Estimation in complex latent transition models with extreme data sparseness

Linda M. Collins and Allison J. Tracy[1]

Abstract
Latent transition analysis (LTA) is a type of latent class model for longitudinal data that emphasizes the use of multiple indicators. It is useful in estimating and testing stage-sequential models of human development in longitudinal data. One practical issue that arises frequently in connection with LTA is whether these models can be tested with modest sample sizes. The question arises because in any latent class model, the sufficient statistics are the cell frequencies of the multiway contingency table formed by cross tabulating all the items. In complex LTA models, such as models involving multiple items at multiple times, this contingency table involves many cells, often in the thousands or tens of thousands. If the sample of subjects is small in relation to the size of the contingency table, there will be many empty cells. Furthermore, these models often involve estimation of a large number of parameters. This is particularly true with second-order LTA models. The purpose of this paper is to investigate the extent to which estimation can proceed successfully under these conditions.

1. Introduction

1a. Introduction to LTA

Latent transition analysis (LTA) is a method for estimating and testing stage-sequential models of human development in longitudinal data. LTA is a type of latent class model (Collins & Wugalter, 1992; Van de Pol & De Leeuw, 1986) that emphasizes the measurement of stage transitions over time through the use of multiple indicators. Consider a model of math skill acquisition tested in Collins and Wugalter (1992). For purposes of the present article, suppose the model is tested on data collected from 1500 United States students in tenth, eleventh, and twelfth grade. Suppose we are interested in testing a model of math skill development over time, where individuals start out with no math skills; first learn single operations on whole numbers; then progress to powers and roots, decimals, and fractions; then learn low level algebra without word problems; then go on to low level geometry and algebra with word problems. In this model, it is possible to advance or to decline. When an individual progresses to the next stage, all skills learned in earlier stages are retained; when an individual declines in skill, skills are lost in the order in which they were gained. This model is depicted in Figure 1, which is taken from Collins and Wugalter (1992). In LTA each stage is called a *latent status*. Further suppose that we wish to compare the math skill acquisition of two groups: those who on a questionnaire indicate a high interest in a career in mathematics, science, or engineering, and those who indicate a low interest in such a career. The changing math skills are a dynamic

---

[1]The Methodology Center and Department of Human Development & Family Studies, Pennsylvania State University, 159 Henderson Building South, University Park, PA 16802, USA. The first author's email address is LMC8@PSU.EDU. This research was supported by a grant from the United States National Institute on Drug Abuse, R01 DA04111.

endogenous latent variable, and the interest group is a static exogenous latent variable.

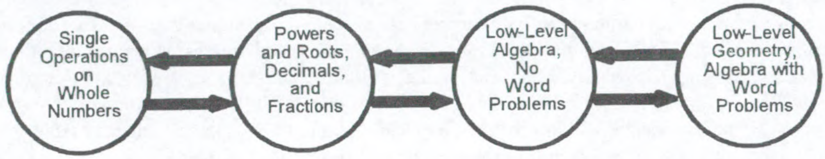

Figure 1. Model of math skill development from Collins and Wugalter (1992).

Like structural equation modeling, LTA is based on the idea of a latent variable measured by manifest indicators. Whereas in covariance structure modeling the latent variable and its indicators are usually continuous, in LTA the latent variable and its indicators are categorical. In the example used in Collins and Wugalter (1992), the four indicators of the math skills acquisition latent variable were four dichotomized "testlets" corresponding to each of the four skills making up the acquisition process.

