Sociological Methods & Research
http://smr.sagepub.com/

A Latent Markov Model to Correct for Measurement Error

FRANK VAN DE POL and JAN DE LEEUW

Sociological Methods & Research 1986 15: 118
DOI: 10.1177/0049124186015001009

The online version of this article can be found at:
http://smr.sagepub.com/content/15/1-2/118

Published by:
SAGE
http://www.sagepublications.com

In classical test theory the reliability of a test can be estimated by test-retest correlation models. These models do not apply to data of the lowest or nominal measurement level. Instead, models for latent Markov chains may be used to correct for measurement error in panel data from three or more waves. In this article it is shown how to use the E-M algorithm for estimating the parameters of a latent Markov chain. Where previous algorithms performed badly on variables with more than two categories this algorithm performs better, although convergence is often slow. The method is applied to two trichotomous questions from the Dutch civil servants panel survey. Generally the assumptions of the model, that is, a latent stationary Markov chain, are reasonably well met by the data. The probability of a correct answer, which can be interpreted as the reliability of a latent response category, is high in most cases (about .8). Also transition tables are presented that are corrected for measurement error according to the model. Standard errors of model parameters are approximated by a finite difference method.

A Latent Markov Model to Correct for Measurement Error

FRANK VAN DE POL
Netherlands Central Bureau of Statistics
JAN DE LEEUW
University of Leyden

1. *INTRODUCTION*

In the social sciences many phenomena are measured with a certain amount of error. This includes opinions, attitudes, and personality traits. Several circumstances may give rise to measure-

AUTHORS' NOTE: *The views expressed in this article are those of the authors and do not necessarily reflect the policies of the Netherlands Central Bureau of Statistics. The authors wish to thank J. Hagenaars, Th. Nijman, D. Sikkel, H. Stronkhorst, and two anonymous reviewers for their comments on previous drafts of this article.*

SOCIOLOGICAL METHODS & RESEARCH, Vol. 15 Nos. 1-2, August-November 1986 118-141

ment error. The respondent may be tired, little interested, or just in a bad mood. Also the respondent may not be sure what his opinion or attitude is in the matter concerned. Furthermore, in the case of an interview, the interaction between the respondent and the interviewer may be an important disturbing factor as well as the presence of others during the interview. Finally, errors may occur during the coding and typing process. Measurement errors, both random and systematic, can hardly be avoided. The problem has been stated and illustrated already often before, both for tables (Petersson, 1974; Schwartz, 1985) and for measures of association (Kraemer, 1985).

Of course, the best thing to do is to prevent measurement errors whenever possible. A measurement model describing the error generating process will never be able to produce perfectly "true" data. On the other hand, results that are corrected by a correctly specified measurement model will generally be less biased than the uncorrected results.

When several consecutive measurements are available for the same respondent, that is, in the case of panel data, there are simple models available that enable the researcher to separate random measurement error and true change (Heise, 1969; Jöreskog, 1970; Van de Pol, 1982). Henry (1973) pointed out that if these models are applied to a dichotomous variable some parameters will be unintentionally fixed because the mean and the variance are not independent for a dichotomy. Moreover, one does not obtain a very detailed account about the location of measurement error by these methods. Just one reliability coefficient is produced, indicating the proportion true score variance.

When data are measured at the nominal level it is necessary to use another approach. In the fifties and sixties, latent structure analysis was developed by Lazarsfeld and Henry (1968). These authors, as well as Wiggins (1955, 1973) applied this class of models also to panel data. In one of their models they assumed a latent Markov chain that was connected with the observed, manifest data by certain probabilities to give a correct or a wrong answer. The model produces a latent discrete probability distribution for every panel measurement that may differ from the manifest probability distribution if the latent classes are measured with unequal reliabilities.[1] The turnover tables of consecutive measurements are corrected for measurement error. Because random measurement error inflates the measured change, this usually means that on a latent level less respondents are found in the "change" categories than on the manifest level.

Until recently the estimation of the parameters for these models was a problem. Lazarsfeld and Henry (1968) proposed an eigenvalue approach for the latent Markov chain that could produce negative "probabilities" as model parameters. Goodman (1974) used a generalization of the iterative proportional fitting procedure for latent structure analysis. In this procedure probabilities are bound in the 0-1 range. Dempster et al. (1977) wrote a thorough mathematical exposition about this procedure for missing data they called the E-M algorithm (expectation—maximization). They proved that this algorithm produces maximum likelihood estimates. The first application of the E-M algorithm *avant la lettre* is due to Wolfe (1970).

Goodman's fitting procedure is implemented in a computer program for latent class analysis according to Goodman, LCAG (Hagenaars, 1985). The model behind this program allows for a log-linear model structure at a latent level. This implies that some latent Markov chain can be specified. An older computer program for Maximum-Likelihood Latent Structure Analysis is MLLSA (Clogg, 1977; Clogg and Goodman, 1985). This program allows for all kinds of factor-analytic structures in one or more populations, but not for a nonsaturated latent log-linear model. So in applications to panel data the analyst using MLLSA is forced to assume that the latent variable does not change in time (Dayton and Macready, 1983), whereas the analyst using LCAG is not inhibited in this sense (Hagenaars, 1978). However, not all Markov models that are described in this article are compatible with LCAG.

Another approach for which a program exists is the one by Coleman (1964). Starting out with a beautiful theory about attitude particles governed by a Markov process in continuous time, he eventually reaches a simple expression for a matrix of reliability coefficients. This matrix can be estimated using only bivariate tables of panel measurements in a way that is very similar to Heise's formula for the reliability of an interval-level variable.[2] An application of the method can be found in Markus (1979). However, the Coleman approach cannot be easily put into the general framework of latent structure analysis.

The manifest Markov chain is introduced in the next section and the latent Markov chain in section three. The fourth section deals with estimation methods. Then the model is applied to two variables from the Dutch civil servants panel survey. We focus on questions on satisfaction with rank and with economic position. Finally the main conclusions are given.

