

9

Structural Models for Categorical and Continuous Latent Variables

This chapter describes what can be reasonably considered the state of the art in structural equation modeling—namely, structural equation models that combine categorical and continuous latent variables for cross-sectional and longitudinal designs. The comprehensive modeling framework described in this chapter rests on the work of B. Muthén (2002, 2004), which builds on the foundations of finite mixture modeling (e.g., McLachlan & Peel, 2000) and conventional structural equation modeling for single and multiple groups as described in Chapter 4.

It is beyond the scope of this chapter to describe every special case that can be accommodated by the general framework. Rather, this chapter touches on a few key methods that tie into many of the previous chapters. The organization of this chapter is as follows. First, we set the stage for the applications of structural equation modeling for categorical and continuous latent variables with a brief review of finite mixture modeling and the expectation-maximization (EM) algorithm, following closely the discussion given in McLachlan and Peel (2000). This is followed by a discussion of applications of finite mixture modeling for categorical outcomes leading to latent class analysis and variants of Markov chain modeling. Next, we discuss applications of finite mixture modeling to the combination of continuous and categorical outcomes, leading to growth mixture modeling. We focus solely on growth mixture modeling because this methodology encompasses structural equation modeling, factor analysis, and growth curve modeling for continuous outcomes. The chapter closes with a brief overview of other extensions of the general framework that relate to previous chapters of this book.

9.1 A Brief Overview of Finite Mixture Modeling

The approach taken to specifying models that combine categorical and continuous latent variables is *finite mixture modeling*. Finite mixture modeling relaxes the assumption that a sample is drawn from a population characterized by a single set of parameters. Rather, finite mixture modeling assumes that the population is composed of a mixture of unobserved subpopulations characterized by their own unique set of parameters.

To fix notation, let $\mathbf{z} = (z'_1, z'_2, \dots, z'_n)'$ denote the realized values of a p -dimensional random vector $\mathbf{Z} = (Z'_1, Z'_2, \dots, Z'_n)'$ based on a random sample of size n . An element Z_i of the vector \mathbf{Z} has an associated probability density function $f(z_i)$. Next, define the finite mixture density as

$$f(\mathbf{z}_i) = \sum_{k=1}^K \pi_k f_k(\mathbf{z}_i), \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, K), \quad [9.1]$$

where $f_k(\mathbf{z}_i)$ are component densities with mixing proportions ($0 \leq \pi_k \leq 1$) and $\sum_{k=1}^K \pi_k = 1$.

It may be instructive to consider how data are generated from a K -class finite mixture model.¹ Following McLachlan and Peel (2000), consider a categorical random variable C_p referred to here as a *class label*, which takes on values $1, 2, \dots, K$ with associated probabilities $\pi_1, \pi_2, \dots, \pi_k$. In this context, the conditional density of Z_i given that the class label $C_i = k$ is $f_k(\mathbf{z}_i)$ and the marginal density of Z_i is $f(\mathbf{z}_i)$.

We can arrange the class label indicators in a K -dimensional vector denoted as $\mathbf{C} = (C'_1, C'_2, \dots, C'_n)'$ with corresponding realizations $\mathbf{c} = (c_1, c_2, \dots, c_n)'$. Here, the elements of \mathbf{c}_i are all zero except for one element whose value is unity indicating that \mathbf{z}_i belongs to the k th mixture class. It follows then, that the K -dimensional random vector \mathbf{C}_i possesses a multinomial distribution, namely,

$$\mathbf{C}_i \sim \text{Mult}_K(1, \boldsymbol{\pi}), \quad [9.2]$$

where the elements of $\boldsymbol{\pi}$ defined earlier arise from the fact that

$$\text{pr}\{\mathbf{C}_i = \mathbf{c}_i\} = \pi_1^{c_{1i}} \pi_2^{c_{2i}} \cdots \pi_K^{c_{Ki}}. \quad [9.3]$$

A practical way of conceptualizing the finite mixture problem is to imagine that the vector \mathbf{Z}_i is drawn from population J consisting of K groups (J_1, J_2, \dots, J_K) with proportions $\pi_1, \pi_2, \dots, \pi_k$. Then, the density function of \mathbf{Z}_i in group J_k given $C_i = k$ is $f_k(\mathbf{z}_i)$ for $k = 1, 2, \dots, K$. Note that the proportion

π_k can be thought of as the prior probability that individual i belongs to mixture class k . Thus, from Bayes's theorem, the posterior probability that individual i belongs to class k given the data \mathbf{z}_i can be written as

$$\tau_k(\mathbf{z}_i) = \frac{\pi_k f_k(\mathbf{z}_i)}{f(\mathbf{z}_i)}, \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, K). \quad [9.4]$$

Estimated posterior probabilities from Equation [9.4] provide one approach for assessing the adequacy of the finite mixture model, as will be demonstrated in the examples below.

In the context of this chapter, it is necessary to provide a parametric form of the finite mixture model described in this section. The parametric form of the finite mixture model in Equation [9.1] can be written as

$$f(\mathbf{z}_i; \boldsymbol{\Omega}) = \sum_{k=1}^K \pi_k f_k(\mathbf{z}_i; \boldsymbol{\theta}_k), \quad (i = 1, 2, \dots, n; k = 1, 2, \dots, K), \quad [9.5]$$

where $\boldsymbol{\Omega}$ is a parameter vector containing the unknown parameters of the mixture model, namely,

$$\boldsymbol{\Omega} = (\pi_1, \pi_2, \dots, \pi_{K-1}, \boldsymbol{\Theta}), \quad [9.6]$$

where $\boldsymbol{\Theta}$ contains the parameters $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_K$, and where

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_K) \quad [9.7]$$

is the vector of mixing proportions defined earlier. Because the probabilities in Equation [9.7] sum to unity, one of them is redundant as represented in Equation [9.6]. As outlined below, the vector $\boldsymbol{\Theta}$ will contain the parameters of the various models under consideration—such as growth mixture models. For now, we consider $\boldsymbol{\Theta}$ to be any general parameter vector whose elements are distinct from $\boldsymbol{\pi}$.

9.2 The Expectation-Maximization Algorithm

Standard estimation algorithms for structural equation models, such as maximum likelihood (ML) and the class of weighted least squares estimators, were covered in Chapter 2. The method of estimation typically employed for finite mixture models is ML using the EM algorithm. The EM algorithm was originally developed as a means of obtaining maximum likelihood estimates in the

context of incomplete data problems (Dempster et al., 1977; see also Little & Rubin, 2002). However, it was soon recognized that a wide array of statistical models, including the latent class model, could be conceptualized as incomplete data problems, including finite mixture modeling. Specifically, in context of finite mixture models, the component label vector \mathbf{c} is not observed. The EM algorithm proceeds by specifying the complete-data vector, denoted here as

$$\mathbf{z}_{\text{comp}} = (\mathbf{z}', \mathbf{c}')'. \quad [9.8]$$

The complete-data log-likelihood must account for the distribution of the class-label indicator vector as well as the distribution of the data. Thus, from Equation [9.5], the complete data log-likelihood for $\mathbf{\Omega}$ can be written as

$$\log L_{\text{comp}}(\mathbf{\Omega}) = \sum_{k=1}^K \sum_{i=1}^n c_{ik} \{\log \pi_k + \log f_k(\mathbf{z}_i | \boldsymbol{\theta})\}, \quad [9.9]$$

where c_{ik} is an element of \mathbf{c} . The form of Equation [9.9] shows the role of c_{ik} as an indicator of whether individual i is a member of class k .