1b. LTA mathematical model

For ease of exposition the latent transition model will be presented for problems involving three occasions of measurement, four manifest indicators (items or variables) of the dynamic latent variable at each occasion, and one exogenous static latent variable measured by one manifest indicator; the extension to other problems is direct. To relate this to the math skills acquisition example, the three occasions of measurement are tenth, eleventh, and twelfth grade, the dynamic latent variable is math skill development, and the exogenous static latent variable is career interest. The first occasion of measurement will be labeled Time t, the second Time t+1, and the third Time t+2. Also suppose the four manifest indicators are Item 1, with $i, i', i''=1,\ldots I$ response categories; Item 2, with $j, j', j''=1,\ldots J$ response categories, Item 3, with $k, k', k''=1,\ldots K$ response categories, and Item 4, with $l, l', l''=1,\ldots L$ response categories, where i, j, k and l refer to responses obtained at Time t; i', j', k' and l' refer to responses obtained at Time t+1; and i'', j'', k'' and l'' refer to responses obtained at Time t+2. The exogenous static latent variable divides the population into latent classes $c = 1,\ldots C$, and is measured by a manifest indicator with $m = 1,\ldots M$ response categories. There are $p, q, r = 1,\ldots S$ latent statuses, with p denoting a latent status at Time t, q denoting a latent status at Time t+1, and r denoting a latent status at Time t+2. Let $Y = \{m,i,j,k,l,i',j',k',l',i'',j'',k'',l''\}$ represent a "response pattern", a vector of possible responses made up of a single response to the manifest indicator of the exogenous variable and responses to the four items at Times t, t+1, and t+2. Then the estimated proportion of a particular response pattern, $P(Y)$, is expressed as follows for a first-order model:

$$P(Y)=\sum_{c=1}^{C}\sum_{p=1}^{S}\sum_{q=1}^{S}\sum_{r=1}^{S}\gamma_c\rho_{m|c}\delta_{p|c}\rho_{i|p,c}\rho_{j|p,c}\rho_{k|p,c}\rho_{l|p,c}\tau_{q|p,c}\rho_{i'|q,c}\rho_{j'|q,c}\rho_{k'|q,c}\rho_{l'|q,c}\tau_{r|q,c}\rho_{i''|r,c}\rho_{j''|r,c}\rho_{k''|r,c}\rho_{l''|r,c}$$

where

γ_c represents the proportion in latent class c. In the example, there would be two gamma parameters, the proportion of students who indicate a high interest in a career in math, science, or engineering, and those who indicate a low interest in this kind of career.

$\delta_{p|c}$ represents the proportion in latent status p at Time t conditional on membership in latent class c. In our example this is the proportion of individuals in each of the stages in the model, for example, the proportion in the no skills latent status, the proportion in the simple operations on whole numbers latent status, and so on, conditional on membership in the high interest or low interest group.

$\tau_{q|p,c}$ is an element of the latent transition probability matrix, representing the probability of membership in latent status q at Time t+1 conditional on membership in latent status p at Time t and membership in latent class c. An example is the probability of membership in the simple operations on whole numbers latent status in twelfth grade, conditional on membership in the no skills latent status in tenth grade and membership in the low interest group.

$\rho_{i|p,c}$ represents the probability of response i to Item 1 at Time t, conditional on membership in latent status p at Time t and on membership in latent class c; $\rho_{i'|q,c}$ represents the probability of response i' to Item 1 at Time t+1, conditional on membership in latent status q at Time t+1 and on membership in latent class c, etc. An example of a ρ parameter is the probability of passing the testlet about power and roots, decimals, and fractions conditional on membership in the no skills latent status and the low interest latent class.

$\rho_{m|c}$ represents the probability of having a value of m on the indicator of latent class membership, conditional on membership in latent class c. In the example this might represent the probability of choosing the "low interest" response to the questionnaire item asking about interest in a career in math, science, or engineering, conditional on membership in the low interest latent class.

Sometimes transitions between latent statuses are conditional not only on the immediately previous latent status, but on the latent status two observations previous. Then the data are best represented by a second-order model:

$$P(Y)=\sum_{c=1}^{C}\sum_{p=1}^{S}\sum_{pq=1}^{S^2}\sum_{r=1}^{S}\gamma_c\rho_{m|c}\delta_{p|c}\rho_{i|p,c}\rho_{j|p,c}\rho_{k|p,c}\rho_{l|p,c}\tau_{q|p,c}\rho_{i'|q,c}\rho_{j'|q,c}\rho_{k'|q,c}\rho_{l'|p,c}\tau_{r|pq,c}\rho_{i''|r,c}\rho_{j''|r,c}\rho_{k''|r,c}\rho_{l''|r,c}$$

where

$\tau_{r|pq,c}$ is an element of the latent transition probability matrix, representing the probability of membership in latent status r at Time t+2 conditional on membership in latent status p at Time t, membership in latent status q at Time t+1, and membership in latent class c.

