

The Integration of Continuous and Discrete Latent Variable Models: Potential Problems and Promising Opportunities

Daniel J. Bauer

North Carolina State University

Patrick J. Curran

University of North Carolina at Chapel Hill

Structural equation mixture modeling (SEMM) integrates continuous and discrete latent variable models. Drawing on prior research on the relationships between continuous and discrete latent variable models, the authors identify 3 conditions that may lead to the estimation of spurious latent classes in SEMM: misspecification of the structural model, nonnormal continuous measures, and nonlinear relationships among observed and/or latent variables. When the objective of a SEMM analysis is the identification of latent classes, these conditions should be considered as alternative hypotheses and results should be interpreted cautiously. However, armed with greater knowledge about the estimation of SEMMs in practice, researchers can exploit the flexibility of the model to gain a fuller understanding of the phenomenon under study.

In recent years, many exciting developments have taken place in structural equation modeling, but perhaps none more so than the development of structural equation models that account for unobserved population heterogeneity, or mixtures of unobserved groups. Structural equation mixture models (SEMMs) are ideally suited for testing theories where qualitatively distinct types of individuals are thought to be characterized by different structural relationships between variables (and/or their means) and where group membership cannot be observed directly. For instance, in the context of marketing research, Jedidi, Jagpal, and DeSarbo (1997a) noted that market segments must be defined by the differential appeal of product

features, rather than by a priori classification. Similarly, Arminger and Stein (1997) noted the desirability of identifying unobserved heterogeneity in individual lifestyles in sociological research. More recently, drawing on the work of Nagin and Land (1993) and Verbeke and Lesaffre (1996), B. O. Muthén (2001) demonstrated how SEMMs could be used in psychological and educational research to capture heterogeneity in developmental pathways.

More formally, in SEMM, *latent groups* or *classes* are defined at the distributional level. Each group is composed of a collection of individuals who may differ in their individual scores but are *homogeneous* with respect to the distribution from which they were sampled. Specifically, their observed scores are thought to be drawn from a common multivariate normal distribution, such that the same structural relationships among variables hold for all individuals within the group. *Population heterogeneity* is then indicated by the presence of two or more latent groups in the population characterized by different distributions. Analyzing this type of population heterogeneity has been the traditional domain of finite normal mixture models (Everitt & Hand, 1981; McLachlan & Peel, 2000; Titterton, Smith, & Makov, 1985). The more recent synthesis of finite normal mixture modeling with structural equation modeling represents a major increase in the analytic capabilities of both models (Ariminger & Stein, 1997; Dolan & van der Maas, 1998; Jedidi et al., 1997a; Jedidi, Jagpal, &

Editor's Note. Scott Maxwell served as action editor for this article.—SGW

Daniel J. Bauer, Department of Psychology, North Carolina State University; Patrick J. Curran, Department of Psychology, University of North Carolina at Chapel Hill.

This work was funded in part by National Institute on Drug Abuse (NIDA) Fellowship DA06062, awarded to Daniel J. Bauer, and NIDA Grant DA13148, awarded to Patrick J. Curran. We thank Kenneth Bollen and the members of the Carolina Structural Equations Modeling Group for their valuable input throughout this project.

Correspondence concerning this article should be addressed to Daniel J. Bauer, Department of Psychology, North Carolina State University, Raleigh, NC 27695-7801. E-mail: dan_bauer@ncsu.edu

DeSarbo, 1997b). B. O. Muthén (2001) even suggested that SEMM represents a second-generation latent variable model because it integrates both continuous and discrete latent variable models. Indeed, we may view SEMM as composed of two submodels, a continuous latent variable submodel involving latent factors and a discrete latent variable submodel composed of latent classes, both of which are estimated simultaneously. Although this new modeling architecture is exciting and highly promising, much work remains to be done to determine the specific strengths and potential limitations of the model.

Our goals here are fourfold. First, we highlight that SEMM represents the end result of more than a century of analytic developments on continuous and discrete latent variable models, each of which, in isolation, has been carefully evaluated. We trace the development of SEMM from these more traditional latent variable models and discuss how SEMM may serve as an integrative framework. Second, we argue that although the assumptions of continuous and discrete latent variable models are relatively well understood when considered independently, the implications of synthesizing the two types of models have not yet been fully investigated. We examine this issue by identifying several analytical relationships that exist between discrete and continuous latent variables that may portend difficulty for some SEMM analyses. Third, we formally extend these analytical developments to SEMM to demonstrate the conditions under which the model may not optimally recover the population structure; specifically, we focus on the spurious identification of unobserved groups that do not actually exist in the population. Here our goal is only to delineate general principles that may lead to the identification of spurious latent classes without attempting to thoroughly evaluate the scope or potential impact of these issues under specific sampling conditions. Finally, we use these same general principles to identify several new opportunities for exploring complex empirical relationships with these models when there is little explicit interest in identifying “true” subgroups within the population.

Classical Continuous and Discrete Latent Variable Models

We begin with an overview of two “first-generation” latent variable models for continuous outcomes, the common factor model of Thurstone (1947) and the latent profile model of Gibson (1959). We

open with these models because they were developed expressly as alternative approaches for latent variable analysis. As such, much work has been conducted to demonstrate their analytical similarities and differences. In addition, these models represent the initial steps in a long line of research ultimately resulting in the SEMM. We thus subsequently consider how the analytical relationships revealed by these models generalize to SEMM.

Factor Analysis

Since its inception, one of the interpretations of the correlation coefficient has been that it indicates two variables share a common cause (e.g., Galton, 1888, p. 135; see Stigler, 1986). Factor analysis, originating with Spearman (1904) and greatly extended by Thurstone (1935, 1947), was developed as a tool for identifying these unobserved common causes or *factors*. Traditionally, the common factor model assumes that all of the shared variance (covariance or correlation) among a set of observed variables reflects the presence of continuously distributed latent factors with a common influence on the variables.¹ For example, suppose that a study of high school youth revealed a positive correlation between truancy and illicit drug use. From the perspective of the common factor model, this correlation might be viewed as evidence for a continuously distributed underlying antisocial behavior dimension. After the presence of this underlying common factor is accounted for, the residual relationship between truancy and drug use would approach zero. This illustrates an idea known as the *axiom of local independence*, which holds that once all of the common factors are estimated, the residual associations among the observed variables should be zero, within sampling error.

The population model for the common factor model may be expressed formally as

$$\Sigma = \Lambda\Phi\Lambda' + \Theta, \quad (1)$$

where Σ is a $q \times q$ covariance matrix of the observed variables, Λ is a $q \times m$ factor loading matrix, Φ is an $m \times m$ covariance matrix of the common factors, and Θ is a $q \times q$ covariance matrix of the residuals net of the common factors. The assumption of local inde-

¹ Though he is best known for developing a model for multiple continuous common factors, Thurstone (1935, pp. 51–52; 1947, p. 343) also considered the possibility of dichotomous latent factors.

pendence is formalized by constraining Θ to be a diagonal matrix. With this assumption, the variances on the diagonal of Σ are partitioned into common factor variance and residual (unique) variance, but the off-diagonal covariances in Σ are reproduced exclusively by the relationship of the observed variables to the common factors.

Two other assumptions are key to the common factor model. The first is that the observed variables can be expressed as a linear combination of the latent factors and the residuals. Nonlinearities in these relationships may force the estimation of *difficulty* or *nuisance* factors that do not reflect the true dimensionality of the factor space (Gibson, 1959; McDonald, 1967; Wherry & Gaylord, 1944). Second is the assumption that the relationships among the measures are sufficiently summarized by the second-order moments (variances and covariances), which is strictly true only if the data are multivariate normally distributed. Because the observed variables are linear combinations of the latent factors and residuals, this would imply that the latent factors and residuals are also multivariate normally distributed.²

Latent Profile Analysis

Latent profile analysis is based on a second interpretation of the correlation coefficient, namely, that the correlation reflects the presence of discrete groups in the population, each characterized by different mean levels on the observed variables. To understand how this interpretation differs from the common factor model, consider again the positive correlation between truancy and illicit drug use. The common factor model posited that this correlation reflects the presence of an underlying antisocial behavior dimension. Alternatively, the same observed correlation could instead indicate the presence of two qualitatively distinct types of adolescents, one characterized by high levels of truancy and illicit drug use, and the other characterized by low levels of truancy and illicit drug use. When mixed together, the total population of adolescents would exhibit the observed positive correlation between truancy and drug use.

