

Comment

JEROEN K. VERMUNT¹

1. INTRODUCTION

I enjoyed very much reading this very well written paper. The topic addressed by Paul Biemer – classification errors in the measurement of employment status – is a very important one. Employment statistics belong to the most important macro-economic indicators and, actually, we would wish they would be free of error. It, however, turns out to be impossible to measure a person's employment without error. The best that can be done is design the data collection in such a manner that the classification errors at the individual level are minimized as much as possible. The current paper contributes to this objective.

An earlier study by Biemer and Bushery (2000) indicated that the 1993 changes in the measurement procedure that intended to reduce classification errors actually increased measurement error. In the current paper, Paul Biemer replicates these former analyses with a longer time series and with an extra employment category obtained by splitting the unemployed group into “on layoff” and “looking for work”. The reported results confirm the earlier conclusions that the new procedure is worse than the old procedure. In a second step, Biemer tries to disentangle the sources of measurement error for the two unemployed categories by modeling the separate questions that are used to determine whether a person is “on layoff” and “looking for work”, respectively. Sources of error are identified that point at possible improvements in the questionnaire.

Because of my background, my commentary will mainly concern methodological and statistical issues. More precisely, I will discuss some methodological problem related to application of the LC Markov model, as well as indicate how the statistical analysis could be somewhat refined. It is, however, not clear whether such a more elegant modeling will yield very different conclusions. I want to stress ones more that this is a great paper. My critical remarks are only meant to stimulate the discussion.

2. LATENT CLASS MARKOV: METHODOLOGY

The main engine of the study performed by Paul Biemer is the LC of hidden Markov model. Several assumptions that may affect the encountered results have to be made

when – as in this study – the model is applied with a single indicator per occasion. The assumption that is discussed in detail by Biemer is the first-order Markov process assumption. Simulation studies by Biemer and Bushery showed that, fortunately, estimates of classification error are not very sensitive to this assumption. Another assumption that is needed here for model identification is that the measurement error is constant over time. This assumption does not seem to be very problematic in the current study since we are looking for a single time-constant measure for classification error. Moreover, there is no good reason to assume that the quality of the measurement procedure changed over time while the procedure itself did not change (of course, apart from the questionnaire redesign). I am much more concerned about the third assumption; that is, the assumption of independent classification errors (ICE) over time (Bassi, Hageaars, Croon and Vermunt 2000). Is it realistic to assume that the occurrence of a certain type of classification error at time point t does not affect the probability of making the same mistake at time point $t + 1$? In my opinion, this assumption is not realistic in the current application. For example, a respondent who makes a mistake because (s)he did not understand one of the questions will most probably (or at least be more likely than others) make the same error again at the next occasion. In my opinion, it is necessary to conduct a simulation study to determine the sensitivity of the estimated classification errors for violations of the ICE assumption.

I have another critical remark concerning the use of the LC Markov model for quantifying measurement error in a person's employment state. According to the model, there is a probabilistic relationship between an individual's true and observed states. What is, however, the true state? Is it the true employment state occupied at a particular time point, or the state that would have been recorded with an error-free or gold-standard instrument? Or is it the state a person would have occupied under “normal conditions”? That is, if also randomness in his/her behavior is filtered out.

I will illustrate my point with a small example. Suppose that there is two types (two latent segments) of coffee consumers: consumers who prefer brand A and consumers who prefer brand B, and that I belong to the brand B segment, which means that under normal circumstances I buy brand B coffee. In an interview, I am asked which

¹ Jeroen K. Vermunt, Department of Methodology and Statistics, Tilburg University, The Netherlands.

brand I bought last week. Suppose I report that I bought a brand A package of coffee, and that am neither lying nor making a mistake. In other words, there is no classification error in the sense of making a mistake: I really bought brand A this week (the researcher doesn't know that of course). On the other hand, my behavior from this week is inconsistent with my preference, which means that in terms of measurement of my preference there is a classification error. This example illustrates that there are two types of "errors" that can be made: an error in the reporting and an "error" in the behavior. The "error" in my behavior of this week may have many causes, such as "brand B was sold out", "brand A was offered at a lower price this week", "I could not find the brand B package because of changes in the arrangement of the supermarket", *etc.* The LC Markov model is not able to distinguish such randomness in the behavior that is uncorrelated across time points from real classification errors.