The first-order model is a special case of the second-order model where $\tau_{r|pq,c}$ for a given q and c is equal across all p's. It is necessary to have at least three occasions of measurement in order to test a second-order model.

1c. The problem addressed in this study

As LTA models are used in an increasing variety of applications, it is inevitable that users will want to apply the procedure to larger models. The term "large" in this context is not what would be considered large in the context of a structural equation model. All current latent class models, including LTA, are intended for problems where there are relatively few indicators of the latent variable. This is because these procedures are special cases of loglinear models, and therefore they build multi-way contingency tables, which can become very large if there are numerous items or times. For example, suppose a problem involves four dichotomous indicators of the latent variable at each time. If there are three times, there are 4,096 possible response patterns. If there are four times, there are 65,536 possible response patterns. If there are five times, there are over one million possible response patterns! The potential problem lies not in the number of response patterns per se, but rather in the number of subjects in relation to the number of response patterns. If there are data available on millions of subjects, over 65 thousand response patterns does not present a problem. However, if the number of response patterns is large and the number of subjects is small, a sparse data matrix results. This presents two problems. First, the goodness-of-fit statistic G^2 is not distributed as a chi-square when data are sparse, creating difficulties for model selection. This problem is not the focus of this article; for a more thorough discussion, see Collins, Fidler, Wugalter, and Long (1994) and Langeheine, Pannekoek, and Van de Pol (1995). Instead, this article focuses on the second problem, which has to do with estimation. Under conditions of sparseness, sometimes there may not be enough information in the data to provide good parameter estimates. This is an important practical consideration, because although studies with thousands of participants exist, it is more typical for the number of subjects taking part in a study to be in the range of 300 to 1500.

The number of parameters to be estimated is another important consideration. Suppose we are interested in a model with five latent statuses. Assuming no constraints are imposed on the parameters, the number of parameters for a first order model increases from 69 with a two-time problem to 204 for a five-time problem; for a second-order model, the increase is from 69 to 504 parameters. In reality, constraints would be put on some parameters, which would reduce the parameter estimation load, but even with constraints added this is potentially a great deal of estimation.

Thus, the effects of increasing the size of the problem, even to what most researchers would consider modest, can be severe in terms of the size of the contingency table and the number of parameters estimated. This raises an important question. Can estimation be successful, that is, unbiased and reasonably efficient, when the problem is large and the sample size is moderate? Collins, Fidler, and Wugalter (1996) investigated the issue of parameter recovery in large problems. They varied N/k, where N is the sample size and k represents the number of cells in the contingency table. They found very good parameter recovery overall, even with N/k as low as .5. Although these results are encouraging, they do not go far enough. A model with four dichotomous indicators and three times involving a sample size of

300 results in an N/k of about .07. Thus, for very large problems and moderate sample sizes, parameter recovery has not been investigated. The present article describes a simulation study that investigates this.

Table 1
Parameter values used to generate data

δ parameters

	Ls 1	Ls 2	Ls 3	Ls 4	Ls 5
	.40	.30	.20	.05	.05

τ parameters

Time t+1 / Time t	Ls 1	Ls 2	Ls 3	Ls 4	Ls 5
Latent status 1	.40	.30	.20	.05	.05
Latent status 2	.05	.40	.30	.20	.05
Latent status 3	.05	.05	.40	.30	.20
Latent status 4	.05	.05	.05	.50	.35
Latent status 5	.05	.10	.10	.10	.65

ρ parameters

	Item 1	Item 2	Item 3	Item 4
Latent status 1	.10	.10	.10	.10
Latent status 2	.90	.10	.10	.10
Latent status 3	.90	.90	.10	.10
Latent status 4	.90	.90	.90	.10
Latent status 5	.90	.90	.90	.90

2. Methods

2a. Overview

The purpose of this study is to investigate whether bias in parameter estimation is introduced in large problems with relatively small sample sizes, and to see whether the mean squared error (MSE) of parameters is acceptable under these conditions. LTA models were used as a basis for generating random data sets with known structure. These known models were estimated in the random data sets, so that the null hypothesis was true in every case. Thus model misspecification was not present to bias the parameter estimates. After the LTA models were estimated in the data, the parameter estimates obtained were compared to the known population parameter values, and bias and MSE were computed.