Latent profile analysis was developed by Gibson (1959) as a means for identifying these latent groups, and was explicitly presented as an alternative to Thurstone's (1947) common factor model. Thus, rather than postulating that continuous latent factors explain the observed associations, the latent profile model holds that the associations are a by-product of differences in the means of the continuous measures over

the latent groups or classes. Paralleling the assumption of local independence in the common factor model, Gibson assumed that, conditional on class membership, the residual association between the observed variables would be zero, again within sampling variability. In this context, local independence is invoked based on the philosophy that the mean vector (or centroid) of a latent class represents the true scores for all individuals of that type. Any deviation from the mean vector should therefore be random and independent.

The fundamental equations of the latent profile model for the variances and covariances of the observed variables were expressed by Lazarsfeld and Henry (1968, chap. 8) in scalar form as

$$\sigma_{ii}^2 = \sum_{k=1}^K \pi_k (\mu_{ik} - \mu_i)^2 + \sum_{k=1}^K \pi_k \sigma_{ik}^2, \quad (2)$$

and

$$\sigma_{ij} = \sum_{k=1}^K \pi_k (\mu_{ik} - \mu_i) (\mu_{jk} - \mu_j), \quad (3)$$

where i and j ($i \neq j$) are index specific variables and k designates a specific latent class, so that μ_{ik} represents the mean and σ_{ik}^2 represents the variance for variable i in group k , K is the total number of latent classes, π_k indicates the proportion of cases belonging to each class (where $\sum_{k=1}^K \pi_k = 1$), and the grand mean of each variable is calculated as a weighted average of the class means, or

$$\mu_i = \sum_{k=1}^K \pi_k \mu_{ik}. \quad (4)$$

Lazarsfeld and Henry (pp. 235–236) noted that Equation 2 represents the familiar decomposition of *total* (or *aggregated*) *variance* into *between-class* and *within-class* components that is used in conventional analysis of variance models. The difference in this case is that group membership is not observed but is estimated instead.

² The assumption of normality is made explicit in maximum-likelihood estimation. Other methods of estimation may permit the inclusion of ordinal indicators and/or relax the assumption of normality, but typically the latent factors are still assumed to be continuously distributed and linearly related, and the second-order moments are treated as sufficient summary statistics.

Estimation of the model relies in part on the assumption of local independence embodied in Equation 3. Note that in contrast to Equation 2, there is no *within-class* component to Equation 3. This means that any association between measures i and j that exists in the aggregate population must be accounted for by the *between-class* component, or the class mean differences. By implication, i and j are assumed to be independent (i.e., orthogonal) within each class k . Of importance, unlike the common factor model, the latent profile model also traditionally uses the higher order moments of the data, including the triple products among three variables that are likewise assumed to be zero (implying higher order independence as opposed to simply pairwise independence; for further detail, see Gibson, 1959; Lazarsfeld & Henry, 1968).

To facilitate further comparison to the common factor model, we can rewrite Equations 2 and 3 more compactly in matrix form in terms of the class covariance matrices and mean vectors so that the aggregated covariance matrix is

$$\Sigma = \sum_{k=1}^K \sum_{l=k+1}^K \pi_k \pi_l (\boldsymbol{\mu}_k - \boldsymbol{\mu}_l) (\boldsymbol{\mu}_k - \boldsymbol{\mu}_l)' + \sum_{k=1}^K \pi_k \Sigma_k, \quad (5)$$

where local independence dictates that Σ_k is diagonal for all k and $\boldsymbol{\mu}_k$ is the centroid of variable means for class k .³ Note that the two terms in Equation 5 directly parallel those in Equation 1. Whereas the factor analysis model decomposes the covariance matrix into common factors and residual variances, the latent profile model decomposes the same covariance matrix into mean differences between discrete latent classes and within-class residual variances.

A Comparison of the Two Models

There are important similarities between the latent profile model and the common factor model, but also key differences. We discuss these relationships here conceptually and illustrate the primary points with several demonstrative artificial data sets. These data sets are used for pedagogical purposes only and are not intended to represent a comprehensive simulation study of the issues at hand.

One similarity between the latent profile model and the common factor model is that both models take local independence to be axiomatic. Because of this shared assumption, both models decompose the covariance matrix into shared variance accounted for by the latent variables and uncorrelated residuals. In fact,

under certain conditions, the two decompositions of the covariance matrix are analytically equivalent. Specifically, a covariance matrix generated to be consistent with an m -factor model can be perfectly reproduced with a $K = m + 1$ class latent profile model and vice versa (Bartholomew, 1987; Gibson, 1959; McDonald, 1967; Molenaar & von Eye, 1994). This fact has led some methodologists to take the stance that the two models should be regarded as mutually complementary (e.g., B. O. Muthén, 2003; B. O. Muthén & Muthén, 2000; but see Meehl, 1995). From this perspective, the common factor model decomposes the covariances to highlight relationships among the variables, whereas the latent profile model decomposes the covariances to highlight relationships among individuals. Because one never knows the true generating model, and because each model can equivalently reproduce the covariances, it could be argued that neither model is superior to the other (i.e., Cudeck & Henly, 2003; B. Muthén, 2003), though we do not necessarily take this view ourselves (Bauer & Curran, 2003b; see also Meehl, 1995).

This equivalency of the two models is represented graphically in Figure 1. The ellipses presented in the figure are 95% confidence ellipses based on the assumption of bivariate normality. The artificially generated sample data consist of 100 cases ($N = 100$), and the sample Pearson product-moment correlation is .67. First consider Figure 1A, representing the common factor model. Here the variables are related to one another in a continuous and linear fashion. The positive correlation between x and y , captured by the upward slope of the major axis of the ellipse, is assumed to reflect the presence of a common underlying latent factor. In comparison, Figure 1B is consistent with a latent profile model (where bivariate normality

³ Equation 5 also represents the multivariate generalization of Meehl's (1965, 1968; Meehl & Golden, 1982) general covariance mixture theorem. Of interest, Meehl's *coherent cut kinetics* method for discriminating latent taxa, like the latent profile model, makes the assumption of independence within categories (i.e., that the covariances between variables are zero within classes). Because the analytic model of coherent cut kinetics appears to be identical to the latent profile model, any empirical differences that arise between the two techniques can probably be ascribed solely to differences in the methods of estimation. Given this relationship between the two models, it is likely that the conclusions we make here regarding the latent profile model can be generalized directly to coherent cut kinetics as well.

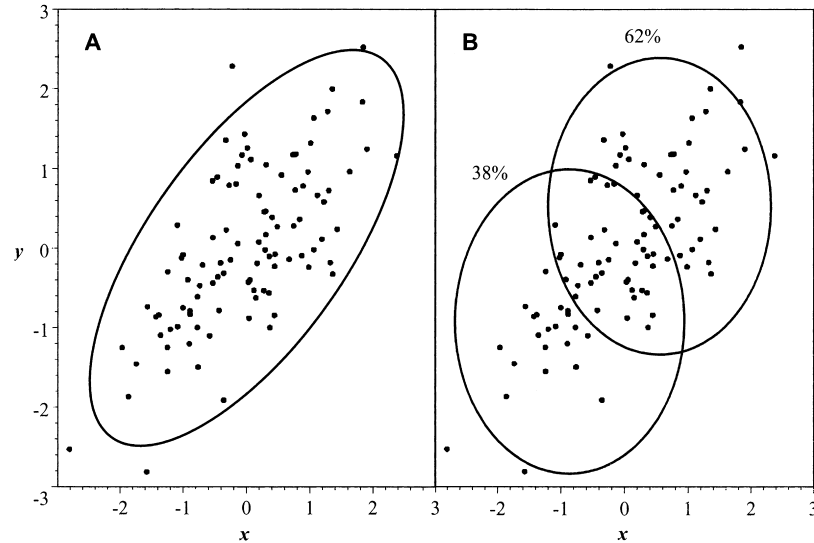


Figure 1. A and B are scatter plots containing 100 data points ($N = 100$) generated from a bivariate normal distribution. The 95% confidence ellipse in A shows how a common factor model would reproduce the data. The 95% confidence ellipses in B show how a two-class latent profile model would reproduce the data (the percentage of cases in each class is indicated).

is assumed within components for convenience). Note that in Figure 1B there is no tilt to the ellipses for the two latent classes, reflecting the assumption of local independence (that x and y are uncorrelated within classes). The overall correlation between x and y in the aggregate data is instead captured by the different class centroids, or mean vectors. Specifically, if one were to draw a line between the class centroids in Figure 1B, it would parallel the major axis of the tilted ellipse in Figure 1A, illustrating that the two models equivalently reproduce the correlation between x and y .