The EM algorithm involves two steps. The E-step begins by taking the conditional expectation of Equation [9.9] given the observed data \mathbf{z} using the current estimates of $\mathbf{\Omega}$ based on a set of starting values, say $\mathbf{\Omega}^{(0)}$. Following McLachlan and Peel (2000), the conditional expectation is written as

$$Q(\mathbf{\Omega}; \mathbf{\Omega}^{(0)}) = E_{\mathbf{\Omega}^{(0)}} \{\log L_{\text{comp}}(\mathbf{\Omega}) | \mathbf{z}\}. \quad [9.10]$$

Let $\mathbf{\Omega}^{(m)}$ be the updated value of $\mathbf{\Omega}$ after the m th iteration of the EM algorithm. Then the E-step on the $(m + 1)$ th iteration calculates $Q(\mathbf{\Omega}, \mathbf{\Omega}^{(m)})$.

With regard to the class-label vector \mathbf{c} , the E-step of the EM algorithm computes the conditional expectation of C_{ik} given \mathbf{z} , where C_{ik} is an element of \mathbf{C} . Specifically, on the $(m + 1)$ th iteration, the E-step computes

$$\begin{aligned} E_{\mathbf{\Omega}^{(m)}}(C_{ik} | \mathbf{z}) &= pr_{\mathbf{\Omega}^{(m)}} \{C_{ik} = 1 | \mathbf{z}\} \\ &= \boldsymbol{\tau}_k(\mathbf{z}_i; \mathbf{\Omega}^{(m)}), \end{aligned} \quad [9.11]$$

where $\boldsymbol{\tau}_k(\mathbf{z}_i; \mathbf{\Omega}^{(m)})$ is the posterior probability of class membership defined in Equation [9.4]. The M-step of the EM algorithm maximizes $Q(\mathbf{\Omega}, \mathbf{\Omega}^{(m)})$ with respect to $\mathbf{\Omega}$ providing the updated estimate $\mathbf{\Omega}^{(m+1)}$. Note that the E-step replaces c_{ik} in Equation [9.9] with $\boldsymbol{\tau}_k(\mathbf{z}_i; \mathbf{\Omega}^{(m)})$. Therefore, the updated estimate of the mixing proportion for class k , denoted as $\pi_i^{(m+1)}$ is

$$\pi_k^{(m+1)} = \sum_{i=1}^n \tau_k(\mathbf{z}_i; \boldsymbol{\Omega}^{(m)}) / n \quad (k = 1, 2, \dots, K). \quad [9.12]$$

9.3 Cross-Sectional Models for Categorical Latent Variables

In this section, we discuss models for categorical latent variables, with applications to cross-sectional and longitudinal designs. This section is drawn from Kaplan (in press). To motivate the use of categorical latent variables consider the problem of measuring reading ability in young children. Typical studies of reading ability measure reading on a continuous scale. Using the methods of item response theory (see, e.g., Hambleton & Swaminathan, 1985), reading measures are administered to survey participants on multiple occasions, with scores equated in such a way as to allow for a meaningful notion of growth. However, in large longitudinal studies such as the Early Childhood Longitudinal Study (NCES, 2001), not only are continuous scale scores of total reading proficiency available for analyses but also mastery scores for subskills of reading. For example, a fundamental subskill of reading is letter recognition. A number of items constituting a cluster that measures letter recognition are administered, and, according to the ECLS-K scoring protocol, if the child receives 3 out of 4 items in the cluster correct, then the child is assumed to have mastered the skill with mastery coded “1” and nonmastery coded as “0.” Of course, there exist other, more difficult, subskills of reading, including beginning sounds, ending sounds, sight words, and words in context with subskill cluster coded for mastery.

Assume for now that these subskills tap a general reading ability factor. In the context of factor analysis, a single continuous factor can be derived that would allow children to be placed somewhere along the factor. Another approach might be to derive a factor that serves to categorize children into mutually exclusive classes on the latent reading ability factor. Latent class analysis is designed to accomplish this categorization.

9.3.1 LATENT CLASS ANALYSIS

Latent class models were introduced by Lazarsfeld and Henry (1968) for the purposes of deriving latent attitude variables from responses to dichotomous survey items. In a traditional latent class analysis, it is assumed that an individual belongs to one and only one latent class, and that given the individual’s latent class membership, the observed variables are independent of one another—the so-called *local independence assumption* (see Clogg, 1995). The

latent classes are, in essence, categorical factors arising from the pattern of response frequencies to categorical items, where the response frequencies play a role similar to that of the correlation matrix in factor analysis (Collins, Hyatt, and Graham, 2000). The analogues of factor loadings are probabilities associated with responses to the manifest indicators given membership in the latent class. Unlike continuous latent variables, categorical latent variables serve to partition the population into discrete groups based on response patterns derived from manifest categorical variables.

9.3.2 SPECIFICATION, IDENTIFICATION, AND TESTING OF LATENT CLASS MODELS

The latent class model can be written as follows. Let

$$P_{ijkl} = \sum_{a=1}^A \delta_a \rho_{i|a} \rho_{j|a} \rho_{k|a} \rho_{l|a}, \quad [9.13]$$

where δ_a is the proportion of individuals in latent class a . The parameters $\rho_{i|a}$, $\rho_{j|a}$, $\rho_{k|a}$, and $\rho_{l|a}$ are the response probabilities for items i , j , k , and l , respectively conditional on membership in latent class a .