2. *THE MANIFEST MARKOV CHAIN*

Before turning to the latent Markov chain we will first consider the manifest Markov chain in discrete time. Suppose there are consecutive measurements of the same variable. The first measurement is denoted by x^1, the second by x^2, and so on. Each variable x^s has the same response categories $(1, \ldots, c, \ldots C)$.

A Markov chain is described by an initial distribution vector p^s of x^s and a set of transition matrices R^{st} for transitions from x^s to x^t $(s < t)$. The elements of this matrix will be denoted as r_{ik}^{13} if, for example, $s = 1$ and $t = 3$. An example of an observed transition matrix is given in Table 1. The relationship between the turnover table P^{st}, with elements p_{ij}^{st} adding to 1, and p^s and R^{st} is given by

$$P^{st} = P^s R^{st} \quad [1]$$

where P^s is a diagonal matrix with the elements of p^s on the diagonal and zeros elsewhere. The main assumption of a Markov model is that a transition matrix R^{st} is independent of the past states through which the process has passed. So a process without memory is assumed. As will be explained below, this implies that the transition matrix of two consecutive periods (s, t) and (t, u) satisfies

$$R^{su} = R^{st} R^{tu} \qquad (s < t < u) \qquad [2a]$$

or, in scalar notation

$$r_{ik}^{su} = \Sigma_j \, r_{ij}^{st} \, r_{jk}^{tu} \qquad [2b]$$

The probability of a respondent being in some latent cell p_{ij}^{st} was already given in equation 1 as the probability of being in category i at time s, p_i^s, times the probability of a transition from category i at time s to category j at time t, r_{ij}^{st}. Now equation 2 states that transition probabilities r_{jk}^{tu} for a respondent in category j at time t are independent of his past state i at time s; $p_{jk}^{tu} = p_j^t r_{jk}^{tu}$ has the same parameters for any category i at a previous time s. So the probability of a respondent being in some cell p_{ijk}^{stu} is obtained by multiplying p_{ij}^{st} by r_{jk}^{tu} for corresponding values of j,

$$p_{ijk}^{stu} = p_{ij}^{st} \, r_{jk}^{tu} = p_i^s \, r_{ij}^{st} \, r_{jk}^{tu} \qquad [3]$$

The model is illustrated by a causal diagram in Figure 1. Tables of higher dimensionality may be described by multiplying with transition probabilities for more periods. Tables of lower dimensionality may be obtained by summing the p_{ijk}^{stu} over nonrelevant dimensions.

In the present article the length of the periods (s, t) and (t, u) is taken to be the same as the time between consecutive waves of the

TABLE 1

Transition Probabilities r_{ik}^{13} and Marginal Probabilities p_i^1 and p_k^3: Civil Servants' Satisfaction with Their Economic Position, 1979-1981

March 1979	March 1981				marginal distribution
	1.	2.	3.	total	March 1979: p^1
1. (very) satisfied	.62	.27	.11	1.00	.55
2. rather satisfied	.29	.42	.29	1.00	.30
3. neither/ not satisfied	.15	.27	.58	1.00	.15
marg.distr. March '81, p^3	.45	.32	.23	1.00	

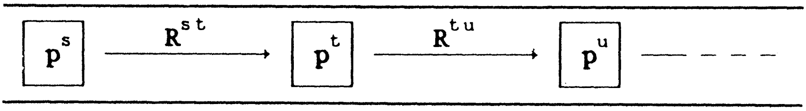

Figure 1: Causal Diagram of a Manifest Markov Chain in Discrete Time

panel-data collection, but this assumption is not necessary. Singer and Spilerman (1978) have pointed out that choosing some arbitrary time unit that is not a whole multiple of these periods may be incompatible with assumption 2. For half the time period (s, t) for instance there should be a transition matrix R* such that $R^{st} = R^* R^*$. They gave an example of a matrix R^{st} that cannot be decomposed in this way. This problem will not occur if the model is not based on transition probabilities for some discrete time unit but on transition rates for an infinitesimal small elapse of (continuous) time (Tuma et al., 1979; Carroll, 1983). However, one will not be inhibited by this drawback of discrete time models as long as one sticks to the observation periods of the panel.

Much of the theory on Markov chains (Anderson, 1980) is derived under the restriction of stationary transition probabilities in time—for periods (s, t), (t, u), and so on, of equal length.

$$R^{st} = R^{tu} = \ldots = R \qquad [4]$$

When fitting manifest Markov chains to empirical data the stationarity assumption is often met by the data. The assumption of a process without memory (2), however, is very often not in accordance with the data. Wiggins (1973) and Logan (1981) pointed out that reversion effects may take place. An advertising campaign, for instance, may only have a temporary effect, influencing the transition probabilities R^{12}. When the campaign stops at time 2 people will tend to return to their old brand preferences at time 1. So the transition probabilities in the following period R^{23} $(x^1 = i)$ will be different for every preference i at time 1. This is called a second-order Markov chain.

Other authors interpreted the violation of assumption 2 in a different way. Blumen et al. (1966) studied labor force dynamics and observed that after many periods the change predicted from a first-order Markov chain was larger than the actual change. This was caused by the fact that different individuals satisfied different Markov processes. They proposed to distinguish two distinct latent categories: movers and stayers. The first group has no transitions at all. The second group satisfies the assumptions of a Markov chain.

The approach that is adopted in this article is to relax the Markov assumption in a different way. In classical test theory the observed (interval level) scores are decomposed in true scores and error. In this tradition, Heise (1969) and Jöreskog (1970) adapted the Markov simplex, which describes test-retest correlations, by introducing a quasi Markov simplex, which is an unreliably measured, and therefore latent, Markov simplex. For discrete data Wiggins introduced in his dissertation (1955) the latent Markov chain, which is an unreliably measured Markov chain. Henry (1973) treated both approaches in one article.

3. *THE LATENT MARKOV CHAIN*

In order to formulate a latent Markov model some latent variables have to be defined. Corresponding with every manifest variable x^t one latent variable y^t is assumed. We will treat the case where the latent variables y^t are assumed to have the same number, c, of latent classes as the manifest x^t. The probability distribution of a manifest variable x^t, p^t, depends on the probability distribution of the corresponding latent variable y^t, v^t, and a matrix of transition probabilities from latent to manifest, the conditional response probabilities Q^t,

$$p^t = v^t Q^t \qquad [5]$$

The sequence of latent response classes is made such that a diagonal element from Q^t denotes the probability of a correct answer, which may be interpreted as the reliability by which a latent class is measured.