Given this fundamental relationship between the two models, Molenaar and von Eye (1994) remarked that the choice of continuous versus discrete scaling for the latent variables is essentially arbitrary, if we confine our analysis to first and second moments (i.e., means and covariances). The common factor model, because it assumes multivariate normality and linearity, only involves the analysis of these moments. In contrast, as Lazarsfeld and Henry (1968, pp. 228–229) noted, the latent profile model uses the higher order moments of the data and so does not assume multivariate normality of the aggregate distribution (though normality within classes may be assumed; e.g., Arminger & Stein, 1997; Lazarsfeld & Henry, 1968, pp. 235–239). For the same reason, unlike the common factor model, the latent profile model does

not require that the variables be linearly related (Gibson, 1959).

Figures 2 and 3 highlight these differences between the common factor and latent profile models. Figure 2 presents the same contrast as in Figure 1, but here the data were generated to be slightly skewed ($N = 100$; $r = .75$; $skew_x = .64$; $skew_y = .82$). Most of the data points are observed in the lower left quadrant, and both the sparseness and the spread of the data increase as we move to the upper right quadrant. The confidence ellipse in Figure 2A does not optimally represent the characteristics of the observed data because it assumes bivariate normality. In contrast, the latent profile model in Figure 2B provides a better summary of the data. The dense data in the lower left quadrant are captured by one large group with small variances for x and y , and the disparate data in the upper right quadrant are captured by a smaller group with larger variances for x and y . Note that both models equivalently recover the overall correlation between x and y , but the latent profile model also captures the nonnormality of the data.

Finally, Figure 3 demonstrates the performance of the two models when x and y are nonlinearly related (which also implies that bivariate normality will be violated). The generated data show a strong positive relationship ($N = 100$, $r = .74$), but y increases ex-

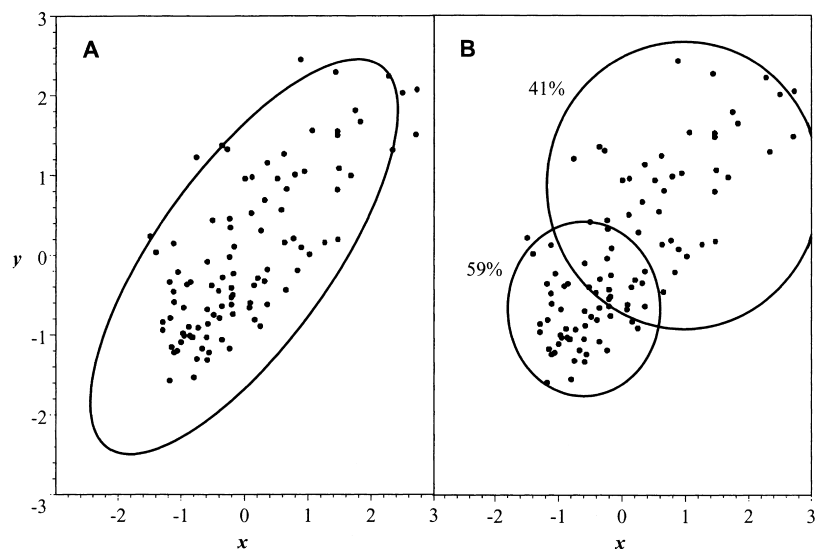


Figure 2. A and B are scatter plots containing 100 data points ($N = 100$) generated from a skewed bivariate distribution. The 95% confidence ellipse in A shows how a common factor model would reproduce the data. The 95% confidence ellipses in B show how a two-class latent profile model would reproduce the data (the percentage of cases in each class is indicated).

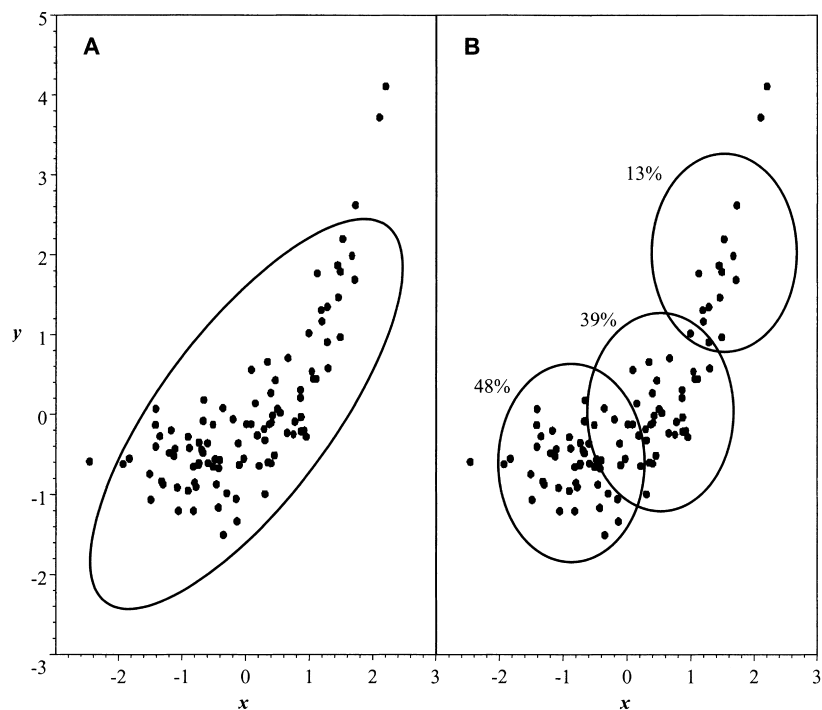


Figure 3. A and B are scatter plots containing 100 data points ($N = 100$) generated to show a nonlinear relationship. The 95% confidence ellipse in A shows how a common factor model would reproduce the data. The 95% confidence ellipses in B show how a three class latent profile model would reproduce the data (the percentage of cases in each class is indicated).

ponentially with x —specifically, $E(y|x) = -.70 + .42e^x$ in the population. The confidence ellipse in Figure 3A represents the common factor model, with its assumptions of normality and linearity. This ellipse captures the overall positive relationship between the two variables but fails to account for the nonlinear form of this relationship. In contrast, the latent profile model depicted in Figure 3B, including three latent classes, captures the nonlinear relationship between x and y well. This result is explained analytically by McDonald (1967), who noted that a nonlinear factor model approximated well by a polynomial of degree m can be equally represented by a latent profile model with $K = m + 1$ points (a result that is related to the fact that $m + 1$ points is the minimum number needed to define a polynomial of degree m), at least at the level of the second-order moments.

In summary, both the common factor model and the latent profile model have the ability to account for the mean and covariance structure of a set of continuous measures. However, the two models are not equivalent beyond the level of the first and second moments. Because the latent profile model uses information about the higher order moments, it can more flexibly capture nonnormality and nonlinearity in the data. Of importance, these differences between the common factor model and latent profile model reflect a general asymmetry between continuous and discrete latent variables. This asymmetry has important implications for models such as the SEMM that involve the simultaneous estimation of both continuous and discrete latent variables. In the next section, we cast the SEMM as a hybrid model that may be viewed as an extension of more traditional continuous and discrete latent variable models (see also Arminger & Stein, 1997; B. O. Muthén, 2002). The common basis of SEMM with these models suggests that the principles we highlighted here for the common factor model and the latent profile model also hold in SEMM. We demonstrate this point with several examples illustrating how the asymmetrical capabilities of discrete and continuous latent variables may influence SEMM analyses.