2b. Design of the study and data generation

All data were generated using the same basic model. This model involved five latent statuses and four dichotomous indicators at each of three times. Number of subjects was either N=300 or N=1500. Table 1 shows the parameter values used to generate the data.

The data generation procedure followed was the same as the one used in Collins et al. (1996). Based on the model and the previously specified parameter values shown in Table 1, it is possible to construct a vector of response pattern probabilities, for example,

Response Pattern	Probability	Cumulative Probability
111111111111	.0203	.0203
111111111112	.0023	.0226
111111111121	.0023	.0249

and so on for the entire vector of possible response patterns. These are the population probabilities of a randomly selected individual contributing a particular response pattern. For our purposes, we wish to draw random samples of either N=300 or N=1500 from this population. To do so, we use this vector of probabilities very much like a multi-sided die, with the sides weighted according to the probabilities. Each artificial "subject" was generated by sampling a number between 0 and 1 from a uniform random distribution. In order to assign a response pattern to the "subject," the random number was compared to the cumulative probability. For example, if the random number is greater than .0226 and less than or equal to .0249, it is assigned to the response pattern 111111111121.

2c. Parameter constraints

In order to ensure identification, we added some constraints to the models. First, in all models tested we constrained the ρ's equal across times. Second, we placed additional constraints on the ρ parameters for all models and in several additional places in the second-order models. The additional constraints on the ρ parameters in the first-order models are summarized in Table 2. Two different sets of constraints were used. We will refer to these as A constraints and C constraints. The A constraints are more parsimonious, resulting in only one ρ parameter estimated. The C constraints are less parsimonious, resulting in four ρ parameters estimated. These same constraints were used in the second-order models, along with the additional constraints shown in Table 3.

Table 2
Constraints on ρ parameters

A Constraints: Probability of passing*

	Item 1	Item 2	Item 3	Item 4
LS1	e	e	e	e
LS2	d	e	e	e
LS3	d	d	e	e
LS4	d	d	d	e
LS5	d	d	d	d

C constraints: Probability of passing*

	Item 1	Item 2	Item 3	Item 4
LS1	e	g	i	k
LS2	d	g	i	k
LS3	d	f	i	k
LS4	d	f	h	k
LS5	d	f	h	j

*Constraints for the ρ parameters corresponding to probability of failing are the complements of the constraints shown in this table.

Table 3
Additional constraints used in second-order models*

δ parameters

	Ls 1	Ls 2	Ls 3	Ls 4	Ls 5
	1	1	1	b	b

τ parameters**

Time t+1 / Time t	Ls 1	Ls 2	Ls 3	Ls 4	Ls 5
Latent status 1	1	1	1	c	c
Latent status 2	d	1	1	1	d
Latent status 3	e	e	1	1	1
Latent status 4	f	f	f	1	1
Latent status 5	1	g	g	g	1

* Parameters denoted by the number 1 are freely estimated. Parameters denoted by the same letter are constrained equal to each other.

** This pattern of constraints was repeated for the entire τ matrix.

One hundred data sets with N=300 and one hundred data sets with N=1500 were generated. Each data set was analyzed four ways: as a first-order model using A constraints; as a first-order model using C constraints; as a second-order model using A constraints; and as a second-order model using C constraints. Thus, there were three independent variables in the simulation: sample size (300 vs. 1500), type of constraints (A vs. C), and order of the process (first-order vs. second-order).

2d. Data analysis

Because we generated the artificial data analyzed in this study, we are in the unique position of knowing exactly which model generated the data. In each data set, this model was tested, so the null hypothesis was always true. Both first-order and second-order versions of each model were tested. The data were generated using a first-order model; however, a first-order model is identical to a second-order model where the probability of a transition between Time 2 and Time 3 does not vary according to latent status membership at Time 1. Therefore, a second-order model fits identically to a first-order model, although it is not a parsimonious model because it involves estimating parameters that are not needed.