In the case of the ECLS-K reading example, there are five dichotomously scored reading subskill measures, which we will refer to here as A , B , C , D , and E . Denote the response options for each of the measures respectively by i , j , k , l , and m , ($i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$; $l = 1, \dots, L$; $m = 1, \dots, M$) and denote the categorical latent variable as ξ . Then, the latent class model can be written as

$$\pi_{ijklmc}^{ABCDE\xi} = \pi_c^\xi \pi_{ic}^{A|\xi} \pi_{jc}^{B|\xi} \pi_{kc}^{C|\xi} \pi_{lc}^{D|\xi} \pi_{mc}^{E|\xi}, \quad [9.14]$$

where π_c^ξ is the probability that a randomly selected child will belong to latent class c ($c = 1, 2, \dots, C$) of the categorical latent variable ξ , $\pi_{ic}^{A|\xi}$ is the conditional probability of response i to variable A given membership in latent class c , and $\pi_{jc}^{B|\xi}$, $\pi_{kc}^{C|\xi}$, $\pi_{lc}^{D|\xi}$, and $\pi_{mc}^{E|\xi}$ are likewise the conditional probabilities for items B , C , D , and E , respectively. For this example, the manifest variables are dichotomously scored, and so there are two response options for each item.²

Identification of a latent class model is typically achieved by imposing the constraint that the latent classes and the response probabilities that serve as indicators of the latent classes sum to 1.0—namely, that

$$\sum_c \pi_c^\xi = \sum_i \pi_{ic}^{A|\xi} = \sum_j \pi_{jc}^{B|\xi} = \sum_k \pi_{kc}^{C|\xi} = \sum_l \pi_{lc}^{D|\xi} = \sum_m \pi_{mc}^{E|\xi} = 1.0, \quad [9.15]$$

where the first term on the left-hand side of Equation [9.15] indicates that the latent class proportions must sum to 1.0, and the remaining terms on the left-hand side of Equation [9.15] denote that the latent class indicator variables sum to 1.0 as well (McCutcheon, 2002).³

To continue with our reading example, suppose that we hypothesize that the latent class variable ξ is a measure of reading ability with three classes (1 = *advanced reading ability*, 2 = *average reading ability*, and 3 = *beginning reading ability*). Assume also that we have a random sample of first semester kindergartners. Then, we might find that a large proportion of kindergartners in the sample who show mastery of letter recognition (items *A* and *B*, both coded 1/0) are located in the beginning reading ability class. A smaller proportion of kindergartners demonstrating mastery of ending sounds and sight words might be located in the average reading ability class, and still fewer might be located in the advanced reading class. Of course at the end of kindergarten and hopefully by the end of first grade, we would expect to see the relative proportions shift.⁴

An Example of Latent Class Analysis

The following example comes from Kaplan and Walpole (2005) using data from the *Early Childhood Longitudinal Study: Kindergarten Class of 1998–1999* (NCES, 2001). The ECLS-K database provides a unique opportunity to estimate the prospects of successful reading achievement (which Kaplan and Walpole define as the ability to comprehend text) by the end of first grade for children with different levels of entering skill and different potential barriers to success. The ECLS-K data available for their example include longitudinal measures of literacy achievement for a large and nationally representative sample—a sample unprecedented in previous investigations of early reading development.

Data used in the Kaplan and Walpole (2005) example consist of the kindergarten base year (Fall 1998/Spring 1999) and first grade follow-up (Fall 1999/Spring 2000) panels of ECLS-K. Only first-time public school kindergarten students who were promoted to and present at the end of first grade were chosen for this study. The sample size for their example was 3,575.⁵

The measures used in their example consisted of a series of reading assessments. Using an item response theory framework, the reading assessment yielded scale scores for (1) letter recognition, (2) beginning sounds, (3) ending sounds, (4) sight words, and (5) words in context.

In addition to reading scale scores, ECLS-K provides transformations of these scores into probabilities of proficiency as well as dichotomous proficiency scores, the latter which Kaplan and Walpole used in their study. The reading proficiencies were assumed to follow a Guttman simplex model, where mastery at a specific skill level implies mastery at all previous skill levels. Details regarding the construction of these proficiency scores can be found in Kaplan and Walpole (2005).

188 STRUCTURAL EQUATION MODELING

Table 9.1 presents the response probabilities measuring the latent classes for each wave of the study separately. The interpretation of this table is similar to the interpretation of a factor loading matrix. The pattern of response probabilities across the subsets of reading tests suggest the labels that have been given to the latent classes—namely, low alphabet knowledge (LAK), early word reading (EWR), and early word comprehension (ERC). The extreme differences across time in the likelihood ratio chi-square tests are indicative of sparse cells, particularly occurring at spring kindergarten. For the purposes of this chapter, I proceed with the analysis without attempting to ameliorate the problem.

Table 9.1 Response Probabilities and Class Proportions for Separate Latent Class Models: Total Sample

<i>Latent Class</i>	<i>Subtest Response Probabilities^a</i>					<i>Class Proportions</i>	χ^2_{LR} (29 df)
	<i>LR^b</i>	<i>BS</i>	<i>ES</i>	<i>SW</i>	<i>WIC</i>		
Fall K							
LAK ^c	0.47	0.02	0.01	0.00	0.00	0.67	3.41
EWR	0.97	0.87	0.47	0.02	0.00	0.30	
ERC	1.00	0.99	0.98	0.97	0.45	0.03	
Spring K							
LAK	0.56	0.06	0.00	0.00	0.00	0.24	4831.89*
EWR	0.99	0.92	0.63	0.05	0.00	0.62	
ERC	0.00	0.99	0.99	0.96	0.38	0.14	
Fall First							
LAK	0.52	0.08	0.01	0.00	0.00	0.15	11.94
EWR	1.00	0.92	0.71	0.05	0.03	0.59	
ERC	1.00	0.99	0.98	0.98	0.42	0.26	
Spring First							
LAK	0.19	0.00	0.00	0.00	0.00	0.04	78.60*
EWR	0.98	0.90	0.79	0.35	0.00	0.18	
ERC	1.00	0.99	0.98	0.99	0.60	0.78	

a. Response probabilities are for passed items. Response probabilities for failed items can be computed from $1 - prob$ (mastery).

b. LR = letter recognition, BS = beginning sounds, ES = ending letter sounds, SW = sight words, WIC = words in context.

c. LAK = low alphabet knowledge, EWR = early word reading, ERC = early reading comprehension.

* $p < .05$. Extreme value likely due to sparse cells.

The last column of Table 9.1 presents the latent class membership proportions across the four ECLS-K waves for the full sample. We see that in fall of kindergarten, approximately 67% of the cases fall into the LAK class, whereas only approximately 3% of the cases fall into the ERC class. This breakdown of proportions can be compared with the results for Spring of first grade; by that time, only 4% of the sample are in the LAK class, whereas approximately 78% of the sample is in the ERC class.

9.4 Longitudinal Models for Categorical Latent Variables: Markov Chain Models

The example of latent class analysis given in the previous sections presented results over the waves of the ECLS-K but treated each wave cross-sectionally. Nevertheless, it could be seen from Table 9.1 that response probabilities did change over time as did latent class membership proportions. Noting these changes, it is important to have a precise approach to characterizing change in latent class membership over time. In this section, we consider changes in latent class membership over time. We begin by describing a general approach to the study of change in qualitative status over time via Markov chain modeling, extended to the case of latent variables. This is followed by a discussion of latent transition analysis, a methodology well-suited for the study of stage-sequential development.

9.4.1 IDENTIFICATION, ESTIMATION, AND TESTING OF MARKOV CHAIN MODELS

In this section, we briefly discuss the problem of parameter identification, estimation, and model testing in Markov chain models. As with the problem of identification in factor analysis and structural equation models, identification in Markov chain models is achieved by placing restrictions on model.