Structural Equation Mixture Models

Relation to the Common Factor Model

The intellectual progression from the common factor model to SEMM is relatively direct, and we thus provide only a cursory review of this development here. See Bollen (1989, pp. 4–9) and Kaplan (2000,

chap. 1) for further details on various steps in this progression. First, the common factor model may be viewed as the antecedent of confirmatory factor analysis (CFA). CFA is predicated on the same notions as the common factor model and follows the same basic form as Equation 1. Where CFA differs from the common factor model is in the imposition of restrictions on the parameters (typically the factor loadings). These restrictions are used to achieve model identification, (Bollen, 1989, pp. 238–253) and may allow the axiom of local independence to be relaxed.⁴ In turn, the CFA model may be viewed as the antecedent of the structural equation model. Whereas in CFA the correlations among latent factors are typically freely estimated (that is, unrestricted), in structural equation modeling these correlations may be structured to reflect causal relations (Jöreskog, 1973; see also Bollen, 1989).

A key assumption of standard CFA and structural equation modeling is that the population is homogeneous, so that a single covariance matrix can be used to summarize the relationships among the variables. To overcome this assumption, Jöreskog (1971) extended the model to multiple samples. These models allow for the simultaneous estimation of structural equation models in two or more groups, each characterized by their own mean vector and covariance matrix. Multiple-groups structural equation modeling is particularly useful for testing hypotheses about the invariance of a model over samples (e.g., the factor structure of a test for males and females). Although the multiple-groups framework explicitly accounts for population heterogeneity, its limitation is the requirement that the groups be defined a priori. SEMM may then be viewed as the logical next step, extending the multiple-groups model to the case in which group membership is not observed and must instead be inferred inductively from the data (Jedidi et al., 1997a; Yung, 1997).

Relation to the Latent Profile Model

Surprisingly, the latent profile model appears to have had little direct influence on the development of SEMM. It is instructive to better understand why this

⁴ For instance, it is common to intercorrelate the residuals of observed variables that share some common measurement property. From the perspective of the common factor model, such a correlation would be viewed as evidence of a common *method factor*.

was the case. As with the common factor model, the latent profile model was predicated on the axiom of local independence. The two models were thus mutually contradictory: Either the observed correlations reflect the presence of continuous underlying factors, or they reflect the presence of unobserved groups, but not both. The mutual reliance of the two models on this axiom precluded their integration.

The development of SEMM may be traced more directly to a different kind of discrete latent variable model, namely, finite normal mixture models. Finite normal mixture models assume that the population consists of a mixture of unobserved groups, each characterized by its own normal distribution for the continuous measures. The univariate normal mixture model was considered by Pearson (1894), with practical applications to multivariate settings awaiting the computer revolution (Wolfe, 1970, 1971). The probability density function (PDF) of the multivariate normal mixture model is

$$f(\mathbf{x}) = \sum_{k=1}^K \pi_k \phi_k(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \quad (6)$$

where $\mathbf{x} = (x_1, x_2, \dots, x_q)$ is a vector of q continuous random variables, π_k are again the class proportions (where $\sum_{k=1}^K \pi_k = 1$), and the variables within each class follow a multivariate normal PDF (denoted ϕ_k) with mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\boldsymbol{\Sigma}_k$. For further detail on this model (as well as mixtures of other distributions), see Everitt and Hand (1981), Titterton et al. (1985), or McLachlan and Peel (2000).

There are two key differences between finite normal mixture models and the latent profile model. First, in the finite normal mixture model, the parametric assumption of normality within groups is explicit and is relied upon to estimate the model parameters. Second, given this, the axiom of local independence need not be invoked for the model to be estimable. As such, the off-diagonal elements of the within-class covariance matrices $\boldsymbol{\Sigma}_k$ may be freely estimated (assuming the model is identified). Despite these differences, the basic formulas for calculating the means $\boldsymbol{\mu}$ and covariances $\boldsymbol{\Sigma}$ of the aggregate distribution $f(\mathbf{x})$ remain identical to Equations 4 and 5 presented earlier for the latent profile model. By implication, finite normal mixture models have many of the same basic properties as latent profile analysis (i.e., the latent classes can serve many of the same functions).

These differences between the latent profile model and finite normal mixture model are deceptively

simple, for they actually represent a radical difference in the underlying conceptual model. In the latent profile model, the assumption of local independence was imposed not simply for analytical convenience but on the philosophy that the latent class variable *should* explain the associations in the data. That is, controlling for the mean differences between the groups, there should be no residual association between the variables. This notion was based on the conceptual model that all individuals within a class should be characterized by identical true scores as represented by the class centroid. Apart from measurement error, all individuals within a class would then be essentially interchangeable and thus truly homogeneous. Residual variability within a class, reflecting only random measurement error, would be uncorrelated by definition.

In contrast, the finite normal mixture model involves a much different conceptual model. Specifically, homogeneity is no longer defined in terms of a common set of true scores (i.e., interchangeability). Instead, homogeneity is defined at the distributional level, by a common mean vector and covariance matrix. From this perspective, the latent class variable is not an explanatory variable (as it was in the latent profile model) but is instead a moderator variable (Jedidi et al., 1997a, 1997b). That is, the magnitude and direction of the relationships between variables may vary as a function of class. In terms of our earlier example, truancy and drug use may be positively correlated in one group of adolescents, orthogonal in another, and negatively correlated in a third (e.g., for whom the school provides a context for drug use); thus, the relation between truancy and drug use depends on class membership, but it is not explained by class membership.

It is this latter definition of homogeneity that permits the integration of finite normal mixture models with contemporary continuous latent variable models. Specifically, given that each latent class is defined by its own mean vector and covariance matrix, a logical next step is to structure these moments to be consistent with a continuous latent variable model. This step was first taken by Blåfield (1980) and Yung (1997), who demonstrated that a mixture of factor analysis models could be estimated by specifying that each within-class covariance matrix $\boldsymbol{\Sigma}_k$ is structured according to Equation 1. By doing so, researchers can estimate latent classes that may differ in factor structure, factor means, variances or covariances, factor loadings, and/or residual variances (subject to identi-

fiability constraints). Jedidi et al. (1997a, 1997b), Arminger and Stein (1997), and Dolan and van der Maas (1998) each independently generalized this approach to mixtures of structural equation models with latent variables. The basic form for SEMM is the same as Equation 6, with the only difference being that the class PDFs are rewritten,

$$\phi_k [\mathbf{x}; \boldsymbol{\mu}_k(\boldsymbol{\theta}_k), \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k)], \quad (7)$$

to indicate that both the class mean vector $\boldsymbol{\mu}_k$ and the class covariance matrix $\boldsymbol{\Sigma}_k$ are modeled as a function of a class-specific vector of model parameters $\boldsymbol{\theta}_k$.⁵ The analytic goal is thus to fit a structural model to the within-class means and covariances. This basic model was recently extended by B. Muthén and Shedden (1999) to include categorical outcomes and covariates of latent class membership (for greater detail, see L. K. Muthén & Muthén, 1998).

