1991}\right)} \tag{4}$$

$$\mathrm{P}\left(y_{it}^{2000} = \ell \mid x_t = s\right) = \frac{\exp\left(\beta_{\ell s}^{2000}\right)}{\sum_{m=1}^{M}\exp\left(\beta_{ms}^{2000}\right)}. \tag{5}$$

Under the identifying restrictions[20] $\alpha_1 = 0$ and $\gamma_{11} = \gamma_{22} = \gamma_{33} = \gamma_{44} = 0$ and $\beta_{11} = \beta_{22} = \beta_{33} = \beta_{44} = 0$, we get the following rather straightforward interpretation of the logit parameters:

$$\alpha_s = \log\frac{\mathrm{P}\left(x_0 = s\right)}{\mathrm{P}\left(x_0 = 1\right)} \tag{6}$$

$$\gamma_{rs} = \log\frac{\mathrm{P}\left(x_t = r \mid x_{t-1} = s\right)}{\mathrm{P}\left(x_t = s \mid x_{t-1} = s\right)} \tag{7}$$

$$\beta_{\ell s}^{1991} = \log\frac{\mathrm{P}\left(y_{it}^{1991} = \ell \mid x_t = s\right)}{\mathrm{P}\left(y_{it}^{1991} = s \mid x_t = s\right)} \tag{8}$$

$$\beta_{\ell s}^{2000} = \log\frac{\mathrm{P}\left(y_{it}^{2000} = \ell \mid x_t = s\right)}{\mathrm{P}\left(y_{it}^{2000} = s \mid x_t = s\right)}; \tag{9}$$

[20] Because these are multinomial logit models, we need one constraint on α, four on γ (one for each value of x_{t-1}) and four on β (one for each value of x); that is, one category of the dependent variable should be treated as the reference category. In our "transition coding" and "error coding" schemes, the reference category changes with the value of the independent variable in the equation, giving the parameters an interpretation that makes very much sense in latent Markov models.

that is, as the log odds of having initial state s rather than 1 (Equation 6), the log odds of making a transition from state s to state r rather than staying in state s ('transition coding'; Equation 7), and the log odds of misclassifying someone with latent state s by assigning observed state ℓ rather than the correct state s ('error coding', Equation 8 for the 1991 response and Equation 9 for the 2000 response). Note that Equations (7), (8), and (9) follow from the fact that the "no transition" and the "no error" categories, respectively, serve as the *reference* categories; that is, from the fact that the reference category changes with the origin state (true state).

The model described in Equation (1) is the basic latent Markov model. It makes certain assumptions about 1) the way people move between latent states (the transitions) and 2) the way the observed states are reported given the true states (the measurement error). The main assumption about the (latent) transitions is that these are in agreement with a first-order Markov process, which implies that the state occupied at time point t depends only on the state occupied at time point t-1 (the previous state), and that it is thus independent of the states occupied at all other time points. In addition, it is assumed that transition probabilities are time homogeneous and that there are no observed or unobserved factors affecting the transition probabilities. Although it would be possible to relax each of these sets of assumptions, we keep the model for the transitions as simple as possible since our main interest is not in the parameters of a transition model but in the measurement error parameters, as well as in posterior estimates of the respondents' true states. It turns out that this information of main interest is not (or very weakly) affected by the model we use for the transitions. This is because we have two measurements of a person's true state (and not just one) and because we have a rather long time series. We obtain very good (almost perfect) estimates for the true event sequences, which cannot be improved by a more sophisticated transition model. Relaxing the first-order Markov assumption, adding covariates (including time) in the model, and modeling unobserved heterogeneity using a time constant latent variable will improve the prior estimates of the transition probabilities for a specific respondent, but will not affect the estimates of the measurement error nor of the posterior estimates of a person's true state. Therefore, we decided to keep our models as simple as possible as far as the transition part of the model is concerned.

The two key assumptions in the measurement part of the basic latent Markov model are that, conditional on a person's true states, the 1991 and 2000 responses at time point t are 1) independent of one another and 2) independent of the responses at the other time points. In latent class analysis, the former assumption - which is connected to the fact that we have two measurements - is referred to as "local independence" (Lazarsfeld and Henry, 1968; Goodman, 1974). The second assumption is sometimes referred to as "independent classification errors" (ICE; Bassi et al., 2000; Biemer and Bushery, 2000). It is especially the ICE assumption that turns out to be problematic in our application, which is why we propose various modifications of the basic latent Markov model for relaxing this assumption. Similar kinds of latent Markov models with correlated measurement errors were proposed by Hagenaars (1988) for the analysis of two-wave panel data and by Bassi et al. (2000) for the analysis of a single employment history collected with a rotation panel design.