On the latent level there is a three-way table of the consecutive variables y^s, y^t, and y^u. It is assumed that the latent variables y^t are interrelated by a Markov chain, as described in the previous section. The latent transition matrix is denoted by M^{tu} (for the manifest Markov chain by R^{tu}) and the latent initial distribution by the diagonal matrix V^{s} (manifest P^{s}). The probability of being in latent class (α, β, γ), $v^{stu}_{\alpha\beta\gamma}$, is described by the latent Markov chain that was assumed: $v^{stu}_{\alpha\beta\gamma} = v^{s}_{\alpha}\, m^{st}_{\alpha\beta}\, m^{tu}_{\beta\gamma}$.

When observed, this three-way table generates a six-way table of three latent variables times three manifest variables. About their relation the usual assumption of local independence is made. A manifest variable at time t, x^{t}, is only dependent on the latent counterpart y^{t}, or, stated otherwise, x^{t} is independent from all variables in the model except y^{t}. Thus for a respondent in latent class (α, β, γ) the probability of answering i, j, and k is $q^{s}_{\alpha i}\, q^{t}_{\beta j}\, q^{u}_{\gamma k}$. Hence the probability of being in cell (α, i, β, j, γ, k) is obtained by multiplying this response probability with $v^{stu}_{\alpha\beta\gamma}$.

$$p^{stu}_{ijk} = \Sigma_{\alpha}\, \Sigma_{\beta}\, \Sigma_{\gamma}\, v^{s}_{\alpha}\, q^{s}_{\alpha i}\, m^{st}_{\alpha\beta}\, q^{t}_{\beta j}\, m^{tu}_{\beta\gamma}\, q^{u}_{\gamma k} \qquad [6]$$

Summing over the latent dimensions α, β, and γ the expected cell proportions p^{stu}_{ijk} of the manifest three-way table are obtained. A causal diagram of the latent Markov chain is given in Figure 2.

4. *ESTIMATION*

Lazarsfeld and Henry (1968) proved that the parameters of the latent Markov chain in principle can be identified. They found an expression for $(Q^{t})^{-1}\, \Lambda^{t}_{j} Q^{t}$ in terms of observed bivariate and trivariate tables. Here Λ^{t}_{j} is defined to be diagonal with one column j from Q^{t}, that is, the elements $q^{t}_{\beta j}$, on the diagonal.

$$A_{j} = P^{st}(P^{su})^{-1}\, P^{su}(x^{t} = j)\, (P^{st})^{-1} = (Q^{t})^{-1}\, \Lambda^{t}_{j} Q^{t} \qquad [7]$$

The product of observed matrices is called A_j. Hence the characteristic equation $A_j(Q^{t})^{-1} = (Q^{t})^{-1}\Lambda^{t}_{j}$ may be solved.

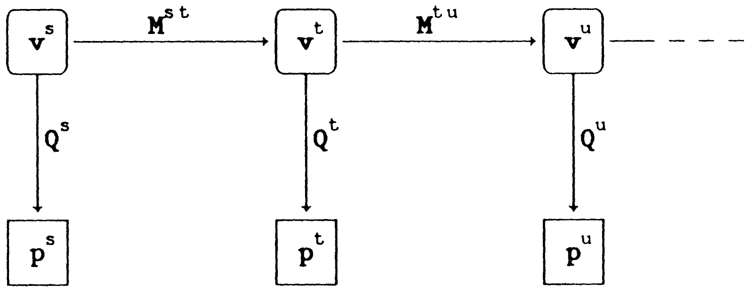

Figure 2: Causal Diagram of a Latent Markov Chain in Discrete Time

However, several difficulties may be encountered when applying this method. First, the study of a small sample and a variable with three or more categories may result in unprecise estimates of the cell probabilities, p_{ijk}^{stu}. Because of this, or because the model does not fit, some of the eigenvalues may be complex; the eigenvalue polynomials may not pass through the zero-axis as often as they should. Then some of the elements in Q^t are not defined by this algorithm. Furthermore, even if Q^t is completely identified, negative probabilities may be found in the parameter matrices V^s, Q^t, or M^{st} and M^{tu}.

The above mentioned problems make it desirable to look for another estimation method. This is done below. Nevertheless, from equation 7 it may be seen that only the matrix Q^t, that is, the middle Q-matrix of three measurements, is identified in the latent Markov chain model, just as in Heise's model (Henry, 1973; Van de Pol, 1982). For latent Markov chains of four or more points in time, the first and the last matrix with reliability coefficients, Q^s and Q^ℓ, are not identified. This forces us to make an assumption like

$$Q^s = Q^{s+1} \quad \text{and} \quad Q^\ell = Q^{\ell-1} \qquad [8]$$

The first reliability matrix is equal to the second and the last reliability matrix is equal to the previous one. For three consecutive measurements (8) reduces to $Q^s = Q^t\ Q^u$.

Looking for a better algorithm one might attempt to maximize the likelihood of the latent Markov chain. The log-likelihood for a three-wave latent Markov chain may be obtained from equation 6 that gives the expected proportions p_{ijk}^{stu} of the manifest three-way table in a population where the model holds. The likelihood of a multinomially distributed discrete random quantity is given by the product of the proportions that are reproduced by the model parameters, $\hat{p}_{ijk}^{stu}$ to the power f_{ijk}^{stu}, the observed frequencies a random sample (Edwards, 1976). Because the natural logarithm of the likelihood function attains its maximum for the same values $\hat{p}_{ijk}^{stu}$ as does the function itself, the log likelihood, ln (L), which has convenient additive properties, may be maximized instead.

$$\ln(L) = \Sigma_i \Sigma_j \Sigma_k \; f_{ijk}^{stu} \ln \Sigma_\alpha \Sigma_\beta \Sigma_\gamma \; v_\alpha^s q_{\alpha i}^s m_{\alpha\beta}^{st} q_{\beta j}^t m_{\beta\gamma}^{tu} q_{\gamma k}^u \qquad [9]$$

Maximum-likelihood estimates for the model parameters may be obtained by putting the first-order derivatives to every parameter to zero. The resulting system of equations may be solved using the iterative Newton-Raphson method that furthermore requires second-order derivatives. This approach is advocated by Haberman (1979) for latent structure analysis. Hagenaars (1985), however, reports that the Haberman algorithm needs very good starting values; otherwise the iterations do not converge. So this algorithm should not be used to find maximum-likelihood estimates of the model parameters, but may nevertheless be useful in estimating their variances, as will be shown later.