SEMM as an Integrative Framework

Implicit in our preceding discussion is that SEMMs represent a synthesis and generalization of many other latent variable models. This is highlighted by considering the specific conditions under which SEMM reduces to the other models we have discussed. For instance, if $K = 1$, then SEMM reduces to a conventional single-group CFA or structural equation model. Alternatively, if $K > 1$, and both $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are fully unconstrained, the model becomes a standard multivariate finite normal mixture model (Jedidi et al., 1997a, 1997b). Finally, if $K > 1$, $\boldsymbol{\mu}_k$ is unconstrained, and $\boldsymbol{\Sigma}_k$ is constrained to be a diagonal matrix, the model reduces to the latent profile model (with a formal assumption of normality within classes; Arminger & Stein, 1997). Thus, both conceptualizations of homogeneity discussed above (interchangeability and distributional homogeneity) can be accommodated within the SEMM. The relation of SEMM to traditional multiple-groups CFA or structural equation modeling is also straightforward. Specifically, in the estimation process, probabilities of group membership (posterior probabilities) are estimated for each case and are used to weight each case's contribution to the estimation of the model parameters for each class. If the grouping variable was observed, these probabilities could be replaced by 1s (indicating inclusion in the group) or 0s (indicating exclusion), and the model would reduce to the standard multiple-groups structural equation model (Jedidi et al., 1997b; L. K. Muthén & Muthén, 1998, Appendix 9; Zhu &

Lee, 2001). SEMM may thus be thought of as multiple-groups structural equation modeling where group membership is an unobserved or latent variable.

To summarize thus far, we note that the development of SEMM may be viewed as the end result of converging lines of research on continuous and discrete latent variable models that began more than a century ago. However, little research has yet been conducted to examine the implications of combining the two models. It is to this issue that we now turn.

Combining Continuous and Discrete Latent Variables

As we noted in our discussion of the common factor model and the latent profile model, the relationship between continuous and discrete latent variables is both complex and asymmetrical. Whereas continuous and discrete latent variables are equivalently capable of reproducing the first two moments of the data (means and covariances), discrete latent variables can also accommodate nonnormal data and nonlinear relationships in a way that continuous latent variables typically cannot. We now show how these same properties of continuous and discrete latent variables surface in important ways in the SEMM. Paralleling our discussion of the common factor model and the latent profile model, we focus on three issues: model specification, distributional assumptions, and the assumption of linearity. We show that, apart from the presence of true latent groups, misspecification of any of these aspects of the model may lead to the estimation of multiple latent classes.

Of importance, although we both analytically and empirically demonstrate that spurious latent classes can be extracted for any one of these reasons, we make no conclusions about the extent to which these problems will arise in applied research. To address the

⁵ For simplicity, we assume all variables in the SEMM are endogenous. Arminger and Stein (1997) and Arminger, Stein, and Wittenberg (1999) have noted that models with categorical or nonnormal exogenous predictors (e.g., gender, marital status) should be based on conditional finite normal mixtures. Otherwise, the distribution of the exogenous variables will influence latent class estimation and potentially bias within-class parameter estimates. The conditional finite normal mixture differs in important ways from the unconditional model we describe here, but we do not dwell on the distinction here because our points generalize directly to the conditional case.

latter issue would require comprehensive simulation studies beyond the scope of the current article to determine more precisely the conditions under which spurious latent classes will be estimated when these assumption violations occur, either alone or, as seems more likely, in combination.

In the examples that follow, we approach the problem of identifying the optimal number of latent classes from a largely exploratory vantage point. Used in an exploratory way, the analysis begins with the hypothesis that multiple groups may exist and focuses on identification of the optimal number of classes based on the comparative fit of models with successively more latent classes. Here we are concerned with whether the number decided upon includes spurious latent classes. A more confirmatory approach to the problem would involve a strong theoretical prediction that the population consists of a specific number of groups. It may be argued that in this confirmatory mode, our theory would in some way protect us from accepting a model with too many (or too few) latent classes. However, even in this case, the theory must be supported by showing that the model with the predicted number of classes fits optimally when compared with models with fewer or more latent classes. We could, of course, elect to estimate the predicted number of classes *perforce*, but this would reverse the typical practice of inferential statistics. That is, we would be using theory to dictate the optimal statistical model for the data rather than using the statistical model to test the theory against the data (Bauer & Curran, 2003a, 2003b). For this reason, we regard the issues we raise as being equally important for exploratory and confirmatory SEMM analyses.

Of course, this raises the question, How does one define the optimally fitting model? This matter is complicated by the fact that regularity conditions do not hold for conducting standard likelihood ratio tests between models with different numbers of latent classes (see McLachlan & Peel, 2000, for a review). A modified version of the likelihood ratio test, based on a sum of chi-square distributions, was recently proposed by Lo, Mendall, and Rubin (2001) and has been advocated by B. Muthén (2003), though we caution that the performance of this test has yet to be investigated in complex models of the kind considered here (Bauer & Curran, 2003b).

In practice, it is still common to determine the optimal number of latent classes through the comparison of information criteria. Commonly used measures of fit include Akaike's information criterion (AIC),

Bayes's information criterion (BIC), the consistent AIC (CAIC), the classification likelihood criterion (CLC), the normalized entropy criterion (NEC), and the integrated completed likelihood criterion with BIC approximation (ICL-BIC). These criteria all balance the improvement in fit associated with adding classes to the model against specific penalty factors. Whereas the AIC, BIC, and CAIC all penalize for the number of parameters in the model (overparameterization), the CLC and NEC instead involve penalties based on the model entropy, which increases as the degree of separation between the latent classes decreases (i.e., when individual probabilities of class membership stray far from zero or one). The ICL-BIC implements both types of penalty and is the most conservative of the criteria presented here, generally favoring fewer classes. For each index, the model with the minimum value is considered optimal. More detail on these measures can be obtained from McLachlan and Peel (2000) or Bauer and Curran (2003a).

These measures too must be used with some caution. Although the behavior of these indices has been well studied for unrestricted finite normal mixtures, it has not yet been investigated thoroughly in an SEMM context. This is especially true of fit indices that depend on the model entropy, such as the CLC, NEC, and ICL-BIC, because they tend to favor models in which classes are distinguished by well-separated mean vectors, as it is typically in these cases that one could assign cases to classes with the most precision (Biernacki, Celeux, & Govaert, 1999, 2000; Celeux & Soromenho, 1996; Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993). In SEMM analyses, however, attention is often focused instead on the different covariance matrices of the classes (i.e., how the latent classes moderate the relationships among the observed variables). For this reason, Jedidi et al. (1997a, 1997b) recommended determining the number of classes based on the BIC and CAIC, with entropy criteria used only subsequently to judge the degree of separation between classes.

In addition to these indices, following Jedidi et al. (1997a), we also present the standardized root-mean-square residual (SRMR) for each model. In the SEMM context, the SRMR is calculated by first computing the model-implied aggregate covariance matrix according to Equation 10. The usual formula for the SRMR is then applied, given as

$$\text{SRMR} = \sqrt{\frac{2}{q(q+1)} \sum_{i=1}^q \sum_{j=1}^i \left[\frac{(s_{ij} - \hat{\sigma}_{ij})}{\sqrt{s_{ii}} \sqrt{s_{jj}}} \right]^2}, \quad (8)$$

where s_{ij} are elements of the observed aggregate covariance matrix, \hat{s}_{ij} are elements of the model-implied aggregate covariance matrix, and q is the total number of observed variables. Hu and Bentler (1999) suggested that an SRMR of .08 or below represents good fit in traditional structural equation models. Because the covariances are not sufficient statistics for normal mixtures, Jedidi et al. (1997a) argued that the SRMR should not be used to determine the number of latent classes in an SEMM. We agree. However, in juxtaposition with the other fit statistics, the SRMR does allow us to see whether the improvement in fit between successive models takes place primarily at the level of the covariance matrix (in which case the SRMR should decrease) as opposed to the higher order moments (in which case the SRMR should be stable).

We now consider in turn how model misspecifications, violations of distributional assumptions, or unmodeled nonlinearity can all lead to the situation in which the optimally fitting model includes spurious latent classes.

Model Specification

Consider again the data presented in Figure 1. Suppose we know that these data were generated to be consistent with a bivariate normal distribution for a single group. We may then view the ellipse in Figure 1A as symbolic of a properly specified model. In this case, the overall association between the two variables is captured using just one ellipse. However, what if we misspecified the model—for instance, by imposing the constraint that x and y are uncorrelated? Then we could not account for the association between x and y using just one ellipse; we would require two, as in Figure 1B. In other words, we can compensate for the lack of fit of the within-class model (which incorrectly imposes orthogonality) by also misspecifying the between-class model (estimating more classes than groups in the population). In a very real sense, the spurious latent classes “absorb” the misspecification of the model.