[Figure 2 about here]

[Table 6 about here]

In our application, we used three kinds of specifications for the measurement error component of the latent Markov model. Figure 2 provides a graphical representation of the models. Table 6 presents their fit measures (log-likelihood value and BIC), as well as their measurement errors estimates.

Model 1 is a standard latent Markov model assuming ICE. In addition, it assumes that the response at the 1991 interview is perfect.[21] Note that such a no measurement error specification for the 1991 response is similar to the implicit model used by the OM approach, which quantified how well the 2000 sequences match the 1991 sequences. The first set of probabilities reported in Table 6 represents the estimates of the probability of being in a particular latent state averaged across time points. The other ones are the response probabilities, which sum to 1 within columns. The results for Model 1 show that the estimated probability of being truly employed is .75. For this state the estimated measurement error in the 2000 report is rather small (less than 5%). The self-employed and non-employed states (containing 7 and 17% of the person periods) have much larger measurement errors (22% and 19%). For the unemployment state (estimated to contain 1.5% of the person months), the measurement error is extremely large with a misclassification rate of 69%.

[21] Note that the probability of making a classification error can be fixed to zero by equating each of the logit parameters $\beta_{\ell s}^{1991}$ appearing in Equations (4) and (8) to a large negative value (e.g. minus 100) for $\ell \neq s$.

Model 2 is also an ICE model, but it relaxes the assumption that the 1991 response is perfect. As indicated by its BIC value, this model should be preferred over Model 1. However, as it can be seen from the measurement error probabilities, it yields the very unrealistic result that the 2000 report is much better than the 1991 report. This is due to an artifact of the basic latent Markov model, which sees the larger stability of the 2000 series as an indication that there is less measurement error. This problem, which is caused by the ICE assumption, becomes visible when both the 1991 and 2000 measurement error probabilities are estimated freely. It can be solved by relaxing the ICE assumption, which is feasible in our application because we have two measurements of a person's state at each occasion.

The reason why the ICE assumption is unrealistic in our application relates to the fact that in a retrospective survey, during the recall process, a respondent could forget or misplace episodes, which is much more likely to happen with increasing recall time. The consequence of such mistakes is that the same mismatch between true and observed state (the same type of error) occurs at consecutive months. As a result, errors at consecutive time points are not independent of one another, and the ICE assumption is thus clearly invalid. Although it is unlikely that such correlated errors are fully absent in the 1991 report, it is reasonable to assume that these are much weaker when reporting over the previous year than when reporting about approximately ten years before. If we were to allow dependent classification errors for both responses, we would run into the risk of fully distorting the structure of the latent Markov model, as well as of running into identification problems. Therefore, we relaxed the ICE assumption only for the 2000 report.

Model 3 relaxes the ICE assumption by including dependencies between errors occurring across months in the 2000 report. More specifically, we allow the probability of a particular type of error at month *t* to depend on the type of error made at month *t*-1. This requires that y_{it}^{2000} depends on both the latent and the observed state at *t*-1. This is achieved by replacing the term $P(y_{it}^{2000} = \ell \mid x_t = s)$ appearing in Equation (1) with the term $P(y_{it}^{2000} = \ell \mid x_t = s, x_{t-1} = r, y_{i,t-1}^{2000} = k)$. Given that *K*=*M*=4, introducing both the lagged latent state and the lagged 2000 reported state (x_{t-1} and y_{it-1}^{2000}) into the measurement error part of the latent Markov model implies that there are 16 (4 times 4) matrices of error probabilities, one for each possible combination of lagged observed and latent state. On the one hand, this shows that this model allows large flexibility. On the other hand, while it is possible to estimate all 16 sets of error probabilities freely, this may yield many parameters possibly difficult to interpret. We therefore opt for imposing more structure on the measurement error parameters for the 2000 report. More specifically, we assume that there are two sets of error probabilities (instead of 16), one for the situation in which the *t*-1 report was correct and one for the situation in which the *t*-1 report was incorrect. Such a more restricted structure can be easily defined using a logit parameterization of the error probabilities; that is,

$$\log \frac{P(y_{it}^{2000} = \ell \mid x_t = s, x_{t-1} = r, y_{i,t-1}^{2000} = k)}{P(y_{it}^{2000} = s \mid x_t = s, x_{t-1} = r, y_{i,t-1}^{2000} = k)} = \beta_{\ell s}^{2000} \delta_{rk} + \lambda_{\ell s}(1 - \delta_{rk}). \qquad (10)$$