With regard to manifest Markov chains, identification is not an issue. All parameters can be obtained directly from manifest categorical responses. In the context of latent Markov chain models with a single indicator, the situation is somewhat more difficult. Specifically, identification is achieved by restricting the response probabilities to be invariant over time. As noted by Langeheine & Van de Pol (2002), this restriction simply means that measurement error is assumed to be equal over time. For four or more time points, it is only required that the first and last set of response frequencies be invariant. As with latent class analysis, parameters are estimated via ML using the EM algorithm as discussed in Sections 9.1 and 9.2.

190 STRUCTURAL EQUATION MODELING

After obtaining estimates of model parameters, the next step is to assess whether the specified model fits the data. In the context of Markov chain models and latent class extensions, model fit is assessed by comparing the observed response proportions against the response proportions predicted by the model. Two statistical tests are available for assessing the fit of the model based on comparing observed versus predicted response proportions. The first is the classic Pearson chi-square statistic. As an example from the latent class framework, the Pearson chi-square test can be written as

$$\chi^2 = \sum_{ijkl} \frac{(F_{ijkl} - f_{ijkl})^2}{f_{ijkl}}, \quad [9.16]$$

where F_{ijkl} are the observed frequencies of the $IJKL$ contingency table and f_{ijkl} are the expected cell counts. The degrees of freedom are obtained by subtracting the number of parameters to be estimated from the total number of cells of the contingency table that are free to vary.

In addition to the Pearson chi-square test, a likelihood ratio statistic can be obtained that is asymptotically distributed as chi-square, where the degrees of freedom are calculated as with the Pearson chi-square test. Finally, the Akaike information criterion (AIC) and Bayesian information criterion (BIC) discussed in Chapter 6 can be used to choose among competing models.

9.4.2 THE MANIFEST MARKOV MODEL

The manifest Markov model consists of a single chain, where predicting the current state of an individual only requires data from the previous occasion. In line with the example given in Section 4, consider measuring mastery of ending letter sounds at four discrete time points. The manifest Markov model can be written as

$$P_{ijkl} = \delta_i^1 \tau_{ji}^{21} \tau_{kj}^{32} \tau_{lk}^{43}, \quad [9.17]$$

where P_{ijkl} is the model-based expected proportion of respondents in the defined population in cell (i, j, k, l) . The subscripts, i, j, k , and l are the manifest categories for times 1, 2, 3, and 4, respectively, with $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$; and $l = 1, \dots, L$. In this study, there are two categorical responses for i, j, k , and l —namely, mastery or nonmastery of ending letter sounds. Thus, $I = J = K = L = 2$. The parameter δ_i^1 is the observed proportion of individuals at time 1 who have or have not mastered ending letter sounds and corresponds to the initial marginal distribution of the outcome. The parameters τ_{ji}^{21} , τ_{kj}^{32} , and τ_{lk}^{43} are the transition probabilities. Specifically, the parameter τ_{ji}^{21} represents the

transition probability from time 1 to time 2 for those in category j given they were in category i at the beginning of the study. The parameter τ_{klj}^{32} represents the transition probability from time 2 to time 3 for those in category k given they were in category j at the previous time point. Finally, the parameter τ_{lk}^{43} is the transition probability from time 3 to time 4 for those in category l given that they were in category k at the previous time point.

The manifest Markov model can be specified to allow transition probabilities to be constant over time or to allow transition probabilities to differ over time. The former is referred to as a stationary Markov chain while the latter is referred to as a nonstationary Markov chain.

Application of the Manifest Markov Model

Table 9.2 presents the results of the nonstationary manifest Markov model applied to the development of competency in ending sounds.⁶ It can be seen that over time, the probabilities associated with moving from nonmastery of ending sounds to master of ending sounds changes. For example, at the beginning of kindergarten and the beginning of first grade, the proportions who have not mastered beginning sounds and the proportion who then go on to master ending sounds is relatively constant. However, the transition from nonmastery of ending sounds to mastery of ending sounds is much greater from the beginning of first grade to the end of first grade. Nevertheless, approximately 25% of the sample who did not master ending sounds at the beginning of first grade does not appear to have mastered ending sounds by the end of first grade.

9.4.3 THE LATENT MARKOV MODEL

A disadvantage of the manifest Markov model is that it assumes that the manifest categories are perfectly reliable measures of a true latent state. In the context of the ending sounds example, this would imply that the observed categorical responses measure the true mastery/nonmastery of ending sounds. Rather, it may be more reasonable to assume that the observed responses are fallible measures of an unobservable latent state, and it is the study of transitions across true latent states that are of interest.

The latent Markov model was developed by Wiggins (1973) to address the problem of measurement error in observed categorical responses and as a result, to obtain transition probabilities at the latent level. The latent Markov model can be written as

$$P_{ijkl} = \sum_{a=1}^A \sum_{b=1}^B \sum_{c=1}^C \sum_{d=1}^D \delta_a^1 \rho_{i|a}^1 \tau_{b|a}^{21} \rho_{j|b}^2 \tau_{c|b}^{32} \rho_{k|c}^3 \tau_{d|c}^{43} \rho_{l|d}^4, \quad [9.18]$$

192 STRUCTURAL EQUATION MODELING

Table 9.2 Results of the Nonstationary Manifest Markov Chain Model Applied to Mastery of Ending Sounds

<i>Ending Sounds Time 1 (Rows) by Ending Sounds Time 2 (Columns)^a</i>		
	1	2
1	0.55	0.45
2	0.10	0.90
<i>Ending Sounds Time 2 (Rows) by Ending Sounds Time 3 (Columns)</i>		
	1	2
1	0.57	0.43
2	0.10	0.90
<i>Ending Sounds Time 3 (Rows) by Ending Sounds Time 4 (Columns)</i>		
	1	2
1	0.25	0.75
2	0.03	0.97
Goodness-of-fit tests^b		
χ_p^2 (8 df) = 133.77, $p < .05$		
χ_{LR}^2 (8 df) = 150.23, $p < .05$		
BIC = 13363.49		

a. 1 = nonmastery, 2 = mastery.

b. χ_p^2 refers to the Pearson chi-square test, χ_{LR}^2 refers to the likelihood ratio chi-square test.

where the parameters in Equation [9.18] taken on slightly different meanings from those in Equation [9.17]. In particular, the parameter δ_a^1 represents a latent distribution having A latent states. The linkage of the latent states to manifest responses is accomplished by the response probabilities ρ . The response probabilities thus serve a role analogous to that of factor loadings in factor analysis. Accordingly, ρ_{ij}^1 refers to the response probability associated with category i given membership in latent state a . The parameter ρ_{jb}^2 is interpreted as the response probability associated with category j given membership in latent state b at time 2. Remaining response probabilities are similarly interpreted.