Optimization of the log likelihood via the E-M algorithm is much simpler. Suppose that the full table of both latent and manifest variables could be observed. Then the log likelihood for the model would be

$$\Sigma_\alpha \Sigma_\beta \Sigma_\gamma \Sigma_i \Sigma_j \Sigma_k \; n\, \theta_{\alpha\beta\gamma ijk} \; \ln(v_\alpha^s q_{\alpha i}^s m_{\alpha\beta}^{st} q_{\beta j}^t m_{\beta\gamma}^{tu} q_{\gamma k}^u) \qquad [10]$$

where the probability of a respondent being in cell $(\alpha, \beta, \gamma, i, j, k)$ is $\theta_{\alpha\beta\gamma ijk}$ and n is the sample size, $\Sigma\, f_{ijk}^{stu}$.

The E-M algorithm consists of two steps. In the E (expectation) step the missing or latent observations are replaced by their expectations given the parameters of the model and the data observed. In the M (maximization of the likelihood) step new estimates of the model parameters are computed. So one iteration of the E-M algorithm consists of finding auxiliary maximum-likelihood estimates for the model parameters according to 10 in the M-step and subsequently finding a new estimate of the full table $\theta_{\alpha\beta\gamma ijk}$ from the parameters and the observed data in the E-step. It can be shown that the auxiliary maximum-likelihood estimates of the model parameters are directly estimated by summation over nonrelevant dimensions of $\theta_{\alpha\beta\gamma ijk}$.

$$v_{\alpha}^{s} = \theta_{\alpha+++++}/\theta_{++++++} \qquad [11a]$$

$$m_{\alpha\beta}^{st} = m_{\beta\gamma}^{tu} = \frac{v_{\alpha\beta}^{st} + v_{\beta\gamma}^{tu}}{v_{\alpha}^{s} + v_{\beta}^{t}} = \frac{\theta_{\alpha\beta++++} + \theta_{+\beta\gamma+++}}{\theta_{\alpha+++++} + \theta_{+\beta++++}} \qquad [11b]$$

$$q_{\alpha i}^{s} = q_{\beta j}^{t} = q_{jk}^{u} = \frac{\theta_{\alpha++i++} + \theta_{+\beta++j+} + \theta_{++\gamma++k}}{\theta_{\alpha+++++} + \theta_{+\beta++++} + \theta_{++\gamma+++}} \qquad [11c]$$

If no stationarity assumption (4) is to be made equation 11b should be replaced by two equations, $m_{\alpha\beta}^{st} = v_{\alpha\beta}^{st}/v_{\alpha}^{s}$ and $m_{\beta\gamma}^{tu} = v_{\beta\gamma}^{tu}/v_{\beta}^{t}$. As stated above these model parameters would be the maximum-likelihood estimates if the full table $\theta_{\alpha\beta\gamma ijk}$ was known. Because this is not the case, we should find the expected values of $\theta_{\alpha\beta\gamma ijk}$ given the new model parameters and the data. This is done in the E-step. In the first part of the E-step the full table is computed on the basis of the model parameters. This intermediate result is called $\xi_{\alpha\beta\gamma ijk}$.

$$\xi_{\alpha\beta\gamma ijk} = v_{\alpha}^{s} q_{\alpha i}^{s} m_{\alpha\beta}^{st} q_{\beta j}^{t} m_{\beta\gamma}^{tu} q_{\gamma k}^{u} \qquad [12]$$

These probabilities $\xi_{\alpha\beta\gamma ijk}$ are not identical with $\theta_{\alpha\beta\gamma ijk}$; equation 12 is not a reversion of 11. The second part of the E-step is to bring into $\xi_{\alpha\beta\gamma ijk}$ the information on the observed frequencies f_{ijk}^{stu}. This is done by proportional fitting. From 6 and 12 it may be seen that the estimated proportions by the model are $\hat{p}_{ijk}^{stu} = \xi_{+++ijk}$. So the new expectations of the full table given the model parameters and the observed data is given by

$$\theta_{\alpha\beta\gamma ijk} = \frac{(f_{ijk}^{stu}/n)}{p_{ijk}^{stu}} \, \xi_{\alpha\beta\gamma ijk} \qquad [13]$$

thus completing one cycle of the E-M algorithm. Reasoning along the same lines the algorithm may easily be extended to more points in time.

Iterations may start with equation 11, substituting some plausible estimates for the model parameters. In order to ensure that the latent classes are ordered in the same way as the manifest classes the Q-matrices should have high starting values on the diagonal. The M-matrices should have the same structure. Zeros should be avoided as starting values. Once the full table contains a zero in some cell this value will not change anymore during the iterations that follow. For a variable with C categories we mainly use as starting values for diagonal elements d and nondiagonal elements nd:

$$d = \sqrt{1/C} \qquad nd = (1 - d)/(C - 1) \qquad [14]$$

Iterations should continue until the improvement of the likelihood per iteration does no longer exceed some convergence criterion, ϵ, or until no parameter changes more than some very low value like 10^{-7}. Dempster et al. (1977) and Wu (1983) gave general conditions for convergence of the E-M algorithm. In our experience with the E-M algorithm on the latent Markov chain there was always convergence, although convergence was rather slow. For a variable with 3 categories usually some 200 iterations were required to obtain accuracy up to the second decimal for all parameters; in some extreme cases even more iterations were needed. For a variable with 2 categories considerably less iterations were required. Furthermore, the algorithm proved to reproduce exactly a latent three-way table from a manifest three-way table that was artificially constructed from known model parameters.