For each analysis performed on each data set, two different sets of start values were used for each run. The same two sets of start values were used for each data set. Convergence for the EM algorithm was defined as a mean absolute deviation between successive parameter estimates smaller than .00001. On the rare occasions where there were differences in fit between the two sets of start values, the run with the smaller G^2 was selected for further analysis.

Often in latent class models a problem comes up that is sometimes called a "naming problem." This occurs when two or more latent classes are not conceptually distinct because the pattern of the ρ parameters across items is very similar. For example, in an ability test the probability of passing might be .05 for all items in one latent class and .2 for all items in another latent class. Both of these latent classes can be interpreted as "No knowledge." This poses a particular problem for simulation studies because when two latent classes are not distinct, it is impossible to tell which parameters "belong to" a particular latent class or latent status. In the present study, data sets were discarded when the LTA analysis resulted in latent statuses that were not distinct enough to determine which of the parameters in the data-generating model were associated with them. In other words, the pattern of the values of the ρ estimates had to correspond at least roughly to the ρ's in the known true model. The naming problem did not occur at all in first-order models, or in second-order models with A constraints. In the second-order models with C constraints, the naming problem occurred in about three percent of data sets with N=1500 and about one-third of data sets with N=300. Data sets discarded for this reason were replaced with additional random data sets until the required number of data sets was obtained.

3. Results

Tables 4-9 contain results presented in two slightly different ways. In each table the first panel contains data from the same 100 data sets in each row. This means that if a data set failed to meet one of the two criteria explained above in <u>any</u> cell, it was removed from <u>all</u> analyses. The data in the first panel facilitate straightforward comparisons, but the inclusion criteria are restrictive. The second panel contains data from the first 100 data sets generated that meet the requirements for a particular cell. Thus a data set with n=300 may appear in one, two, three, or four cells of the second panel. This makes comparisons across cells more difficult, but allows considerably less restrictive inclusion criteria.

Table 4 shows the average number of iterations in each cell required for convergence. It is clear that overall fewer iterations are needed for first-order than for second-order models. There does not seem to be a large difference between the average number of iterations needed for A constraints as opposed to C constraints, or for the smaller N as opposed to the larger N.

Table 4
Average number of iterations required for convergence

Same 100 data sets in all cells

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	46.21	47.44	109.78	121.22
N = 1500	36.14	32.70	114.49	120.39

First 100 data sets meeting inclusion criteria in each cell

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	47.79	49.37	108.49	121.22
N = 1500	36.01	32.91	115.35	120.39

Bias is defined as the average difference between the parameter estimate and the true parameter. Mean squared error (MSE) is defined as the average squared difference between the parameter estimate and the true parameter. There are many parameters that could be examined in this results section. In order to save space, and because the results are generally consistent across parameters of the same type, we have selected a subset of parameters to present here. Tables 5 - 9 show bias and MSE for these selected parameters.

Tables 5 and 6 show the results for the δ (parameter=.4) and ρ (parameter=.9) respectively. The results for these two parameters are very similar. There is virtually no bias associated with either parameter when N=1500, even in the second order models. When N=300 there is somewhat more bias overall. However, nowhere does the bias exceed .004 in absolute value. MSE associated with these parameters is generally small in every cell, not exceeding .001.

Table 5
Bias and mean squared error (in parentheses) for the δ parameter

Same 100 data sets in all cells

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.003 (.001)	-.002 (.001)	-.002 (.001)	-.002 (.001)
N = 1500	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)

First 100 data sets meeting inclusion criteria in each cell

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.001 (.001)	-.001 (.001)	-.001 (.001)	-.002 (.001)
N = 1500	-.000 (.000)	-.000 (.000)	-.000 (.000)	-.000 (.000)

Table 6
Bias and mean squared error (in parentheses) for the ρ parameter

Same 100 data sets in all cells

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.000 (.000)	.002 (.000)	.002 (.000)	.004 (.000)
N = 1500	-.000 (.000)	-.000 (.000)	.000 (.000)	.000 (.000)

First 100 data sets meeting inclusion criteria in each cell

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.000 (.000)	.003 (.000)	.002 (.000)	.004 (.000)
N = 1500	-.000 (.000)	-.000 (.000)	.000 (.000)	.000 (.000)