As with the manifest Markov model, the transition from time 1 to time 2 in latent state membership is captured by τ_{ji}^1 . At time 2, the latent state is measured by the response probabilities ρ_{jb}^2 . Remaining response and transition probabilities are analogously interpreted. Note that an examination of Equation [9.18] reveals that if the response probabilities were all 1.0 (indicating perfect measurement of the latent variable), then Equation [9.18] would essentially reduce to Equation [9.17]—the manifest Markov model.

Application of the Latent Markov Model

Table 9.3 compares the transition probabilities for the manifest Markov model and the latent Markov model under the assumption of a stationary Markov chain. The results show small but noticeable differences in the transition probabilities when taking account measurement error in the manifest categorical responses.

9.4.4 LATENT TRANSITION ANALYSIS

Although the application of Markov models for the analysis of psychological variables goes back to Anderson (1959; as cited in Collins & Wugalter, 1992), most applications focused on single manifest measures. However, as with the early work in the factor analysis of intelligence tests (e.g., Spearman, 1904), it was recognized that many important psychological variables are

Table 9.3 Comparison of Transition Probabilities for Manifest and Latent Markov Chain Model With Homogenous Transition Probabilities: Application to Ending Letter Sounds

<i>Manifest Markov Chain</i>			<i>Latent Markov Chain</i>		
<i>Ending Sounds Time 1 (Rows) by Ending Sounds Time 2 (Columns)^a</i>					
	1	2		1	2
1	0.50	0.50	1	0.47	0.53
2	0.38	0.62	2	0.38	0.62
<i>Ending Sounds Time 2 (Rows) by Ending Sounds Time 3 (Columns)</i>					
	1	2		1	2
1	0.50	0.50	1	0.47	0.53
2	0.38	0.62	2	0.38	0.62
<i>Ending Sounds Time 3 (Rows) by Ending Sounds Time 4 (Columns)</i>					
	1	2		1	2
1	0.50	0.50	1	0.47	0.53
2	0.38	0.62	2	0.38	0.62
Goodness-of-fit tests					
$\chi^2_p (13 df) = 6946.62, p < .05$			7040.50, $p < .05$		
$\chi^2_{LR} (13 df) = 6169.320, p < .05$			6299.62, $p < .05$		
BIC = 19341.68			19471.99, $p < .05$		

a. 1 = nonmastery, 2 = mastery.

latent—in the sense of not being directly observed but possibly measured by numerous manifest indicators. The advantages to measuring multiple latent variables via multiple indicators are the known benefits with regard to reliability and validity. Therefore, it might be more realistic to specify multiple manifest categorical indicators of the categorical latent variable and combine them with Markov chain models.

The combination of multiple indicator latent class models and Markov chain models provides the foundation for the latent transition analysis of *stage-sequential dynamic latent variables*. In line with Collins and Flaherty (2002), consider the current reading example where the data provide information on the mastery of five different skills. At any given point in time, a child has mastered or not mastered one or more of these skills. It is reasonable in this example to postulate a model that specifies that these reading skills are related in such a way that mastery of a later skill implies mastery of all preceding skills. At each time point, the child's latent class membership defines his or her *latent status*. The model specifies a particular type of change over time in latent status. This is defined by Collins and Flaherty (2002) as a "model of stage-sequential development, and the skill acquisition process is a stage-sequential dynamic latent variable" (p. 289). It is important to point out that there is no fundamental difference between latent transition analysis and latent Markov chain modeling. The difference is practical, with latent transition analysis being perhaps better suited conceptually for the study of change in developmental status.

The model form for latent transition analysis uses Equation [9.18] except that model estimation is undertaken with multiple indicators of the latent categorical variable. The appropriate measurement model for categorical latent variables is the *latent class model*.

Application of Latent Transition Analysis

Using all five of the subtests of the reading assessment in ECLS-K, this section demonstrates a latent transition analysis. It should be noted that a specific form of the latent transition model was estimated—namely, a model that assumes no forgetting or loss of previous skills. This type of model is referred to as a longitudinal Guttman process and was used in a detailed study of stage sequential reading development by Kaplan and Walpole (2005).

A close inspection of the changes over time in class proportions shown in Table 9.1 points to transition over time in the proportions who master more advanced reading skills. However, these separate latent class models do not provide simultaneous estimation of the transition probabilities, which are crucial for a study of stage-sequential development over time.

In Table 9.4, the results of the latent transition probabilities for the full latent transition model are provided. On the basis of the latent transition analysis, we see that for those in the LAK class at Fall kindergarten, 30% are predicted to remain in the LAK class, while 69% are predicted to move to the EWR class and 1% are predicted to transition to ERC in Spring kindergarten. Among those in the EWR class at Fall kindergarten, 66% are predicted to remain in that class, and 34% of the children are predicted to transition to the ERC class in Spring kindergarten.

Among those children who are in the LAK class at Spring Kindergarten, 59% are predicted to remain in that class at Fall of first grade, while 40% are predicted to transition to the EWR class, with 1% predicted to transition to the ERC class. Among those children who are in the EWR class in Fall kindergarten, 82% are predicted to stay in the EWR class while 18% are predicted to transition to the ERC class.

Finally, among those children who are in the LAK class in Fall of first grade, 30% are predicted to remain in that class at Spring of first grade, while 48% are predicted to transition to the EWR class by Spring of first grade, with 22%

Table 9.4 Transition Probabilities From Fall Kindergarten to Spring First Grade

<i>Wave</i>	<i>LAK^a</i>	<i>EWR</i>	<i>ERC</i>
<i>Fall K</i>		<i>Spring K</i>	
LAK	0.30	0.69	0.01
EWR	0.00	0.66	0.34
ERC	0.00	0.00	1.00
<i>Spring K</i>		<i>Fall First</i>	
LAK	0.59	0.40	0.01
EWR	0.00	0.82	0.18
ERC	0.00	0.00	1.00
<i>Fall First</i>		<i>Spring First</i>	
LAK	0.30	0.48	0.22
EWR	0.01	0.13	0.86
ERC	0.00	0.00	1.00

Goodness-of-fit tests

$$\chi_p^2 (1048528 \text{ df}) = 12384.21, p = 1.0$$

$$\chi_{LR}^2 (1048528 \text{ df}) = 6732.31, p = 1.0$$

$$\text{BIC} = 44590.80$$

a. LAK = low alphabet knowledge, EWR = early word reading, ERC = early reading comprehension.

transitioning to the ERC class. Among those children in the EWR class at fall of first grade, 13% are assumed to remain in that class with 86% transitioning to the ERC class by Spring of first grade.