It is unknown whether or not the likelihood of a latent Markov chain can have local maxima. If so, the algorithm might run, or better: walk, into one of these maxima when bad starting values are used. But, using several sets of starting values, no local maximum has been recognized in the empirical data sets we used, except the artificial one that is caused by zeros in the starting values.

In order to evaluate the fit of the model one may look at the likelihood ratio (LR). For three measurements the LR as compared to the unrestricted model may be written as

$$LR = 2\, \Sigma_i \Sigma_j \Sigma_k\, f_{ijk}^{stu} \ln(f_{ijk}^{stu}/(n\, p_{ijk}^{stu})) \qquad [15]$$

This LR is χ^2 distributed with $C(C^2 - 2C + 1)$ degrees of freedom for a stationary latent Markov chain (4) and $C(C^2 - 3C + 2)$ degrees of freedom for a nonstationary latent Markov chain. Hence no goodness-of-fit statistic is available for three measurements of a dichotomous question, when no stationarity is assumed: $2(2^2 - 6 + 2) = 0$. If one does assume stationarity in the dichotomous case the LR is only a test on stationarity (with 2 df). For variables with $C \geq 3$ the LR is also a test on the Markov assumption (2) for the latent variable. A better rest on this assumption can be obtained by analyzing more than three measurements. Then also the assumption of stationarity of the reliability matrices, Q, (formula 8) may be tested.

Once the maximum-likelihood estimates of the model parameters have been computed their variances may be found from the information matrix, which we call B. This is the matrix of second-order derivatives of the likelihood toward all parameters. Stacking all independent parameters from v_α^s, $q_{\beta j}^t$, $m_{\alpha\beta}^{st}$ (and in case of nonstationarity also from $m_{\beta\gamma}^{tu}$) into one vector ψ, the covariance matrix[3] of these parameters is according to likelihood theory

$$-B^{-1} = \left(-\frac{\partial\partial \ln(L(\psi))}{\partial\psi\,\partial\psi}\right)^{-1} \quad [16]$$

This result can be found among others in Andersen (1980). In the present article the second-order derivatives have been computed numerically. For those who are not familiar with this simple method the formula will be given below. By definition the second-order derivative of the ln(L)-function (LN-function for short) to some parameters ψ_i and ψ_j is given by the derivative of $\partial LN(\psi)/\partial\psi_j$ to ψ_i

$$\frac{\partial\partial LN(\psi)}{\partial\psi_i\,\partial\psi_j} = \frac{\partial}{\partial\psi_i}\left(\frac{LN(\psi_j+\epsilon, \psi_{\bar{j}}) - LN(\psi)}{\epsilon}\right)$$

$$= \frac{LN(\psi_i+\epsilon, \psi_j+\epsilon, \psi_{\bar{ij}}) - LN(\psi_i+\epsilon, \psi_{\bar{i}}) - LN(\psi_j+\epsilon, \psi_{\bar{j}}) + LN(\psi)}{\epsilon^2}. \quad [17]$$

Here $\psi_{\bar{i}}$ denotes all elements from ψ except ψ_i. The symbol $\psi_{\bar{ij}}$ denotes all elements from ψ other than both ψ_i and ψ_j. Using double precision $\epsilon = 10^{-4}$ turned out to be small enough. For the computation of first-order derivatives $\epsilon = 10^{-7}$ is to be preferred. As a by-product of these computations also the first-order derivatives are obtained, which should be near zero for maximum-likelihood estimates.

When no stationary of the M-matrices is assumed (4) it is not necessary to use the algorithm that was given here. In that case the latent Markov structure may be specified as a latent log-linear model. Because a nonstationary first-order Markov model will reproduce exactly only the bivariate subtables of consecutive measurements, V^{st} and V^{tu}, one should only assume bivariate interactions between consecutive y-variables (y^s, y^t) and (y^t, y^u) and so on, and not interactions between y-variables that lie further apart in time like (y^s, y^u) and so on, or higher order interactions (y^s, y^t, y^u), and so on (Bishop et al., 1977). Computations may be carried out using LCAG (latent class analysis according to Goodman), a program for latent class analysis, suitable for models with several latent variables that are related by some latent log-linear model (Hagenaars, 1985). This program also uses the E-M algorithm. Unfortunately no estimates of the variances of model parameters are produced.

Application of this form of the E-M algorithm may become problematic for variables with many response categories and for data on more than three points in time. First, there should be enough observations (respondents) to obtain reasonably precise estimates of the frequencies in every cell. f^{stu}_{ijk} (or f^{stuv}_{ijkl} for four points in time). Second, the full table, $\theta_{\alpha\beta\gamma ijk}$ for three measurements, may be so big that a computer program performing the algorithm cannot be loaded into the central memory.

5. AN APPLICATION TO A PANEL OF CIVIL SERVANTS: TWO QUESTIONS ON INCOME SATISFACTION

As an example the method will be applied to questions on income satisfaction from a panel survey among civil servants. On request of the Ministry of Internal Affairs, the Netherlands Central Bureau of Statistics has carried out for several years this panel survey that was aimed at the measurement of well-being. The questions on income satisfaction are worded as follows.

- Are you satisfied with your rank?
  {very satisfied, satisfied, rather satisfied,
  neither satisfied nor dissatisfied, not satisfied}
- Are you satisfied with your economic position?
  {very satisfied, satisfied, rather satisfied,
  neither satisfied nor dissatisfied, not satisfied}

The response categories were made nonsymmetrical because, at the beginning of the panel, in 1979, most civil servants were (very) satisfied.

Three consecutive measurements were available on about 3,000 panel members. In order to avoid empty cells in the three-way table the number of response categories has been brought down to three: (1) (very) satisfied (2) rather satisfied (3) neither satisfied nor dissatisfied/not satisfied.