This relationship between the two parts of the model can be expressed more formally by considering the implied means and covariances of the aggregate distribution:

$$\boldsymbol{\mu}(\boldsymbol{\pi}, \boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \boldsymbol{\mu}_k(\boldsymbol{\theta}_k), \quad (9)$$

$$\begin{aligned} \boldsymbol{\Sigma}(\boldsymbol{\pi}, \boldsymbol{\theta}) = & \sum_{k=1}^K \sum_{l=k+1}^K \pi_k \pi_l [(\boldsymbol{\mu}_k(\boldsymbol{\theta}_k) - \boldsymbol{\mu}_l(\boldsymbol{\theta}_l))(\boldsymbol{\mu}_k(\boldsymbol{\theta}_k) - \boldsymbol{\mu}_l(\boldsymbol{\theta}_l))'] \\ & + \sum_{k=1}^K \pi_k \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_k), \end{aligned} \quad (10)$$

where $\boldsymbol{\pi}$ is the vector of class proportions, $\boldsymbol{\theta}$ is a vector of model parameters from all latent classes, and all other notation remains as before. Note that these equations follow directly from Equations 4 and 5 given for the latent profile model, demonstrating a general property of discrete latent variable models for continuous data. Of importance, as before, the aggregate covariance matrix is partitioned into two components. The first term in Equation 10 is the between-class component, reflecting the mean differences between the classes, whereas the second term reflects the within-class component, or the covariance structure within each class.

There is a fundamental interdependency between the two components of Equation 10. Recall that both discrete latent classes and continuous latent factors hold the potential to reproduce the aggregate covariances. This interdependency becomes problematic if the structural model applied within the latent classes is misspecified. In this case, the within-class component will fail to fully reproduce the covariances, and the model will not fit well. One possible way to improve the fit of the model would be to modify the within-class portion, for instance, by adding new paths. However, we could also compensate for the poor fit of the within-class model by estimating additional latent classes in the between-class component.⁶

We provide an example of this point with a single artificial data set generated from the linear latent curve model presented in Figure 4. The repeated measures span five equally spaced time points, they are multivariate normally distributed in the population, and 50% of their variance is accounted for by the underlying growth process. Note that the fixed values of the factor loadings imply linear change over time. The means of the latent factors indicate that the average trajectory is increasing over time. Both the intercept and slope parameters have significant variances and are positively correlated. A single sample of 600 cases was generated from the model, and the

⁶ Additional latent classes may also be required when the mean structure is misspecified, though we do not explore that possibility here.

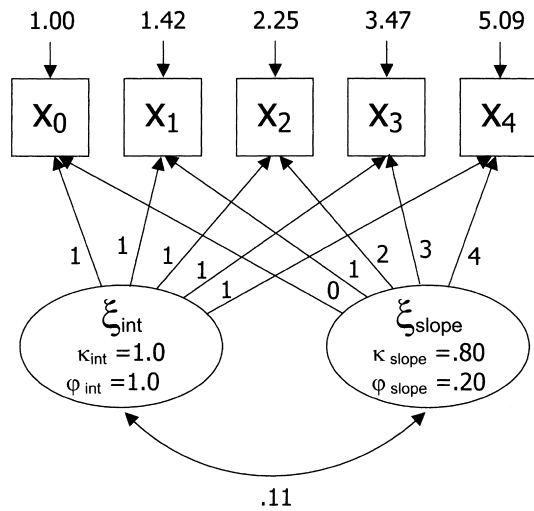


Figure 4. Path diagram displaying the population latent curve model used to generate data for the demonstration that model misspecification may induce the estimation of latent classes. κ designates the factor mean and φ designates the factor variance.

resulting sample statistics are presented in Table 1. Fitting a properly specified single group latent curve model to the data yielded good fit, $\chi^2(10) = 4.67$, $p = .91$; root-mean-square error of the approximation (RMSEA) = .00.⁷ We then extended this model to two classes, and as shown in Table 2, the addition of a second class was rejected by all measures of fit except the AIC. Thus, as one would expect, under proper specification of the within-class structural model, misspecification of the number of latent classes was not supported by the data.

We next fit a misspecified model to this same sample, allowing between one and eight latent classes. The misspecification was to restrict the covariance

Table 1
Intercorrelations, Means, Standard Deviations, Skew, and Kurtosis of Data Generated From the Linear Latent Curve Model ($N = 600$)

Variable	1	2	3	4	5
1. x_0	—				
2. x_1	.477	—			
3. x_2	.403	.507	—		
4. x_3	.368	.504	.533	—	
5. x_4	.291	.440	.493	.510	—
M	1.031	1.851	2.618	3.438	4.235
SD	1.362	1.671	2.226	2.621	3.139
Skew	-.180	-.055	-.152	.078	.094
Kurtosis	.207	-.086	-.230	-.111	.106

matrix of the latent growth factors to a null matrix, that is, to permit no individual variability in either intercepts or slopes within classes. This specification corresponds to one popular variant of the SEMM, referred to as latent class growth analysis (LCGA) by B. O. Muthén (2001), and inspired by the semi-parametric groups-based trajectory model of Nagin and colleagues (Nagin, 1999). By imposing these restrictions to the factor covariance matrix, the LCGA model implies that all of the individual variability in growth is captured by the class mean trajectories (i.e., fixed effects), and any individual deviations from the class mean trajectories are random error. This assumption is commensurate with the latent profile model conceptualization of homogeneity as interchangeability: Individuals of a given class share a single trajectory of change over time, and any individual deviations from that trajectory are construed as random error. In fact, the LCGA can be viewed as a restricted latent profile model for repeated measures that constrains the within-class means to follow a specific time trend (e.g., linear).

Given the characteristics of the population generating model, the single class LCGA model contains the correct number of groups but a misspecified structural model that does not allow for systematic individual variability around the mean trajectory. LCGA models with more classes are then doubly misspecified, as they are estimated with both the incorrect number of groups and a misspecified within-class structural model. Intuitively, one might expect that this additional misspecification would lead to a decrement in model fit. In reality, as Equation 10 demonstrates, estimation of additional latent classes (i.e., the between-class component) predictably allows us to better recover the aggregate covariance matrix of the observed data in the presence of a misspecified structural model. We should thus expect the overall model fit to improve as we estimate more latent trajectory classes even though we are increasingly misspecifying the number of classes in the model.

The results of fitting the LCGA models are provided in Table 2. Large and significant improvements in the log-likelihood were obtained with the additions of a second and third latent trajectory class, but the improvements thereafter were nonsignificant and increasingly negligible. It is also worth noting that with

⁷ Mplus (Version 2.12) was used to estimate this model and all models reported hereafter (L. K. Muthén & Muthén, 1998).