9.4.5 MIXTURE LATENT MARKOV MODEL (THE MOVER-STAYER MODEL)

A limitation of the models described so far is that they assume that the sample of observations arises from a single population that can be characterized by a single Markov chain (latent or otherwise) and one set of parameters—albeit perhaps different for certain manifest groups such as those children living above or below poverty. It is possible, however, that the population is composed of a finite and unobserved mixture of subpopulations characterized by qualitatively different Markov chains. To the extent that the population consists of finite mixtures of subpopulations, then a “one-size-fits-all” application of the Markov model can lead to biased estimates of the parameters of the model as well as incorrect substantive conclusions regarding the nature of the developmental process in question. A reasonable strategy for addressing this problem involves combining Markov chain–based models under the assumption of a mixture distribution (see, e.g., McLachlan & Peel, 2000 for an excellent overview of finite mixture modeling). This is referred to as the *mixture latent Markov model*.⁷

An important special case of the mixture latent Markov model is referred to as the *mover-stayer model* (Blumen, Kogan, & McCarthy, 1955). In the mover-stayer model, there exists a latent class of individuals who transition across stages over time (movers) and a latent class that does not transition across stages (stayers). In the context of reading development, the stayers are those who never move beyond, say, mastery of letter recognition. Variants of the mover-stayer models have been considered by Van de Pol and Langeheine (1989; see also Mooijaart, 1998).

The mixture latent Markov model can be written as

$$P_{ijkl} = \sum_{s=1}^S \sum_{a=1}^A \sum_{b=1}^B \sum_{c=1}^C \sum_{d=1}^D \pi_s \delta_{a|s}^1 \rho_{i|as}^1 \tau_{b|as}^{21} \rho_{j|bs}^2 \tau_{c|jbs}^{32} \rho_{k|cs}^3 \tau_{l|ks}^{43} \rho_{l|ds}^4 \quad [9.19]$$

where π_s represents the proportion of observations in Markov chain s ($= 1, 2, \dots, S$), and the remaining parameters are interpreted as in Equation [9.18], with the exception that they are conditioned on membership in Markov chain s .

The model in Equation [9.19] is the most general of those considered in this article with the preceding models being derived as special cases. For example, with $s = 1$, Equation [9.19] reduces to the latent Markov model in Equation [9.18]. Also, with $s = 1$, and no transition probabilities, the model in Equation [9.19] reduces to the latent class model of Equation [9.13].

Application of the Mover-Stayer Model

For this example, we estimate the full latent transition analysis model with the addition of a latent class variable that is hypothesized to segment the sample into those who do transition over time in their development of more complex reading skills (movers) versus those that do not transition at all (stayers). The results of the mover-stayer latent transition analysis are given in Table 9.5. In this analysis, it is assumed that the stayer class has zero probability of moving. An alternative specification can allow the “stayers” to have a probability that is not necessarily zero but different from the mover class.

From the upper panel of Table 9.5, it can be seen that 97% of the sample transition across stages, with 71% of the movers beginning their transitions to full literacy from the LAK status, 26% beginning EWR status, and 2% already in the ERC status. The stayers represent only 3% of the sample, corresponding to approximately 90 children. These children are in the low alphabet knowledge class and are not predicted to move.

The lower panel of Table 9.5 gives the transition probabilities for the whole sample. In many cases, it would be necessary to compute the transition probabilities separately for the movers, but because all the stayers are in the LAK class, they do not contribute to the transition probabilities for the movers. The slight differences between the mover transition probabilities compared with the transition probabilities in Table 9.4 are due to the fact that 3% of the sample is in the stayer class. Finally, it may be interesting to note that based on a comparison of the BICs the results of the mover-stayer specification provides a better fit to the manifest response frequencies than the latent transition analysis model in Table 9.4. However, the discrepancy between the likelihood ratio chi-square and Pearson chi-square is, again, indicative of sparse cells and would need to be inspected closely.

9.5 Models for Categorical and Continuous Latent Variables

Having introduced the topic of categorical latent variables, we can now move to models that combine categorical and continuous latent variables. The basic idea here, as before, is that a population might be composed of finite mixtures of subpopulations characterized by their own unique parameters, but where the parameters are those of models based on continuous latent variables—such as factor analysis and structural equation models. For this section, we focus on finite mixture modeling applied to growth curve modeling because growth curve modeling encompasses many special cases, including factor analysis, structural equation modeling, and MIMIC modeling.

198 STRUCTURAL EQUATION MODELING

Table 9.5 Transition Probabilities for the Mover-Stayer Model: Total Sample

<i>Movers and Stayers (Rows) by Time 1 Classes (Columns)</i>				<i>Proportion of Total Sample</i>
	LAK	EWR	ERC	
Movers	0.71	0.26	0.02	0.97
Stayers	1.00	0.00	0.00	0.03

Results for Movers*Fall K Classes (Rows) by Spring K Classes (Columns)*

	LAK	EWR	ERC
LAK	0.34	0.65	0.01
EWR	0.00	0.62	0.38
ERC	0.00	0.00	1.00

Spring K Classes (Rows) by Fall First Classes (Columns)

	LAK	EWR	ERC
LAK	0.61	0.39	0.00
EWR	0.00	0.84	0.16
ERC	0.00	0.00	1.00

Fall First Classes (Rows) by Spring First Classes (Columns)

	LAK	EWR	ERC
LAK	0.22	0.55	0.23
EWR	0.01	0.12	0.87
ERC	0.00	0.00	1.00

Goodness-of-fit tests

$$\chi_p^2 (1048517 \text{ df}) = 10004.46, p = 1.0$$

$$\chi_{LR}^2 (1048517 \text{ df}) = 5522.87, p = 1.0$$

$$\text{BIC} = 43397.29$$

a. LAK = low alphabet knowledge, EWR = early word reading, ERC = early reading comprehension.

9.5.1 GENERAL GROWTH MIXTURE MODELING

Conventional growth curve modeling and its extensions were discussed in Chapter 8. The power of conventional growth curve modeling notwithstanding, a fundamental constraint of the method is that it assumes that the manifest growth trajectories are a sample from a single finite population of trajectories characterized by a single average level parameter and a single average growth rate. However, it may be the case that the sample is derived from a mixture of populations, each having its own unique growth trajectory. For

example, children may be sampled from populations exhibiting very different classes of math development—some children may have very rapid rates of growth in math that level off quickly, others may show normative rates of growth, while still others may show very slow or problematic rates of growth.

An inspection of Figure 9.1 reveals heterogeneity in the shapes of the growth curves for a sample of 100 children who participated in the Early Childhood Longitudinal Study. If such distinct growth functions are actually present in the data, then conventional growth curve modeling applied to a mixture of populations will ignore this heterogeneity in growth functions and result in biased estimates of growth. Therefore, it may be preferable to relax the assumption of a single population of growth and allow for the possibility that the population is composed of mixtures of distinct growth trajectory shapes.

Growth mixture modeling begins by unifying conventional growth curve modeling with latent class analysis (e.g., Clogg, 1995) under the assumption that there exists a finite mixture of populations defined by unique trajectory classes. An extension of latent class analysis sets the foundation for growth mixture modeling. Specifically, latent class analysis can be applied to repeated measures at different time points. This is referred to as latent class growth analysis (see, e.g., B. Muthén, 2001; Nagin, 1999). As with latent class analysis, latent class growth analysis assumes homogenous growth within classes. Growth mixture modeling relaxes the assumption of homogeneous growth within classes and is capable of capturing two significant forms of heterogeneity. The first form of heterogeneity is captured by individual differences in growth through the specification of the conventional growth curve model. The second form of heterogeneity is more basic—representing heterogeneity in classes of growth trajectories.