TABLE 2
The Fit of the Latent Markov Chain for Three Panel Waves (6 Degrees of Freedom): Two Income-Satisfaction Questions from the Dutch Civil Servants Panel

	likelihood ratio χ^2	probability level	number of cases
Satisfied with rank?	4.8	.56	2989
Satisfied with economic position?	5.0	.54	2981

In Table 2 the fit of the latent Markov model is given. For both questions the fit is very satisfactory. It should be noted, however, that this is only a partial test of the Markov assumption (2). Using only three consecutive measurements this assumption will only restrict the three way interaction (if the variable has three or more categories). A test that also poses restrictions on bivariate interactions can only be obtained using four or more consecutive measurements. No stationarity assumption has been made because the three measurements were not spread evenly in time. The first measurement was in March 1979, the second in November of the same year, and the third in March 1981.

In Table 3 the probability distribution of the latent classes is given for the first measurement, p^1. For the latent Markov chain model these classes may be given the same labels as the corresponding manifest classes. Furthermore, Table 3 displays the response probabilities Q. If all latent classes are measured with perfect reliability this matrix should be the identity matrix with 1's on the diagonal and zeros elsewhere.

Table 3 shows that the reliabilities of the extreme latent classes are high, about 90% of the answers are on the diagonal. There is, however, more response uncertainty for respondents pertaining to the middle latent classes, "rather satisfied," where only some 60% of the answers are on the diagonal. There are two explanations for the lower reliability of the middle latent class, both of which may be relevant for the present data. First, there is a boundary effect for respondents in extreme latent classes: response uncertainty can manifest itself only in the direction of a less extreme answer.

Second, the extreme latent classes correspond to manifest categories that were taken together and the middle class does not.

TABLE 3
The Latent Distribution at Time 1 (March, 1979), v^1, and the Response Probabilities, Q. Two Income Satisfaction Questions from the Dutch Civil Servants Panel

	Are you satisfied with your rank ?				Are you satisfied with your economic position ?			
	lat.dist.	1.	2.	3.	lat.dist.	1.	2.	3.
1.(very) satisfied	.32 (.02)	.91 (.02)	.08 (.02)	.02 (.01)	.50 (.03)	.90 (.02)	.09 (.02)	.01 (.01)
2.rather satisfied	.32 (.03)	.24 (.03)	.54 (.03)	.22 (.04)	.36 (.03)	.29 (.03)	.62 (.03)	.09 (.03)
3.neither/ not satisf.	.35 (.03)	.02 (.01)	.10 (.02)	.88 (.03)	.15 (.02)	.02 (.02)	.19 (.04)	.79 (.05)

NOTE: n ≈ 3000 (with standard deviations).

Some respondents belonging to, for instance, a latent class "very satisfied" presumably will incorrectly have answered "satisfied." This is, however, not translated into a lower reliability of the collated latent class "very satisfied" and "satisfied," because the corresponding manifest categories were taken together.

When the latent classes of a question are measured with unequal reliabilities the latent marginal probability distribution will presumably differ from the manifest one because of the relation that exists between these two (see equation 5).

Table 4 shows that the middle category, "rather satisfied," is better filled on a latent level than on a manifest level. This is because the middle latent class is less reliably measured than the other latent classes. The respondents in this class spread more evenly over the manifest response categories than the respondents in other latent classes do.

Table 4 also shows that the income satisfaction of Dutch civil servants has decreased between 1979 and 1981. This decrease is probably due to the decrease of their real income during the 1980s (Van de Stadt et al., 1985). That explains why the decrease is more clearly visible in the question on the economic position than in the question on the rank (which did not change dramatically).

Finally, also turnover tables may be corrected for measurement error. From Table 5 latent and manifest transition matrices can be compared. Only the transitions for the second period (from

TABLE 4
The Latent and Manifest Distribution, v and p, at Time 1 (March, 1979) and at Time 3 (March, 1981): Two Income Satisfaction Questions from the Dutch Civil Servants Panel

	Are you satisfied with your rank ?				Are you satisfied with your economic position ?			
	latent		manifest		latent		manifest	
	1979	1981	1979	1981	1979	1981	1979	1981
1.(very) satisfied	.32 (.02)	.31	.38 (.01)	.36	.50 (.03)	.38	.56 (.01)	.45
2.rather satisfied	.32 (.03)	.29	.23 (.01)	.22	.36 (.03)	.37	.29 (.01)	.31
3.neither/ not satisf.	.35 (.03)	.40	.39 (.01)	.42	.15 (.02)	.25	.15 (.01)	.23

NOTE: n ≈ 3000 (with standard deviations).

TABLE 5
The Latent and Manifest Transition Probabilities from November 1979 to March 1981, M^{23} and R^{23}, Together with the Initial Distribution, v^2 and p^2: Two Income Satisfaction Questions from the Dutch Civil Servants Panel

	Are you satisfied with your rank ?							
	latent				manifest			
	init.dist.	1.	2.	3.	init.dist.	1.	2.	3.
1.(very) satisfied	.32	.82 (.03)	.17 (.04)	.00 (.02)	.38 (.01)	.66 (.01)	.21 (.01)	.13 (.01)
2.rather satisfied	.33	.09 (.04)	.62 (.06)	.29 (.04)	.24 (.01)	.27 (.02)	.31 (.02)	.42 (.02)
3.neither/ not satisf.	.35	.04 (.02)	.09 (.05)	.87 (.04)	.39 (.01)	.12 (.01)	.18 (.01)	.70 (.01)

	Are you satisfied with your economic position ?							
	latent				manifest			
	init.dist.	1.	2.	3.	init.dist.	1.	2.	3.
1.(very) satisfied	.50	.77 (.03)	.20 (.03)	.03 (.02)	.55 (.01)	.64 (.01)	.26 (.01)	.09 (.01)
2.rather satisfied	.35	.00 (.04)	.76 (.06)	.24 (.03)	.29 (.01)	.29 (.02)	.44 (.02)	.27 (.02)
3.neither/ not satisf.	.16	.05 (.04)	.02 (.09)	.93 (.07)	.16 (.01)	.11 (.01)	.27 (.02)	.62 (.02)

NOTE: n ≈ 3000 (with standard deviations).