Here, δ_{rk} is an indicator variable (a dummy) taking on the value 1 if *k*=*r* and 0 otherwise. Hence, compared to the ICE model, we have two sets of parameters, one for the situation in which the report at t-1 was correct ($\delta_{rk} = 1$) and one for the situation in which it was incorrect ($\delta_{rk} = 0$). As in the ICE model, for identification, we set $\beta_{\ell s}^{2000} = \lambda_{\ell s} = 0$ for $\ell = s$. This yields 2 sets of 12 measurement error coefficients.

Model 3 has a much lower BIC value than Model 2, showing that including dependencies between classification errors improves the model fit. The estimated measurement errors for the 1991 response show that there is basically no error for the employment and self-employment states, very small error for the not-in-employment state, and much larger error (51%) for the unemployment state. The much larger measurement error in the unemployment state is well documented in the literature (Magidson, Vermunt and Tran, 2009). The estimated measurement error probabilities show that the error at the second interview is very small if a respondent correctly reported the state at the previous time point. For unemployment though, the misreport is larger than for the other states (about .19), also when no error in the lag occurred. When any type of classification error occurred at the previous time point instead, much more errors occur at the current time point. This applies to all four true states. In this situation, self-employment is mainly confused with employment and not employment and unemployment with employment and self-employment, while employment and not in employment are often reversed.

An interesting result is that when no error occurred at t-1 (which is the case for 92% of the person-months, with more than 82% of the respondents who never made a mistake at all),[22] the quality of the response at the second, more distant, interview (2000) is similar to

[22] We are able to identify errors comparing the observed responses with the posterior mode estimates at each time point t.

the one of the first interview (1991), and sometimes even better (error for unemployment is smaller). On the other hand, when error occurred at a previous time point, the quality of the response greatly deteriorates (in more than 91% of the person-months with an error at *t*-1 errors occur at *t* as well; in terms of respondents, more than the 99% of those who never made a mistake at *t*-1 made a mistake at *t*). These results confirm that this model specification makes sense: after filtering out the bad cases, the quality of the 1991 and 2000 reports are similar, which means for the shorter and longer recall periods.

The specific constraints we impose on the $P(y_{it}^{2000} = \ell \mid x_t = s, x_{t-1} = r, y_{i,t-1}^{2000} = k)$ represent one of the many possible restrictions that can be used. However, it turned out that the results of main interest are robust to the exact specification that is used. One of the other specifications we tried, rather than distinguishing reports on and off the main diagonal at *t*-1 (no error versus any error), looks at the type of mistake and in particular at whether at the previous time point the respondent made the same mistake. The underlying idea is that a particular type of mistake is more likely to occur if the same mistake occurred at the previous time point: the log odds of a mistake at *t* depend on having had the same mistake at *t*-1. This specification is obtained with a logit model of the form $\beta_{\ell s}^{2000} + \lambda_{\ell s}\delta_{\ell srk}$, where $\delta_{\ell srk} = 1$ if $\ell \neq s \,\&\, \ell = k \,\&\, s = r$, and 0 otherwise.

After estimating the parameters of the latent Markov model of interest, one can obtain estimates of the true state occupied by individual *i* at time point *t* using the parameter estimates and the observed responses y_i^{1991} and y_i^{2000}. These estimates are based on the posterior class membership probabilities $P(x_t = s \mid y_i^{1991}, y_i^{2000})$, where, as in a standard latent class analysis, a subject is usually assigned to the class with the largest probability (Goodman, 1974). The estimated proportion of correct classification is obtained as the average of $\max_{s} P(x_t = s \mid y_i^{1991}, y_i^{2000})$ across person-months, a number which turned out to be extremely large in our application. That is, we obtain an estimated proportion of correct true state predictions of 99.8%.

We used the true state predictions for a further analysis in which this estimated sequence of true states is compared with the two observed sequences. Note that this represents an improvement compared to the OM approach in which we quantified how well the 2000 sequence matches the 1991 sequence, which means that we implicitly assumed that the 1991 sequence is free of errors. Tables 7 to 9 provide the relevant results obtained by comparing the estimated true states according to Model 3 with both the 1991 and 2000 sequences.