Table 2
Fit Statistics of Correctly Specified One- and Two-Class Latent Curve Models and Latent Class Growth Analyses With One Through Eight Classes

Model	LL	SRMR	AIC	BIC	CAIC	CLC	NEC	ICL-BIC
Latent curve model (individual trajectories distributed around mean trajectory)								
One class	-6,050.73	.027	12,121.46	12,165.43	12,175.43	12,101.46	1.00	12,165.43
Two classes	-6,038.69	.026	12,119.38	12,211.72	12,232.72	12,383.52	12.72	12,517.86
Latent class growth model (no individual trajectories—fixed class mean trajectories only)								
One class	-6,492.43	.374	12,998.85	13,029.63	13,036.62	12,984.84	1.00	13,029.63
Two classes	-6,175.36*	.110	12,380.72	12,446.67	12,461.67	12,570.83	0.35	12,666.78
Three classes	-6,089.23*	.060	12,224.47	12,325.60	12,348.59	12,544.36	0.45	12,691.49
Four classes	-6,059.81	.050	12,181.62	12,317.93	12,348.92	12,553.69	0.50	12,751.99
Five classes	-6,043.50	.035	12,165.00	12,336.48	12,375.48	12,659.14	0.64	12,908.61
Six classes	-6,030.25	.033	12,154.50	12,361.15	12,408.16	12,755.87	0.75	13,056.53
Seven classes	-6,014.40	.029	12,138.79	12,380.63	12,435.63	12,738.86	0.74	13,090.69
Eight classes	-6,009.77	.028	12,145.54	12,422.54	12,485.55	12,769.38	0.78	13,172.38

Note. For AIC, BIC, CAIC, CLC, NEC, and ICL-BIC, the minimum value obtained from the series of models is in boldface to indicate the model with optimal fit. LL = log-likelihood; SRMR = standardized root-mean-square residual; AIC = Akaike's information criterion; BIC = Bayes's information criterion; CAIC = consistent AIC; CLC = classification likelihood criterion; NEC = normalized entropy criterion; ICL-BIC = integrated completed likelihood criterion using BIC approximation. Asterisks attached to LL values indicate a significant improvement in log-likelihood over the preceding model as judged by the test of Lo, Mendell, and Rubin (2001).

five or more latent classes the log-likelihood actually dipped below the value of the correctly specified one-group latent curve model, suggesting that the latent class model was overfitting the data (capturing random sampling fluctuations). The fit criteria varied in terms of their indication of the optimal model, ranging from seven classes (the AIC) to two classes (the ICL-BIC). This lack of consensus among fit criteria is not uncommon in empirical applications and simply reflects differences in the penalty factors applied.

Our immediate concern, however, is not to arrive at a definitive conclusion regarding the optimal number of latent classes but to show how our ability to reproduce the aggregate correlation matrix improves as new classes are added to the model (even though such classes do not exist in the population). This is most directly measured by the SRMR. The SRMR of the one-class model indicates that, on average, we are either under- or overestimating the aggregate correlations in Table 1 by about .375. The reason is that the one-class model implies that the off-diagonal elements of the aggregate correlation matrix are zero, when in reality (i.e., within the population) they are not. The additional latent classes present in the two- to eight-class models allow these off-diagonal elements to take on nonzero values, where these values are implied by the between-class mean differences as specified in Equation 10. Predictably, we see that the

precision with which we recover the aggregate correlations improves rapidly as we add latent classes, and we attain remarkably good fit with three or four classes ($SRMR_3 = .06$; $SRMR_4 = .05$).

These results provide a demonstration of the key dependency between the within-class and between-class components of an SEMM that we identified analytically in Equation 10. In our example, a properly specified single-group latent curve model provided good fit to the data (as would be expected given its correspondence to the population model). However, when the variance-covariance matrix of the latent growth factors was constrained to be zero, multiple trajectory classes were needed to attain comparable fit. This difference is evident in Figure 5. Figure 5A presents the 95% confidence bands for the individual trajectories implied by the correctly specified latent curve model. For comparison, Figure 5B presents the results of the four-class LCGA. Because individual variability around the mean trajectories is not permitted in LCGA, multiple classes are required to coarsely approximate the essential features of the correctly specified model in Figure 5A.

It is worth noting that a different number of latent classes might have appeared optimal had we imposed a different model misspecification. For instance, in the context of LCGA, White, Johnson, and Buyske (2000) suggested placing an autoregressive structure

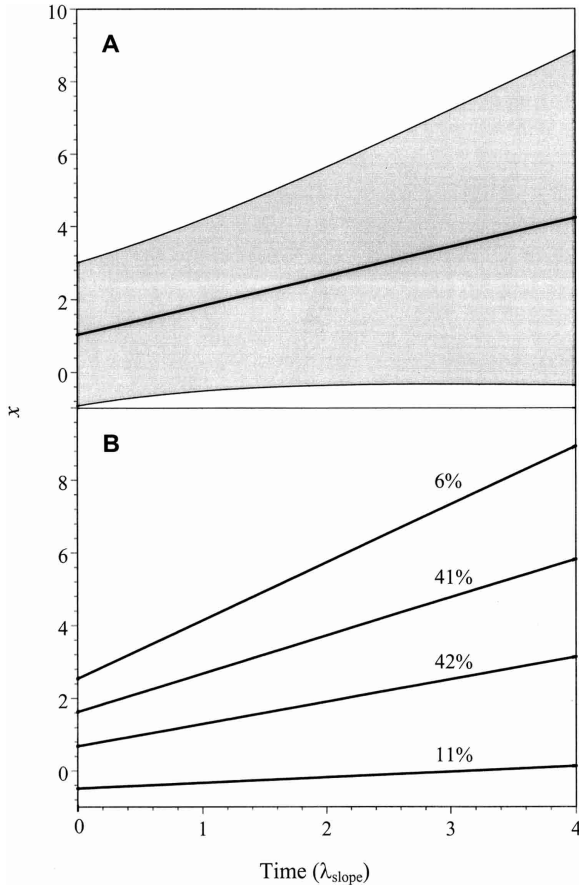


Figure 5. Comparison of single-group latent trajectory model (A; shaded region is within the implied nonsimultaneous 95% confidence bands of the individual trajectories) and the four-class latent trajectory class analysis (B; the percentage of cases in each class is indicated).

on the covariance matrix of the residuals within classes. Because this model would allow for nonzero off-diagonal covariances within classes, it would likely require fewer latent classes to adequately recover the aggregate covariances of a latent curve model. More generally, our first observation is that the number of latent classes needed to optimally fit a given set of data will depend on the degree of misspecification of the within-class structural model.

Distributional Assumptions

As we saw in Figure 2, unlike most standard continuous latent variable models, discrete latent variable models are capable of capturing nonnormality in the observed variables. This point is widely recognized in the general literature on finite normal mixture models. For instance, one key use of finite normal mixtures is to approximate nonnormal distributions of an unspeci-

fied form (e.g., Escobar & West, 1995; Ferguson, 1983; Land & Nagin, 1996; Nagin, 1999; Roeder & Wasserman, 1997; for a review, see Everitt & Hand, 1981, pp. 118–124; Titterton et al., 1985, pp. 18–34). By extension, when the normality assumption is violated in a standard one-group structural equation model, elaboration of the model into an SEMM with multiple latent classes may improve the ability of the model to reproduce the observed data distribution. In this case, even if the structural equation model is not structurally misspecified, the additional latent classes will improve the fit of the model because the summation of the normally distributed class distributions will provide a better approximation to the nonnormality of the aggregate distribution. Bauer and Curran (2003a) extensively explored this point both analytically and empirically in the context of growth mixture modeling. We thus illustrate the role of nonnormality here (in the broader context of SEMM) with a single empirical example.

Our example is a CFA model of two correlated factors, each with three indicators. The population model and associated parameter values are presented in Figure 6. The indicators (x_1 – x_6) were assigned unit loadings and zero intercepts, and 60% of their variance was explained by the associated latent factor. The residuals were independently and normally distributed with variances of .67, and the latent factors were given zero means, variances of one, and a correlation of .71 ($\rho^2 = .50$). All of these parameter values were selected to be reflective of a model that might commonly be encountered in applied research. Of importance, the population distributions of the latent variables were modestly nonnormal, with univariate skew = 2 and kurtosis = 8. As shown in Figure 7, these distributions were marked by a high leftward peak and long right tail. A total of 400 cases were generated from the model. Data were first generated for the latent variables (including the residuals), and these values were then used to construct the model-implied observed scores (the technique of Vale & Maurelli, 1983, was used to generate values for the nonnormal latent factors). The sample skew and kurtosis for the first latent factor, ξ_1 , were 1.63 and 4.69, respectively, whereas the corresponding values for the second latent factor, ξ_2 , were 2.02 and 6.42. The sample statistics for the observed variables are provided in Table 3. Note that the sample skew and kurtosis of the observed measures are well within the range commonly encountered in applied psychological research (Micceri, 1989).