November 1979 to March 1981) are displayed because only this period was long enough to have considerable changes taking place on a latent level. Table 5 shows that the stability, the diagonal elements, would be underestimated if the manifest transition probabilities were taken for the truth. When measurement error is accounted for according to the latent Markov model the stability appears to be a lot higher. Stated differently, the change is much smaller on the latent level than on the manifest level. Part of the change may be labeled as regression to the mean or regression to the modal category. This applies for the middle column of the transition matrix, change from 3 to 2 and from 1 to 2. According to the model, about one-third or one-half of the regression to the modal category is "true," the rest being caused by measurement error.

As to the precision of the parameter estimates one may observe from Table 5 that some of the standard errors in the latent transition matrix are quite high. This occurs especially when the corresponding latent class is small. Here is another argument not to analyse the full $5 \times 5 \times 5$ table but only the smaller $3 \times 3 \times 3$ table.

In the first section it was stated that the E-M algorithm will behave nicely if the maximum-likelihood estimates of probabilities are outside the 0-1 range. An example of this may be found in the latent transition matrix of the question on the economic position. This matrix contains a parameter, r_{21}, that is estimated by the E-M algorithm to be zero. The first-order derivative (not displayed in a table) is quite high, −18. Here a Newton-Raphson algorithm would have produced a negative parameter estimate, which is clearly undesirable for a transition probability.

At last a word of caution about the comparison of distinct rows in a transition matrix may be useful. The decrease in satisfaction cannot easily be seen from the transition matrix unless the initial distribution is taken into account, thus obtaining an ordinary turnover table. In the transition matrix every row has the same weight; in an ordinary turnover table the transition probabilities are weighted with the initial distribution. The turnover table may be computed by applying formula 1.[4]

6. *DISCUSSION*

The latent Markov chain is a measurement model. It is a model that describes the turnover in some characteristic from the point of view that the turnover is governed by a process without memory and that this characteristic is measured with error. There are parallels with certain models for interval data. Although the quasi Markov simplex (Heise, 1969; Jöreskog, 1970) may be used for interval data, the latent Markov chain is to be used when the measurement level is clearly lower, that is, ordinal or nominal.

This model may be especially suitable for analyzing the reliability and turnover concerning opinions, attitudes, preferences, and the like, when panel data are available. Reliabilities are obtained for every latent class. More precisely: the probability of a correct answer is obtained, given the latent class that corresponds to some response category. Furthermore, the marginal distribution at a specific point in time is corrected for measurement error and "latent" turnover tables are computed, which are corrected for attenuation due to measurement error. Thus reliability and stability are separated.

The latent Markov chain was first formulated by Wiggins (1955). Lazarsfeld and Henry (1968) used an algorithm based on the computation of eigenvalues. The main drawback of this algorithm is that negative estimates of probabilities might be produced especially when analysing variables with more than two categories. In this article an alternative algorithm, which does not have this drawback, was treated and applied. This algorithm, which is a version of the E-M algorithm, produces maximum-likelihood estimates. The new algorithm appeared to be slow but trustworthy when applied to a trichotomous question that is measured three times.

For a dichotomous question that is measured three times there is no test on the goodness of fit of the Markov property, only for the stationarity of the chain. However, when a trichotomous question is analyzed, a test on the Markov property does exist, although the power seems to be small. An extension of the analysis to four measurements would probably increase the power of the goodness-of-fit test. The first order Markov property could be tested upon more thoroughly then, also for dichotomous variables.

Computations on test data gave some insight into the standard deviations of the model parameters that may be expected. For a random sample of 3,000 cases the standard error of latent proportions ranged from .01 to .09. Of course the probability of a transition from a *small* latent class will be estimated with a *high* standard deviation, .03 being a typical value for this sample size. As a rule of thumb, the response categories of the variable to be analyzed should contain at least 500 cases. So a sample of at least 1,000 cases should be recommended for a dichotomous variable and a sample of some 2,000 cases for a trichotomous variable. These are large sizes compared to the size of the data sets that were analyzed by, for instance, Wiggins (1973).

To sum up, the new algorithm performed reasonably well[5] and the application of the model gave valuable insights in the weak points of the variables that were analysed, especially in the (un)reliabilities of the latent classes. Also latent marginal distributions and transition tables were obtained that should be considered as better than the manifest marginals and turnover tables.

NOTES

1. As will be shown in section 5 (Tables 3 and 4), the latent probability distribution will differ from the manifest probability distribution if there is a substantial difference between the reliabilities by which the latent classes are measured. The number of people in unreliably measured latent classes will be underestimated on the manifest level and the number of people in more reliably measured latent classes will be overestimated on the manifest level.

2. Heise's formula for the reliability r_{tt} at time t can be written in terms of correlations between consecutive measurements s, t, and u as $r_{tt} = r_{tu}(r_{su})^{-1} r_{st}$. Coleman's matrix P^{tt} for the turnover at time t with a time lag approaching zero can be expressed in terms of bivariate turnover tables P as $P^{tt} = P^{tu}(P^{su})^{-1} P^{st}$. This analogy was also noted by Henry (1973).

3. Of course there are dependencies in the model parameters like $\Sigma_{c=1}^{C} v_c^t = 1$, that prevent the direct computation of variances of all model parameters. Leaving out for instance the last latent class v_C^t from the computation of the information matrix, the variance of v_C^t may be obtained from the covariances of the other v^t-parameters as follows. Because $v_C^t = 1 - \Sigma_{c=1}^{C-1} v_c^t$ we find

$$\mathrm{var}(v_C^t) = E\left(1 - \Sigma_{c=1}^{C-1} v_c^t\right)^2 - \left[E\left(1 - \Sigma_{c=1}^{C-1} v_c^t\right)\right]^2 \ldots$$

$$= \Sigma_{a=1}^{C-1} \Sigma_{b=1}^{C-1} \mathrm{cov}(v_a^t, v_b^t).$$

4. Analogous to formula 1 one may write for the latent parameters $V^{tu} = V^t M^{tu}$, where V^t is a diagonal matrix with v^t on the diagonal.