[Table 7 about here]

In Table 7,[23] we present an OM-based dissimilarity measure between the observed and the estimated true sequences, obtained in the same way as when we compared the two observed sequences (see results in Table 5). While in the latter case the observed response in 1991 was used as a reference, we now take the true states as reference, and we compare both the 1991 and 2000 responses to it. As it can be seen in Table 7, the average distance to the true

[23] The distance measures presented in Table 7 are based on the 'default' substitution cost matrix as defined in subsection 4.2.

sequence is larger in the 2000 than in 1991 report and, as indicated by the bootstrap confidence interval, this difference is statistically significant. Moreover, comparison of these numbers with the estimated distance between the two observed sequences (in 1991 and 2000) shows that the mismatch of 2000 sequences with the true sequence is smaller than its mismatch with the 1991 sequence. This indicates that the amount of recall bias in the 2000 interview is overestimated when it is assessed based on a comparison with another better but still imperfect observed sequence.

[Table 8 about here]

Similarly to what we did in subsection 4.2 (Table 5), we used regression analysis to assess the effect of respondent and career characteristics on the quality of the reports. Results are shown in Table 8. In the first column, we show the results of a regression analysis on the full set of sequence distances, where we look at the overall effect of the covariates of interest, controlling for the source of observed response, i.e. whether the dissimilarity with the latent sequence refers to the 1991 or 2000 report. We find that the complexity of the career and the presence of episodes of unemployment have a significant effect, increasing the amount of errors. However, their effects seem different at different recall times. The last column shows whether the effects of the covariates are significantly different in the two equations.[24] Controlling for other differences, episodes of unemployment increase report

[24] The significance of the differences was assessed in a simultaneous analysis of both reports using a model including report-covariate interactions. The Stata cluster option was used to correct the standard error for the fact that the 1991-latent and 2000-latent distances are dependent observations.

errors to a significantly greater extent for short recall distances (1991 interview), although their effect is positive and strongly significant also for longer recall distances (2000). Career complexity, instead, seems to make the quality of the report worse only with longer recall times.

[Table 9 about here]

Table 9 provides the last results we would like to discuss. For both the 1991 and 2000 interview, it presents the two-way cross-tabulation of the type of observed sequence with estimated true sequence. These tables show that sequences reported at the first interview (1991) are closer to the latent sequences than sequences from 2000. However this seems to be especially the case for careers involving (more) transitions. These results show again that comparing the two observed responses with one another instead of with the true sequence overestimates the amount of recall bias.

5. CONCLUSIONS

This article discussed methods for studying memory bias in retrospective reports when multiple reports are available for the period of interest. In addition to descriptive and exploratory matching techniques, we proposed using a model-based approach which makes use of latent Markov models. The main advantage of the latter approach is that it does not require to assume that one of the responses is perfect; that is, all reports may contain measurement errors. A novel element in the proposed latent Markov modeling approach is that it allows for dependent classification errors across adjacent time points. Moreover, after the estimation of a latent Markov model, matching techniques can be used to quantify the dissimilarity between the true and observed sequences. These dissimilarities can be used as dependent variable in a regression analysis to detect factors related to the occurrence of recall bias.

Our analyses of the Swedish Level of Living Survey showed that recall errors lead to simplified employment careers, especially for longer recall times, resulting in recorded careers with fewer episodes and transitions than they actually had. One of the consequences of this simplification is that career stability may be overestimated with retrospective data. Furthermore, we found that recall accuracy depends on career complexity and the occurrence of unemployment spells: Errors are more likely to occur in more complex careers, especially for longer recall periods, as well as in careers containing unemployment spells, also for shorter recall periods. It should be noted that unemployment is more likely to be misreported than employment (and other states) independent of spell duration. In fact, a short spell duration does not make the report of unemployment any worse, while it increases the measurement error for the other states, which seem to be mistakenly reported mainly for short(er) spell durations.[25] This shows that although unemployment spells are

[25] Results are not shown here but can be obtained from the first author upon request.

shorter and shorter spells contain more errors, unemployment and duration effects are not confounded.[26]

The described approach, which combines latent Markov modeling with matching techniques, allows us to quantify recall bias and to identify its determinants. Although this yields relevant information for studies addressing substantive research questions using retrospective life-history data, a next step would be to perform substantive analysis while correcting for measurement errors. The latent Markov modeling framework offers various possibilities to allow this. The most straightforward approach would be to build the substantive model interest for the transitions across the true states using the latent Markov model with covariates proposed by Vermunt, Langeheine, and Böckenholt (1999), and recently applied by among others (Bartolucci, Lupparelli, and Montanari, 2009; Paas, Vermunt, and Bijmolt, 2007). The transition model may also be extended by allowing for unobserved heterogeneity, for example, in the form of a mover-stayer model (Goodman, 1961, Yamaguchi, 2008).