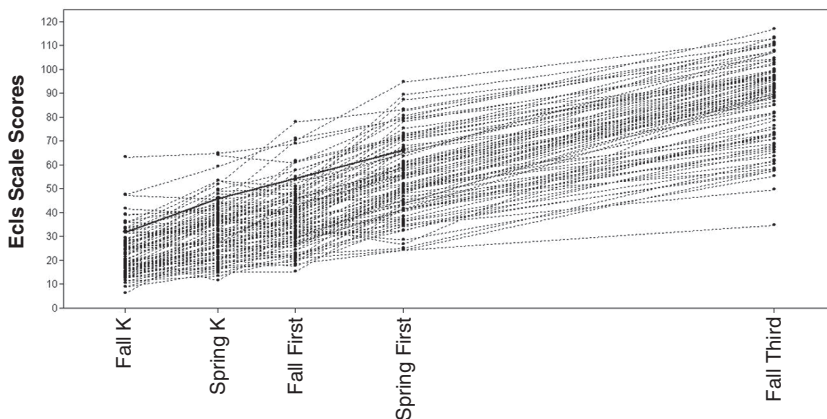


Figure 9.1 Sample of 100 Empirical Growth Trajectories

9.5.2 SPECIFICATION OF THE GROWTH MIXTURE MODEL

The growth mixture model is similar to that given for the conventional growth curve model. The difference lies in allowing there to be different growth trajectories for different classes. Thus, in line with Equations 8.5 and 8.6, we can represent the presence of trajectory classes as

$$y_i = \mathbf{v} + \mathbf{A}\boldsymbol{\eta}_i + \mathbf{K}\mathbf{x}_i + \boldsymbol{\varepsilon}_i \quad [9.20]$$

and

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_c + \mathbf{B}_c\boldsymbol{\eta}_i + \boldsymbol{\Gamma}_c\mathbf{x}_i + \boldsymbol{\zeta}_i, \quad [9.21]$$

where the subscript c represents trajectory class ($c = 1, 2, \dots, C$).

The advantage to using growth mixture modeling lies in the ability to characterize across-class differences in the shape of the growth trajectories. Assuming that the time scores are constant across the classes, the different reading trajectory shapes are captured in $\boldsymbol{\alpha}_c$. Relationships among growth parameters contained in \mathbf{B}_c are also allowed to be class-specific. The modeling framework is flexible enough to allow differences in measurement error variances ($\boldsymbol{\Theta}$) and structural disturbance variances ($\boldsymbol{\Psi} = \text{Var}(\boldsymbol{\zeta})$) across classes as well. Finally, of relevance to this chapter, the different classes can show different relationships to a set of covariates \mathbf{x} . In the context of our example, Equation [9.21] allows one to test whether poverty level has a differential effect on growth depending on the shape of the growth trajectories. Again, one might hypothesize that there is a small difference between poverty levels for children with normative or above average rates of growth in math, but that poverty has a strong positive effect for those children who show below normal rates of growth in math.

Application of Growth Mixture Modeling

The results of the conventional growth curve modeling provide initial information for assessing whether there are substantively meaningful growth mixture classes. To begin, the conventional growth curve model can be considered a growth mixture model with only one mixture class. From here, we specified two, three, and four mixture classes. We used three criteria to judge the number of classes that we decided to retain. The first criterion was the proportion of ECLS-K children assigned to the mixture classes. The second criterion was BIC, which was used to assess whether the extraction of additional latent classes improved the fit of the model. The third criterion was the adequacy of classification using the average posterior probabilities of classification. On the basis of these three criteria, and noting that the specification of

the model did not include the covariate of poverty level, we settled on retaining three growth mixture classes. A plot of the three classes can be found in Figure 9.2.

From Table 9.6 and Figure 9.2, we label the first latent class, consisting of 35.5% of our samples, as “below average developers.” Students in this class evidenced a spring kindergarten mean math achievement score of 23.201, a linear growth rate of 1.317, and a de-acceleration in growth of .005. We labeled the second latent class, comprising of 58.3% of our sample, as “average developers.” Students in this class evidenced a spring kindergarten mean math achievement score of 33.646, a linear growth rate of 1.890, and a de-acceleration of .006. Finally, we labeled the third latent class, consisting of 35.5% of our sample as, “above average developers.” Students in this class evidenced a spring kindergarten mean math achievement score of 54.308, a linear growth rate of 1.988, and a de-acceleration of $-.016$.

When poverty level was added into the growth mixture model, three latent classes were again identified.⁸ The above average developer class started significantly above their peers and continued to grow at a rate higher than the rest of their peers. Interestingly, the above average achiever group was composed entirely of students living above the poverty line. The average achiever group