What does this imply for the employment application? It implies that an individual's true state may be "on layoff", but for some reason (by chance) this particular month (s)he has worked. If this "some reason" is uncorrelated with other "some reasons" for being in the "wrong" observed state at other occasions, it will be labeled classification error by the LC Markov model. While in the case of the measurement of preferences based on revealed (or stated) preferences correcting for randomness in behavior seems to be exactly what we wish to accomplish, this is clearly not the case in the measurement of employment status. I, therefore, have the strong feeling that the error rates reported by Biemer might be somewhat overestimated because of randomness in employment behavior, for instance, caused by randomness in the functioning of the labor market.

A well-known consequence of modeling individual change by means of a LC Markov model is that the estimated number of latent transitions is much smaller than the corresponding observed numbers. The reason for this is that both independent classification errors and independent random behavior is filtered out; that is, part of the observed change is attributed to these phenomena.

3. LATENT CLASS MARKOV: MODEL SPECIFICATION

Paul Biemer estimated a separate three-occasion LC Markov model for each of the 30 three-month data sets. Interview mode was used as a grouping variable in order to take into account some of the heterogeneity in the true employment distributions and classification errors. The reported error rates in the tables are averages over interview modes and rotation groups.

I would have set up the model in a somewhat more elegant and less ad hoc manner. Instead of running a separate analysis for each of the rotation groups, I would have tried to build a simultaneous model for all rotation groups. The main problem of doing a series of separate analyses is that parameters that should actually be equated across rotation groups are now estimated without constraints. For example, the employment distribution in March 1994 should be the same in the rotation groups that were interviewed between January and March, February and April, and March and May, respectively. Moreover, the transition probabilities between March and April should be the same in the February–April and March–May rotation groups. This has also implications for the Parallel Survey groups: their time-specific latent distributions and transitions should be assumed to be equal to the ones of the standard CPS. That would have been a much better manner to test whether measurement error differ between the two questionnaires. Especially for the period in which the questionnaire forms overlap, it is crucial to assume equal latent distributions in order to be able to prevent that differences in measurement error appear partially as differences in true states.

A similar problem of the separate analyses applies for the estimation of the classification errors. These are assumed to be time-constant within the 3-month period that a rotation group is interviewed, but are allowed to differ across rotation groups, even if they are interviewed in the same month. It would, of course, be much better to impose equality constraints across rotation groups. A consistent application of the time-homogeneity assumption would imply that – both for the old and the new questionnaire form – the measurement errors are constant within the full investigation period.

What we, actually, need is a LC Markov model covering all 30 months; that is, a model for 30 instead of 3 time points. Such a simultaneous model for all rotation groups is as easily specified as a model for 3 time points. Of course, for each rotation group, only 3 of the 30 months are observed, which means that the other time points have to be treated as missing values. This is not a problem in the maximum likelihood estimation of the model parameters since we can simply assume that the data are missing at random (Vermunt 1997). Questionnaire type (old/new) serves as grouping variable (in addition to interview mode) and affects the time-homogenous classification error probabilities. In other words, we estimate only two sets of classifications errors, one for the old and one for the new questionnaire. Transition probabilities may change over time, but will be equal across rotation groups interviewed at the same occasions. Moreover, the initial state probabilities of a rotation group are not estimated as separate parameters

since they are defined by the current state of the latent Markov chain.

A practical problem of the simultaneous modeling is that with so many time points it no longer possible to estimate the model parameters with the standard EM algorithm. With a variant of EM called the Baum-Welch algorithm, however, the model can also be applied with many time points (Vermunt 2003; Paas, Bijmolt and Vermunt 2003). This algorithm is implemented in an experimental version of the Latent GOLD program (Vermunt and Magidson 2000, 2003) and will be available in a next version of this program.

An alternative way to implement a simultaneous model is as a LC Markov model for 3 occasions in which rotation group serves as grouping variable and in which the relevant across rotation group equality restrictions are imposed on the classification errors, transition probabilities, and initial state probabilities. The most complicated part of this approach is that it requires the use of restrictions on marginal probabilities (Vermunt, Rodrigo and Ato-Garcia 2001). More precisely, the initial state probabilities should be in agreement with the marginal class sizes in the rotation groups that are interviewed at the same occasion.