[26] The posterior mode estimates of the true state show that average length was 9 months for unemployment spells and 16 months for employment spells. The average duration of spells containing errors was about 4 months shorter that spells without errors.

Tables and Figures

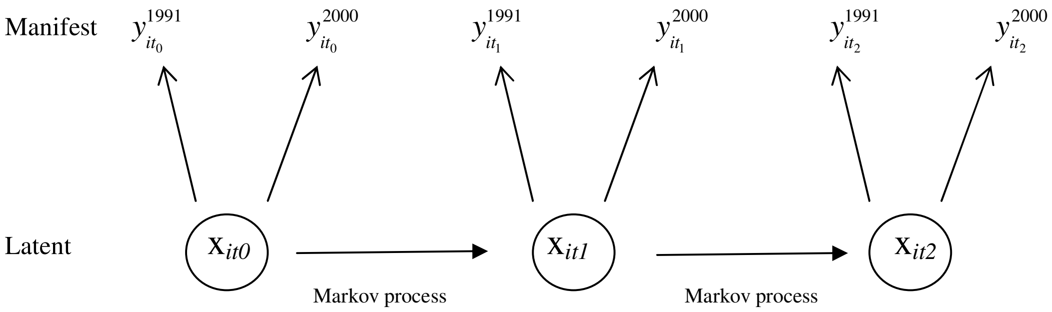
Figure 1. Representation of the data structure

year	…	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
restrospective distance																	
interview 1 (1991)	…	6	5	4	3	2	1	0									
interview 2 (2000)							10	9	8	7	6	5	4	3	2	1	0

Figure 2. Models Graphical Representation

For t=0, 1, 2, ...T

Model 1 and Model 2



Model 3

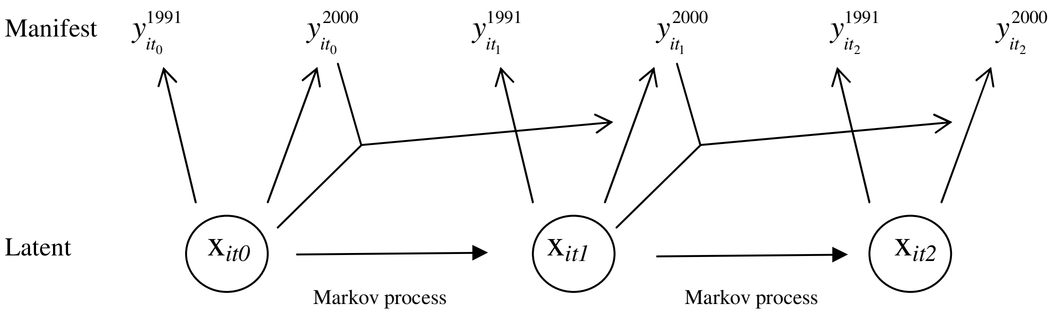


Table 1. Number of reported episodes and transitions, by interview

	1991	**2000**	% reduction in 2000
Episodes			
Employment	1732	1652	4.62
Self-Employment	155	127	18.06
Non employment	534	474	11.24
Unemployment	96	55	42.71
Total	2517	2308	8.30
Transitions	544	335	38.42

Table 2. Descriptive sequences characteristics, by interview

	1991		**2000**		% reduction in 2000
	mean	St. dev	mean	St. dev	
Number of different elements	1.21	0.44	1.15	0.377	4.96
Number of episodes	1.28	0.64	1.17	0.45	8.59
Number of employment episodes	0.88	0.44	0.84	0.41	4.55
Number of unemployment episodes	0.05	0.24	0.03	0.18	40.00
N	1973		1973		