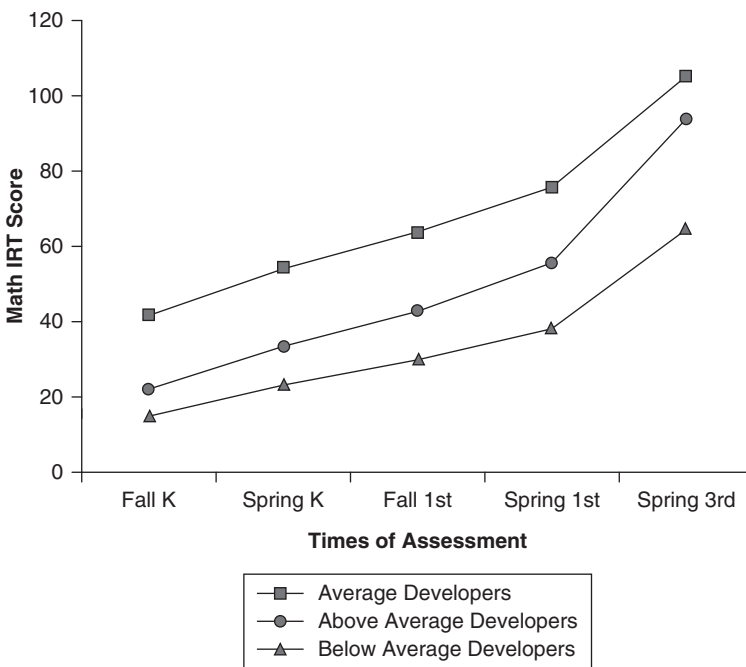


Figure 9.2 The Three-Class Growth Mixture Model

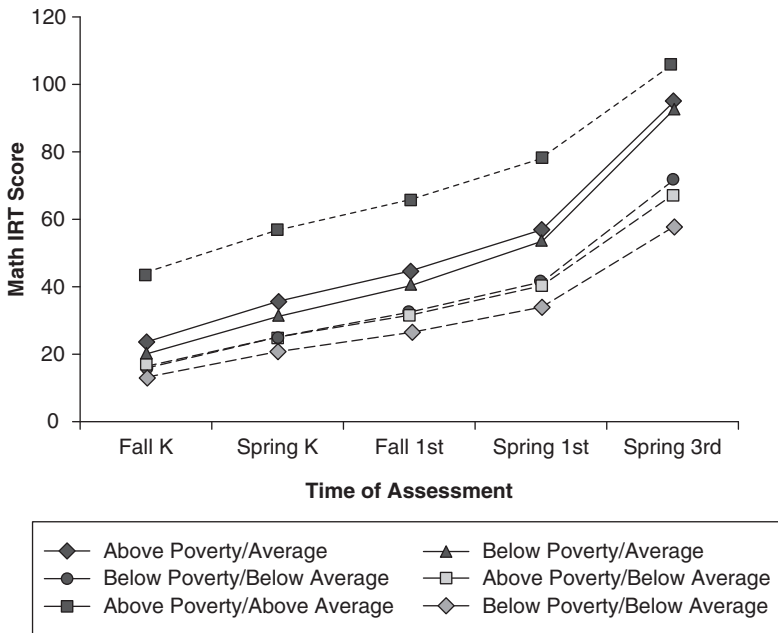
202 STRUCTURAL EQUATION MODELING

Table 9.6 Results of Three-Class Growth Mixture Model

Coefficient	Class 1		Class 2		Class 3	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Intercept (I)	23.201	24.968	33.646	34.943	54.308	56.081
Linear slope (S)	1.317	1.365	1.890	1.912	1.988	1.989
Quadratic (Q)	-0.005	-0.006	-0.006	-0.007	-0.016	-0.017
I on below poverty		-4.513		-10.418		-24.376
S on below poverty		-0.194		-0.434		-0.129
Q on below poverty		0.002		0.002*		0.012

*Not statistically significant.

was composed of both students who lived above and below the poverty line. The below average achiever group was composed disproportionately of below poverty students but did contain some above poverty students. A plot of the three-class solution with poverty added to the model can be found in Figure 9.3.

**Figure 9.3** The Three-Class Growth Mixture Model With Poverty Status Added

The posterior probabilities of classification without and with poverty added to the model can be found in Tables 9.7 and 9.8, respectively. We observe that students who should be classified above average achievers had a .882 probability of being correctly classified as below average developers. Students in the average developer class had a .855 probability of being correctly classified by the model as average achievers. Finally, students in the above average class had a .861 probability of being correctly classified by the model into the below average class. The posterior probabilities do not change dramatically with the addition of poverty to the model, as seen in Table 9.8.

Table 9.7 Average Posterior Probabilities for the Three-Class Solution for Baseline Model

	<i>Class 1</i>	<i>Class 2</i>	<i>Class 3</i>
Class 1	0.882	0.027	0.090
Class 2	0.138	0.855	0.007
Class 3	0.138	0.001	0.861

NOTE: Class 1 = average developing; Class 2 = above average; Class 3 = below average.

Table 9.8 Average Posterior Probabilities for the Three-Class Solution With Poverty Status Included

	<i>Class 1</i>	<i>Class 2</i>	<i>Class 3</i>
Class 1	0.858	0.041	0.101
Class 2	0.155	0.826	0.019
Class 3	0.191	0.008	0.801

NOTE: Class 1 = average developing; Class 2 = above average; Class 3 = below average.

9.6 Conclusion

This chapter provided an overview of models for categorical latent variables and the combination of categorical and continuous latent variables. Methodologies that were reviewed in this section included latent class modeling, manifest and latent Markov modeling, latent transition analysis, and mixture latent transition analysis (the mover-stayer model). In the context of combining continuous and categorical latent variables, we focused on growth mixture modeling.

The general framework that underlies these methodologies recognizes the possibility of population heterogeneity arising from finite mixture distributions. In the case of the mover-stayer model, the heterogeneity manifests itself in a subpopulation of individuals who do not exhibit any stage transition over time. In the case of growth mixture modeling, the heterogeneity manifests itself as subpopulations exhibiting qualitatively different growth trajectories.

As we noted in the introduction to this chapter, the general framework developed by Muthén and his colleagues is quite flexible, and covering every conceivable special case of the general framework is simply not practical. Suffice to say here that the general framework can be applied to all of the models discussed prior to this chapter—including mixture factor analysis, mixture structural equation modeling in single and multiple groups, mixture MIMIC modeling, and perhaps most interestingly, mixture multilevel structural equation modeling. This latter methodology allows for heterogeneity in the parameters of multilevel models. An application to education would allow models for students nested in schools to exhibit unobserved heterogeneity that might be explained by unique student and school characteristics.

Still another powerful application of the general framework focuses on estimating causal effects in experimental studies—the so-called complier average causal effects (CACE) method (see, e.g., Jo & Muthén, 2001). For example, in a field experiment of an educational intervention, not all individuals who receive the experimental intervention will comply with the protocol. Standard approaches analyze the treatment and control groups via an intent-to-treat analysis, essentially ignoring noncompliance. The result of such an approach can, in principle, bias the treatment effect downward. A viable alternative would be compare the treatment compliers to those in the control group who would have complied had they received the treatment. However, this latter group is unobserved. The CACE approach under the general framework uses finite mixture modeling and information about treatment compliers to form a latent class of potential complier, and forms the experimental comparison between these two groups.⁹

While certainly not exhaustive, it is hoped that this chapter provides the reader with a taste the modeling possibilities that the general framework allows. The models in this chapter scratch only the surface of what has been described as “second-generation” structural equation modeling.

Notes

1. From here on, we will use the term “class” to refer to components of the mixture model. The term is not to be confused with latent classes (e.g., Clogg, 1995) although finite mixture modeling can be used to obtain latent classes (McLachlan & Peel, 2000).

Structural Models for Categorical and Continuous Latent Variables 205

2. Note that latent class models can handle polytomously scored items.
3. For dichotomous items, it is only necessary to present the value of one latent class indicator.
4. Methods for assessing latent class membership over time are discussed in Section 10.4.
5. The sampling design of ECLS-K included a 27% subsample of the total sample at Fall of first grade to reduce the cost burden of following the entire sample for four waves but to allow for the study of summer learning loss (NCES, 2001).
6. A nonstationary Markov model is one that allows heterogeneous transition probabilities over time. In contrast, stationary Markov models assume homogeneous transition probabilities over time.
7. It should be noted that finite mixture modeling has been applied to continuous growth curve models under the name general growth mixture models (B. Muthén, 2004). These models have been applied to problems in the development of reading competencies (Kaplan, 2002), and math competencies (Jordan, Kaplan, Nabors-Olah, & Locuniak, 2006).
8. It is sometimes the case that adding covariates can change the number of mixture classes. See Kaplan (2002) for an example of this problem in the context of reading achievement.
9. This is an admittedly simple explanation. The CACE approach makes very important assumptions—including random assignment, and stable unit treatment value (Jo & Muthén, 2001).