5. A (ISO) PASCAL computer program was written that uses the E-M algorithm. Stationarity (equation 4) may be assumed and tested upon by a likelihood ratio test. Three, four, or five measurements can be analyzed of a variable with two or more categories. First-order derivatives and standard errors of the model parameters are computed by numerical differentiation of the likelihood. The program, called LATMARK, runs on a CDC-Cyber and uses the NAG-library for the computation of standard errors (matrix inversion) and for the evaluation of χ^2 values. A tape may be obtained at cost from the Netherlands Central Bureau of Statistics, F. van de Pol, P.O. Box 959, 2270 AZ Voorburg, the Netherlands. A user guide is in preparation.

REFERENCES

ANDERSON, E. B. (1980) Discrete Statistical Models With Social Science Applications. Amsterdam: North-Holland.

BISHOP, Y.M.M., S. E. FIENBERG, and P. W. HOLLAND (1977) Discrete Multivariate Analysis. Cambridge, MA: MIT Press.

BLUMEN, I., M. KOGAN, and P. J. McCARTHY (1966) "Probability models for mobility," pp. 318-334 in P. F. Lazarsfeld and N. W. Henry (eds.) Readings in Mathematical Social Science. Cambridge, MA: MIT Press.

CARROLL, G. C. (1983) "Dynamic analysis of discrete dependent variables: a didactic essay." Quality and Quantity 17: 425-460.

CLOGG, C. C. (1977) "Unrestricted and restricted maximum likelihood latent structure analysis: a manual for users." Working paper 1977-09, Population Issues Research Office, Pennsylvania State University.

———and L. A. GOODMAN (1985) "Simultaneous latent structure analysis in several groups," pp. 81-110 in N. B. Tuma (ed.) Sociological Methodology 1985. San Francisco: Jossey-Bass.

COLEMAN, J. S. (1964) Models of Change and Response Uncertainty. Englewood Cliffs, NJ: Prentice-Hall.

DAYTON, C. M. and G. B. MACREADY (1983) "Latent structure analysis of repeated classifications with dichotomous data." British J. of Mathematical and Stat. Psychology 36: 189-201.

DEMPSTAR, A. P., N. M. LAIRED, and D. B. RUBIN (1977) "Maximum likelihood estimation from incomplete data via the EM-algorithm." J. of the Royal Stat. Society B, 39: 1-22.

EDWARDS, A.W.F. (1976) Likelihood. Cambridge, UK: Cambridge Univ. Press.

GOODMAN, L. A. (1974a) "The analysis of systems of qualitative variables when some of the variables are unobservable, part I—a modified latent structure approach." Amer. J. of Sociology 79: 1179-1259.

———(1974b) "Exploratory latent structure analysis using both identifiable and unidentifiable models." Biometrika 61: 215-231.

HABERMAN, S. J. (1979) Analysis of Qualitative Data, Vol. 2, New Developments. New York: Academic Press.

HAGENAARS, J. A. (1978) "Latent probability models with direct effects between indicators." Quality and Quantity 12: 205-222.

———(1985) LCAG—"loglinear modelling with latent variables; a modified LISREL approach." To be published in the proceedings of the Conference on Methodological Research, Amsterdam.

HEISE, D. R. (1969) "Separating reliability and stability in test-retest correlation." Amer. Soc. Rev. 34: 93-101.

HENRY, N. W. (1973) "Measurement models for continuous and discrete variables," in A. S. Goldberger and O. D. Duncan (eds.) Structural Equation Models in the Social Sciences. New York: Seminar.

JÖRESKOG, K. G. (1970) "Estimating and testing simplex models." British J. of Mathematical and Stat. Psychology 23: 121-145.

KRAEMER, H. C. (1985) "The robustness of common measures of 2 × 2 association to bias due to misclassifications." Amer. Statistician 39: 286-290.

LAZARSFELD, P. F. and N. W. HENRY (1968) Latent Structure Analysis. Boston: Houghton-Mifflin.

LOGAN, J. A. (1981) "A structural model of the higher order Markov process incorporating reversion effects." J. of Mathematical Sociology 8: 75-89.

MARKUS, G. B. (1979) Analyzing Panel Data. Beverly Hills, CA: Sage.

PETERSSON, O. (1974) "Change and unreliability." Quality and Quantity 8: 247-266.

POL, F.J.R. van de (1982) "Betrouwbaarheidsbepaling met herhaalde metingen." Mens en Maatschappij 57: 70-95.

SCHWARTZ, J. E. (1985) "The neglected problem of measurement error in categorical data." Soc. Methods & Research 13: 435-466.

SINGER, B. and S. SPILERMAN (1978) "Some methodological issues in the analysis of longitudinal surveys," pp. 261-296 in K. E. Taubner et al. (eds.) Social Demography. New York: Academic Press.

STADT, H. van de, A. TEN CATE, A. J. HUMDEPOOL, and W. J. KELLER (1985) "Koopkrachtontwikkelingen 1977-1983." Economisch Statistische Berichten 70: 288-296.

TUMA, N. B., M. T. HANNAN, and L. P. GROENEVELD (1979) "Dynamic analysis of event histories." Amer. J. of Sociology 84: 820-854.

WOLFE, J. H. (1970) "Pattern clustering by multivariate mixture analysis." Multivariate Behavioral Research 5: 329-350.

WIGGINS, L. M. (1955) "Mathematical models for the analysis of multi-wave panels." Ph.D. dissertation, Columbia University, New York.

———(1973) Panel Analysis, Latent Probability Models for Attitude and Behavior Processes. Amsterdam: Elsevier.

WU, C.F.J. (1983) "On the convergence properties of the EM algorithm." Annals of Statistics 11: 95-103.

Frank van de Pol is Research Associate at the Netherlands Central Bureau of Statistics. He is writing about design and analysis of panel surveys and worked on time series analysis with intervention effects (with Joop de Beer). Recent publications in Sociometric Research 1985 *(to appear), (W. E. Saris and I. N. Gallhofer, eds.) and in Dutch journals.*

Jan de Leeuw is Professor of Data Theory at the University of Leyden. He is interested in multivariate data analysis techniques for qualitative and quantitative data. Recent publications appeared in Psychometrika, Journal of Econometrics, Computational Statistics Quarterly, *and* Journal of Classification.