Table 3. (Ten most common) Sequences as reported in 1991 and 2000

sequence in 1991	***sequence in 2000*** **E**	**N**	**S**	**EN**	**NE**	**EU**	**ES**	**ENE**	**EUE**	**NEN**	***else***	**N**
E	**95.5**	0.8	0.5	1.2	0.6	0.3	0.2	0.3	0.2	-	*0.4*	1282
N	5.3	**85.6**	-	5.3	2.3	-	2.7	-	-	-	*1.5*	188
S	5.4	9.0	**79.3**	-	-	-	-	-	-	-	*3.6*	111
EN	14.2	18.1	-	**59.1**	1.6	-	-	1.6	-	3.1	*2.3*	127
NE	29.3	7.3	1.2	1.2	**48.8**	-	-	4.9	-	3.7	*3.6*	82
EU	52.4	-	-	9.5	-	**19.1**	-	4.8	-	-	*14.2*	21
ES	52.9	5.9	-	-	-	5.9	**35.9**	-	-	-	-	17
ENE	54.6	-	3.0	3.0	6.1	3.0	-	**24.2**	3.0	-	*3.1*	33
EUE	64.3	-	-	-	14.3	-	-	-	**7.1**	-	*14.3*	14
NEN	-	35.7	-	35.7	14.3	-	-	-	-	**14.3**	-	14
else	*5.3*	*85.64*	-	*5.3*	*2.3*	-	-	-	-	-	*1.5*	84
N	1356	230	103	111	69	11	11	21	6	9	46	1973

Table 4. Sequence turbulence, by interview

	turbulence	
	mean	bootstrap C.I.
1991	1.698	(1.632 - 1.763)
2000	1.481	(1.417 - 1.545)

Table 5. OLS regression on 1991-2000 sequence (standardized) distance

	OM distance $^{\gamma}$	Elzinga's distance
Female	-0.001	0.006
Birth year	-0.002	-0.001
Episodes of unemployment	0.644**	0.473**
Career complexity	0.575**	0.811**
Constant	2.943	1.865
N	1973	1973
R squared	0.20	0.32

$^{\gamma}$ default subcost matrix

Table 6. Markov models results

Model 1: ICE with y1991 no error Log-likelihood=-14813; parameters=27; BIC=29831					Model 2: ICE Log-likelihood=-11850; parameters=39; BIC=23995					Model 3: non-ICE Log-likelihood=-9277; parameters=51; BIC=18940				
	true state					true state					true state			
	E	S	N	U		E	S	N	U		E	S	N	U
Size	0.75	0.07	0.17	0.02	Size	0.71	0.07	0.17	0.05	Size	0.77	0.05	0.15	0.03
y1991	E	S	N	U	y1991	E	S	N	U	y1991	E	S	N	U
E	1.00	0.00	0.00	0.00	E	1.00	0.07	0.15	0.11	E	1.00	0.00	0.02	0.35
S	0.00	1.00	0.00	0.00	S	0.00	0.88	0.00	0.00	S	0.00	1.00	0.00	0.01
N	0.00	0.00	1.00	0.00	N	0.00	0.03	0.84	0.65	N	0.00	0.00	0.98	0.15
U	0.00	0.00	0.00	1.00	U	0.00	0.01	0.01	0.24	U	0.00	0.00	0.00	0.49
y2000	E	S	N	U	y2000	E	S	N	U	y2000 without error at t-1	E	S	N	U
E	0.96	0.11	0.17	0.47	E	1.00	0.10	0.00	0.79	E	1.00	0.01	0.02	0.14
S	0.01	0.78	0.01	0.06	S	0.00	0.80	0.00	0.00	S	0.00	0.99	0.00	0.02
N	0.03	0.10	0.81	0.16	N	0.00	0.09	1.00	0.00	N	0.00	0.00	0.98	0.03
U	0.00	0.01	0.01	0.31	U	0.00	0.00	0.00	0.21	U	0.00	0.00	0.00	0.81
										y2000 with error at t-1	E	S	N	U
										E	0.23	0.44	0.80	0.39
										S	0.00	0.15	0.00	0.47
										N	0.77	0.41	0.20	0.12
										U	0.00	0.00	0.00	0.03

Table 7.OM distances

	OM distance	
	mean	bootstrap C.I.
1991-latent	0.035	(0.0268-0.0435)
2000 - latent	0.165	(0.144-0.185)
1991-2000	0.195	(0.172-0.217)

Table 8. OLS regression on the distance between latent and observed response, by interview

	latent-observed	(a) latent-1991	(b) latent-2000	sig. of (a)-(b)
Female	0.011	-0.003	0.025	
Birth year	-0.0006	0.0006*	0.0007	
Episodes of unemployment	0.394**	0.492**	0.296**	*
Career complexity	0.078**	-0.067**	0.224**	**
Later report (2000)	0.130**			
Constant	0.127	-1.147*	1.525	
N	3946	1973	1973	
R squared	0.17	0.33	0.14	

Table 9. Observed and true (most common) sequences.