Results for the first and second τ parameters appear in Tables 7 and 8 respectively. The first τ (parameter=.4) considered here corresponds to the probability of membership in latent status 1 at Time 2, conditional on membership in latent status 1 at Time 1. This the only one of the τ parameters considered here that is directly comparable across the first-order and second-order models. This is because the Time 1-Time 2 transition probability matrix is the same for first-order and second-order models. However, the Time 2-Time 3 transition probability matrix and any subsequent τ matrices differ. First-order models contain a simple Time 2-Time 3 transition probability matrix, while in second-order models the Time 2-Time 3 transition probability matrix is conditioned on Time 1 latent status. This means that there are many more τ parameters in a second-order model than in a first-order model, and that the parameters have a slightly different meaning. The second τ (parameter=.4) selected for the first-order model is the probability of membership in latent status 1 at Time 3, conditional on membership in latent status 1 at Time 2. In contrast, the second τ (parameter=.4) selected for the second-order model corresponds to the probability of membership in latent status 1 at Time 3, conditional on membership in latent status 1 at Time 2 *and* Time 1.

Tables 7 and 8 show that somewhat more bias is evident in the τ parameters than appears in the δ or ρ parameters. For the first τ parameter the largest bias in absolute magnitude is approximately -.013. This occurs in the second order models where N=300, for the data based on the same 100 data sets. Bias is reduced considerably when N=1500, and appears to be somewhat less in the data based on the first 100 data sets meeting the inclusion criteria in each cell. MSE is around .003 when N=300 and around .001 when N=1500 for both first-order and second-order models. For the second τ parameters bias is larger, with the largest bias occurring in the second-order models. The MSE associated with these parameters is also considerably larger, although still not exceeding approximately .011.

Table 7
Bias and mean squared error (in parentheses) for the first τ parameter

Same 100 data sets in all cells

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.009 (.003)	-.010 (.003)	-.013 (.003)	-.011 (.003)
N = 1500	.000 (.001)	-.000 (.001)	-.000 (.001)	-.000 (.001)

First 100 data sets meeting inclusion criteria in each cell

	1st order		2nd order	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.004 (.003)	-.004 (.003)	-.007 (.003)	-.011 (.003)
N = 1500	.000 (.001)	.000 (.001)	.000 (.001)	-.000 (.001)

Table 8
Bias and mean squared error (in parentheses) for second τ parameter

Same 100 data sets in all cells

	1st order A constraints	1st order C constraints	2nd order A constraints	2nd order C constraints
N = 300	.019 (.008)	.017 (.008)	.025 (.011)	.027 (.011)
N = 1500	.004 (.002)	.004 (.002)	.003 (.002)	.002 (.002)

First 100 data sets meeting inclusion criteria in each cell

	1st order A constraints	1st order C constraints	2nd order A constraints	2nd order C constraints
N = 300	.009 (.008)	.008 (.008)	.014 (.010)	.027 (.011)
N = 1500	.006 (.002)	.006 (.002)	.004 (.002)	.002 (.002)

Two additional τ parameters are examined. These parameters are taken from the second-order models; no corresponding parameters exist in first-order models. The third τ parameter examined is the probability of membership in latent status 1 at Time 3, conditional on membership in latent status 1 and Time 2 *and* membership in latent status 3 at Time 1; and the fourth τ parameter examined is the probability of membership in latent status 1 at Time 3, conditional on membership in latent status 1 at Time 2 *and* membership in latent status 5 at Time 1. Table 9 combines the results for these τ parameters in one table. The table shows a pattern of increasing bias and MSE. Overall, there is more bias in the third τ parameter than in the second, and more in the fourth than in the third. On the average, MSE's are larger in the fourth τ parameter, but the largest MSE's occur in the third τ parameter when N=300. In sharp contrast to the other parameters, in the fourth τ parameter the MSE does not differ very much between the N of 300 and the N of 1500. In fact, the MSE's observed in the fourth τ parameter and in the smaller N in the third τ parameter would probably be considered unacceptable by most researchers.