Other aspects of the modeling that could be refined are the treatment of missing values and the coding of the interview mode. It is not necessary to eliminate cases with missing values from the analysis as is done by Paul Biemer because ML estimation with missing values is straightforward. As far as the interview mode is concerned, it would be much more elegant to work with only two categories – proxy and self – instead of four categories and let the interview mode vary across occasions within cases. In other words, interview mode could be used as a time-varying covariate. Vermunt, Langeheine and Böckenholt (1999) proposed such a latent class Markov model with time-varying covariates.

4. MODEL FOR RESPONSE PROCESS

It is a very nice idea to try to disentangle which questions in the questionnaire are causing the classification errors by modeling the response process itself. This may yield lots of valuable information for redesigning the questionnaire. I, however, think that the extended models for the employment statuses “on layoff” and “looking for work” are formulated in an overly complicated manner.

The form of the created variable R is the same as of the outcome variable in a sequential choice analysis or in a discrete-time survival analysis. Answering the next question is fully determined by whether the current one is answered positively or not. The information we have is how many steps a person takes, which is conceptually equivalent to a

discrete survival time. A person “surviving” till the end is classified as being “on layoff” (“looking for work”).

In my opinion, it is not very helpful to treat this variable as being generated by K latent variables (Ts). This only makes sense if theoretically there should be a response hierarchy at the latent level, which, however, because of measurement error, is not encountered at the manifest level. That is, if at the manifest level there are 2^K instead of K possible responses. Even if is the case, it often suffices to conceptualize the model as a model with a latent variable with $K + 1$ classes and K indicators, a structure that is sometimes referred to as a probabilistic Guttman model.

Paul Biemer recognizes the complexity of the K latent and K manifest variables formulation and decides to simplify the model. However, I assume because of his starting point, he decided to keep $K + 1$ latent classes. I do not see why so many latent classes are needed. There are not even so many employment states. More logical would be to have only two classes – “on layoff” and “not on layoff” (“looking for work” and “not looking for work”) – since the questions are only intended to make this particular distinction. It can, of course, happen that the questions turn out to be informative about the type of “not on layoff” (“not looking for work”) status, in which case an extra latent class might be needed. What is clear to me is that $K + 1$ classes are far too many.

I was wondering how many persons were classified as “on layoff” (“looking for work”) at the various time points in the analysis with composite variable R as indicator. Are these numbers, as well as the number of transitions into and out of this state similar to the ones obtained with the standard four-state LC Markov model. In my opinion, this is a requisite for the validity of the calculation performed to obtain the figures presented in Tables 3 and 4.

A final thing that occurred to me is the following. Why not building a LC Markov model using the full questionnaire information as is done in the second part of the analysis. In other words, an alternative to using the observed constructed classification consisting of 4 employment categories would be to use the full set of CPS employment questions answered by the respondents. Such an analysis with multiple indicators would not only be much more informative, it would also make it possible to test and relax some of the assumptions that were made in the current analysis. For example, the ICE assumption could be relaxed for some of the questionnaire items.

REFERENCES

- 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.

- PAAS, L.J., BIJMOLT, T.H. and VERMUNT, J.K. (2003). Extending dynamic Segmentation with Lead Generation: A Latent Class Markov Approach. Center Paper, Tilburg University (submitted for publication).
- VERMUNT, J.K. (1997). *Log-linear models for event histories*. Techniques in the Social Sciences Series, Thousand Oakes: Sage Publications. 8.
- VERMUNT, J.K. (2003). Multilevel latent class models. *Sociological Methodology*, 33. In press.
- VERMUNT, J.K., LANGEHEINE, R. and BÖCKENHOLT, U. (1999). 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. (2000). *Latent GOLD User's Manual*. Boston: Statistical Innovations Inc.
- VERMUNT, J.K., and MAGIDSON, J. (2003). *Addendum to Latent GOLD User's Guide: Upgrade for Version 3.0*. Boston: Statistical Innovations Inc.
- VERMUNT, J.K., RODRIGO, M.F. and ATO-GARCIA, M. (2001) Modeling joint and marginal distributions in the analysis of categorical panel data. *Sociological Methods and Research*, 30, 170-196.