true	***sequence in 2000***										
sequence	**E**	**N**	**S**	**U**	**EN**	**NE**	**EU**	**ES**	**SE**	**UE**	***else***
E	**97.0**	0.8	-	-	1.5	0.7	-	-	-	-	0.0
N	4.9	**84.7**	-	-	7.4	3.0	-	-	-	-	*0.0*
S	5.5	9.1	**80.0**	-	-	-	-	2.7	-	-	*2.7*
U	5.6	5.6	50.0	**27.8**	-	-	-	-	-	5.6	*5.5*
EN	16.7	15.2	-	-	**62.9**	5.3	-	-	-	-	*0.0*
NE	29.0	6.5	-	-	10.8	**53.8**	-	-	-	-	*0.0*
EU	49.0	-	-	-	6.1	4.1	**32.7**	4.1	-	-	*4.1*
ES	61.5	-	-	-	-	-	-	**38.5**	-	-	0.0
SE	100.0	-	-	-	-	-	-	-	-	-	0.0
UE	21.1	-	-	-	-	-	-	-	10.5	**63.2**	*5.3*
else	*11.59*	*17.39*	*21.74*	7.3	4.4	*7.25*	1.5	1.5	1.5	*2.9*	***23.18***
N	1356	230	103	5	111	79	7	11	3	14	54
true	***sequence in 1991***										
sequence	**E**	**N**	**S**	**U**	**EN**	**NE**	**EU**	**ES**	**SE**	**UE**	***else***
E	**98.6**	-	-	-	0.9	0.4	-	0.1	-	-	*0.1*
N	-	**91.1**	-	-	3.9	4.9	-	-	-	-	*0.0*
S	-	-	**100.0**	-	-	-	-	-	-	-	*0.0*
U	38.9	-	-	5.6	5.6	5.6	5.6	-	-	11.1	*27.8*
EN	-	-	-	-	**100.0**		-	-	-	-	*0.0*
NE	-	-	-	-	8.6	**90.3**	-	-	-	-	*1.1*
EU	16.3	-	-	-	-	-	**69.4**	2.0	-	-	*12.2*
ES	-	-	-	-	-	-	-	**100.0**	-	-	*0.0*
SE	-	-	-	-	-	-	-	-	**100.0**	-	*0.0*
UE	26.3	-	-	-	5.3	5.3	10.5	-	5.3	**47.4**	*0.0*
else	*10.14*	*4.35*	*1.45*	*1.45*	*7.25*	*4.35*	*1.45*	4.4	1.5	*2.9*	***60.9***
N	1282	188	111	1	166	103	37	18	8	11	48

References

Abbott, A. (1995) A comment on "Measuring the agreement between sequences", *Sociological Methods and Research, 24(2), 232–243.*

Abbott, A. and Forrest, J. (1986) Optimal matching methods for historical sequences, *Journal of Interdisciplinary History, XVI(3), 471–494*.

Abbott, A. and Tsay, A. (2000) Sequence analysis and optional matching methods in sociology, *Sociological Methods and Research, 29(1), 3–33*.

Babbie, E. (1973). *Survey Research Methods*. Belmont, California: Wadsworth Publishing Company.

Bartolucci, F., Lupparelli, M. and Montanari, G.E. (2009). Latent Markov model for longitudinal binary data : An application to the performance evaluation of nursing homes. *The Annals of Applied Statistics, 3(2), 611-636*.

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

Biemer, Paul P. and John M. Bushery (2000). On the Validity of Markov Latent Class Analysis for Estimating Classification Error in Labor Force Data. *Survey Methodology, 26, 139-152.*

Björklund, A. and Holmlund, B. (1987) Worker Displacement in Sweden: Facts and Policies. Paper presented at the American Economic Association meeting, Stockholm, December.

Blomskog, S. (1997) *Essays on the functioning of the Swedish labour market* (Vol. 27). Stockholm: Swedish Institute for Social Research.

Brzinsky-Fay, C., Kohler, U. (2001) Sequence analysis using Stata, *The Stata Journal (1)*.

Breen, R. and Moisio, P. (2004) Poverty dynamics corrected for measurement error. *Journal of Economic Inequality,* Vol. 2 (3).

Büchtemann, C.F. (1993) Employment security and labor markets. In: Büchtemann, C.F. (Ed.), Employment Security and Labor Market Behavior. International Industrial and Labor Relations Report 23. ILR Press, Ithaca.

Collins, L. M., and Wugalter, S. E. (1992) Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research, 27, 131–157*.

De Graaf, N. D., Wegener, B., and Liebig, S. (1989). Reliability of retrospective survey data: A comparison of job career and social resource variables. *Unpublished paper.*

Dex, S. (1991) The Reliability of Recall Data: A Literature Review. ISER Working Paper. Colchester, University of Essex.