Table 9
Bias and mean squared error (in parentheses) for third and fourth τ parameters, second-order models

Same 100 data sets in all cells

	Third τ		Fourth τ	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.047 (.131)	-.044 (.131)	-.077 (.117)	-.083 (.119)
N = 1500	-.007 (.057)	-.008 (.056)	-.072 (.120)	-.067 (.119)

First 100 data sets meeting inclusion criteria in each cell

	Third τ		Fourth τ	
	A constraints	C constraints	A constraints	C constraints
N = 300	-.011 (.138)	-.044 (.131)	-.087 (.112)	-.083 (.119)
N = 1500	.004 (.059)	-.008 (.056)	-.065 (.122)	-.067 (.119)

4. Discussion

The results of this simulation are very encouraging for estimation of complex LTA models, and, by extension, latent class and loglinear models. There is very little bias in the δ or ρ parameters, even in the complex second-order models. In addition, the MSE's associated with these parameters are small, rounding to .000. Bias and MSE's are generally better in the conditions with the larger N (which is an N/k of only about .37). There is a slight tendency for bias to improve when the more restrictive A constraints are imposed.

The most bias and the largest MSE's were found in the second-order models in the third and fourth τ parameters. The reason for this has to do with the nature of these parameters in the models used here to generate the data. As explained above, the third and fourth τ parameters represent the probability of membership in the first latent status at Time 3, conditional on membership in the first latent status at Time 2 and membership in the third or fifth latent status, respectively, at Time 1. The unconditional probability of membership in the third latent status at Time 1 is .2; the unconditional probability of membership in the fifth latent status at Time 1 is .05. The probability of membership in the first latent status at Time 3 *and* one of these latent statuses at Time 1 *and* another latent status at Time 2 is small, and therefore the estimation of these quantities is less stable. In other models for other applications, parameters like these will occur elsewhere in the model.

We attempted to replicate this simulation using ρ=.65, and encountered many problems. We found that the programs took a very long time to run, and about one-third failed to converge after 1000 iterations of the EM algorithm. This suggests that maintaining reasonably strong ρ parameters is important for estimating large and complex models.

5. Conclusions

The results of this study suggest that estimation of parameters for complex LTA models, and, by implication, for other latent class and loglinear models, is surprisingly robust even with relatively small N's. While bias and unacceptably large MSE's were a problem for some τ parameters, the δ and ρ parameters and most of the τ parameters were estimated with little bias and acceptably small MSE's. On a less encouraging note, our study points to some limitations researchers using LTA and related models. First, several of the parameters in the second-order models had degrees of bias and MSE that would make most researchers uncomfortable. This suggests that second-order models should be used with extreme caution. Second, it is important to note that the ρ parameters used in this study were all .9 or .1. Our attempt to replicate this study with considerably weaker ρ's indicates that estimation can be difficult under such conditions. Thus it is worthwhile to spend time and attention constructing good items to serve as manifest indicators, because this will help strengthen the ρ parameters. This strategy is greatly aided by having a good background in the substantive field to which the model pertains.

It should also be reiterated that although this study supports the contention that parameter estimation can be robust under conditions of data sparseness, serious model selection problems remain because the likelihood ratio statistic G^2 is not distributed as a chi-square. A new approach to model selection is needed. In the meantime, Collins et al. (1994) and Langeheine et al. (1995) outline Monte Carlo simulation methods for approximating the correct probability of the obtained G^2.

6. References

Collins, L.M., Fidler, P.L., & Wugalter, S.E. (1996). Some practical issues related to estimation of latent class and latent transition parameters. In A. von Eye & C. C. Clogg (Eds.), *Analysis of categorical variables in developmental research*, pp. 133-146.

Collins, L.M., Fidler, P.L., Wugalter, S.E., & Long, J.L. (1993). Goodness-of-fit testing for latent class models. *Multivariate Behavioral Research, 28*, 375-389..

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

Langeheine, R., Pannekoek, J., & Van de Pol,. F. (1995). Bootstrapping goodness-of-fit measures in categorical data analysis (in preparation). Voorburg/Kiel: NCBS/IPN.

Van de Pol, F., & De Leeuw, J. (1986). A latent Markov model to correct for measurement error. *Sociological Methods and Research, 15*, 118-141.

ontvangen: 15- 6-95
geaccepteerd: 23- 8-96