DiPrete, T. (2002) Life course risks, mobility regimes, and mobility consequences: A comparison of Sweden, Germany, and the United States. *American Journal of Sociology, 108, 2, 267-309.*

Efron, B., and R. J. Tibshirani (1993) *An introduction to the bootstrap.* New York: Chapman & Hall.

Elias, P. (1997) Who forgot they were unemployed? ESRC Working Paper

Elzinga, C. H. (2007) CHES*A 2.1 User manual.* Amsterdam, Vrije Universiteit Amsterdam.

Elzinga, C. H. and Liefbroer, A. C. (2007) De-standardization and differentiation of family life trajectories of young adults: A cross-national comparison'. *European Journal of Population* 23: 225-250.

Elzinga, C.H. (2009) Complexity of Categorical Time Series, *Sociological Methods & Research, in press.*

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association, 56, 841-868.*

Goodman, L.A. (1974). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: A modified latent structure approach, *American Journal of Sociology, 79, 1179-1259.*

Grubb, D. and Wells, W. (1993).Employment Regulation and Patterns of Work in EC Countries, OECD Economic Studies, No. 21, Winter, pp. 7-58.

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

Hagenaars, J. A. (1990). *Categorical Longitudinal Data - Loglinear Analysis of Panel, Trend and Cohort Data*. Newbury Park: Sage.

Huber, M. and Schmucker, A. (2009) Identifying and Explaining Inconsistencies in Linked Administrative and Survey Data: The Case of German Employment Biographies*, Historical Social Research, 34, 3, 230-241.*

Jonsson, J. O., and Mills, C. (2001). *Cradle to grave. Life-course change in modern Sweden*: Sociology press.

Lazarsfeld, P.F., and Henry, N.W. (1968). *Latent Structure Analysis*. Boston: Houghton Mill.

Magidson, J., Vermunt, J.K., and Tran, B. (2009). Using a mixture latent Markov model to analyze longitudinal U.S. employment data involving measurement error. In: K. Shigemasu, A. Okada, T. Imaizumi, and T. Hoshino (eds.), *New Trends in Psychometrics*. Universal Academy Press, Inc. pp. 235-242.

Mayer, K. U. & Huinink, J. (1990). Alters-, Perioden- und Kohorteneffekte in der Analyse von Lebensverläufen oder: Lexis ade? In: K. U. Mayer (ed.), *Lebensverläufe und sozialer Wandel.*Opladen: Westdeutscher Verlag, pp. 442–459.

OECD (1994): 'The OECD Jobs Study: Evidence and Explanations' 2 vols, Paris: OECD.

Paull, G. (2002). Biases in the reporting of labour market dynamics. London, The Institute for Fiscal Studies.

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 (Statistics in Society)*, 170 (4), 955-974.

Powers, E. A. et al. (1978). Congruence between panel and ecall data in longitudinal research. *Public Opinion Quarterly* 42, 380-389

Poulsen, C. S. (1982). Latent structure analysis with choice modeling applications. Arhus: The Arhus School of Business Administration and Economics.

Reimer, M.(2004). Collecting Event History Data About Work Careers Retrospectively: Mistakes That Occur and Ways to Prevent Them. Kognition und Kommunikation dei der Laengsschnittdatenerhebung, Berlin, Max-Planck-Institut für Bildungsforschung.

Rubin, D. C., Baddeley, A. D. (1989). Telescoping Is Not Time Compression - A Model Of The Dating Of Autobiographical Events. *Memory & Cognition 17 (6)*

Sudman, S., and Bradburn, N. M. (1973). Effects of Time and Memory Factors on Response in Surveys. *Journal of the American Statistical Association 68, 805-815.*

Van de Pol, F., and Langeheine, R. (1990) Mixed Markov latent class models. *Sociological Methodology, 20, 213–247.*

Vermunt, J.K. Langeheine, R., and Böckenholt, U. (1999). Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates. *Journal of Educational and Behavioral Statistics*, 24, 178-205.

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., Tran, B., and Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of Longitudinal Research: Design, Measurement, and Analysis*. Elsevier. Pp 373-385.

Yamaguchi, K. (2008). Four useful finite mixture models for regression analyses of panel data with a categorical dependent variable. *Sociological Methodology*, 38, 283-328.

Wu, Lawrence L. (2000) Some comments on "Sequence analysis and optimal matching methods in sociology: review and prospect". *Sociological Methods & Research* 29:41-64.