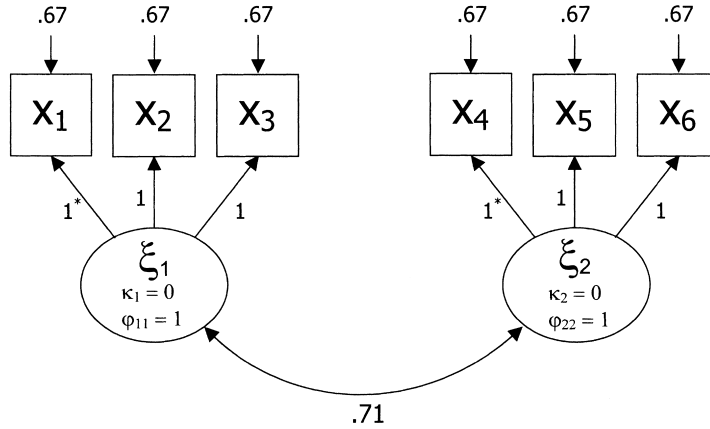


Figure 6. Path diagram displaying the population confirmatory factor model used to generate data for the demonstration that nonnormality may induce the estimation of latent classes. In the estimated models the loadings with asterisks were fixed to scale the latent factors. κ designates the factor mean and φ designates the factor variance.

We began our analysis by fitting the population generating model to the data. The factors were scaled by setting the factor loadings for x_1 and x_3 to one and their intercepts to zero. All other factor loadings and intercepts were estimated, as were the factor means. Even with the nonnormality of the data, the overall fit of the model was quite good, $\chi^2(8) = 3.70, p = .88$; RMSEA = .00. This is not surprising, as the degree

of nonnormality in the observed variables was modest, and at the level of the first and second moments, the fitted model was correctly specified.

We then proceeded to fit SEMMs with two and three latent classes (with the knowledge that the population consists of only one group). The within-class structural models were of the same form as the population model (i.e., correctly specified), but the param-

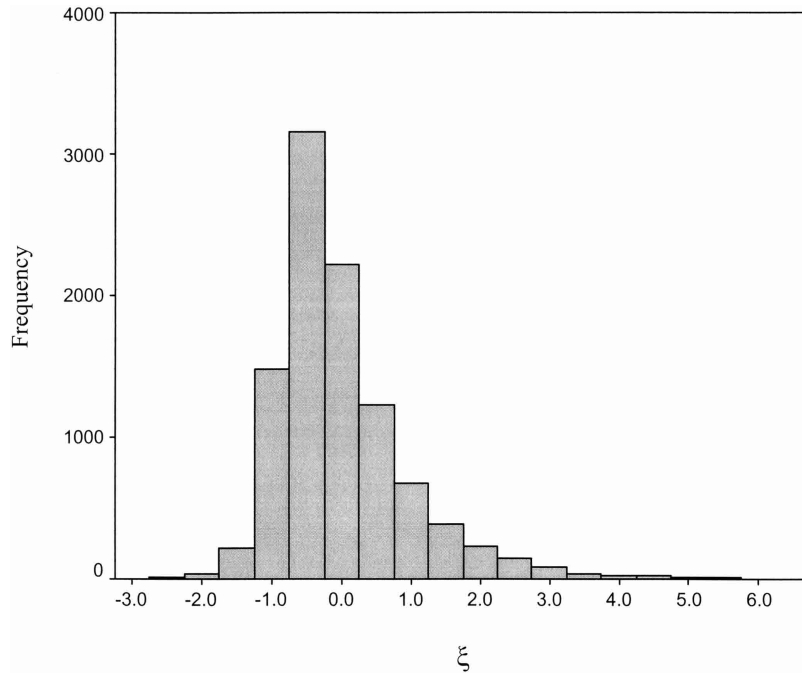


Figure 7. Frequency histogram displaying the distribution of the latent factors (ξ) in the confirmatory factor model with nonnormal latent factors ($N = 10,000$).

Table 3
Intercorrelations, Means, Standard Deviations, Skew, and Kurtosis of Data Generated From the Nonnormal Correlated Factors Model (N = 400)

Variable	1	2	3	4	5	6
1. x_1	—					
2. x_2	.589	—				
3. x_3	.604	.586	—			
4. x_4	.439	.469	.454	—		
5. x_5	.423	.450	.438	.641	—	
6. x_6	.443	.468	.443	.642	.688	—
<i>M</i>	.048	-.037	.059	-.060	.068	.012
<i>SD</i>	1.339	1.284	1.241	1.329	1.328	1.344
Skew	0.635	0.973	0.816	1.042	1.057	1.159
Kurtosis	0.902	2.799	1.690	2.281	3.453	3.085

eter values freely varied over classes. Thus, the latent class variable served to moderate the levels and relationships of the observed variables, consistent with the view that a homogeneous latent group is defined by its own distinct distribution. The change in log-likelihood between the one- and two-class models was large and significant; however, the improvement in log-likelihood between two and three classes was not. As Table 4 shows, the AIC, BIC, CAIC, CLC, and NEC all also supported the two-class model over the one-class model, though the more conservative ICL-BIC did not. The CLC and NEC, which do not penalize for overparameterization, both indicated the three-class model to be optimal, though only slightly better than the one- or two-class alternative models. Further inspection of the three-class model revealed an improper (negative) residual variance estimate. On the basis of the sum of these results, we concluded that the two-class model was optimal for the data.⁸ Given the proper specification of the within-class model, the improvement in fit associated with estimating two classes was due almost entirely to the

ability of the model to approximate the nonnormality of the data rather than to improved recovery of the covariance structure, and this was reflected in the stability of the SRMR.

Figure 8 further clarifies how the presence of the two latent classes allows the model to capture the nonnormality of the data. Figure 8, A and B, plots the distribution of ξ_1 that is implied by the standard CFA model and the two-class model, respectively, and is comparable to the histogram presented earlier in Figure 7. Figure 8A shows that the usual assumption of normality provides a poor representation of the actual distribution of the latent variable. In contrast, in Figure 8B, the weighted component distributions of the two latent classes, each with a different mean and variance, combine to imply a nonnormal aggregate distribution with a much greater resemblance to the histogram in Figure 7. In fact, the model-implied skew and kurtosis for the first factor were 1.13 and 2.02, respectively, whereas for the second factor, they

⁸ It is common to estimate mixtures of CFA models constraining the factor loading matrices to be invariant over classes so that the latent variables will have the same meaning in each class (e.g., Arminger & Stein, 1997; Arminger et al., 1999; Jedidi et al., 1997a, 1997b; Yung, 1997). Invariance constraints are sometimes also placed on the item intercepts and residual variances. Both sets of constraints were found tenable for the present model. The fit indices we examined changed predictably with the equality constraints in place. Those indices that balance fit with parsimony (e.g., AIC, BIC, CAIC) began to favor the three-class model, as far fewer parameters were required per class. In contrast, those indices that balance fit with classification quality (e.g., NEC, CLC) began to favor the one-class model, as the equality constraints served to make the classes less distinctive from one another. These results point to the inherent complexities associated with judging the fit of SEMMs.

Table 4
Fit Statistics of a Properly Specified Confirmatory Factor Model With One to Three Classes Applied to Nonnormally Distributed Data Generated From a Homogeneous Population

Model	LL	SRMR	AIC	BIC	CAIC	CLC	NEC	ICL-BIC
One class	-3,498.54	.013	7,035.08	7,110.92	7,129.92	6,997.08	1.00	7,110.92
Two classes	-3,405.12*	.012	6,888.24	7,043.90	7,082.90	6,993.23	0.98	7,226.89
Three classes	-3,388.11	.013	6,894.21	7,129.71	7,188.71	6,991.54	0.97	7,345.04

Note. For AIC, BIC, CAIC, CLC, NEC, and ICL-BIC, the minimum value obtained from the series of models is in boldface to indicate the model with optimal fit. LL = log-likelihood; SRMR = standardized root-mean-square residual; AIC = Akaike's information criterion; BIC = Bayes's information criterion; CAIC = consistent AIC; CLC = classification likelihood criterion; NEC = normalized entropy criterion; ICL-BIC = integrated completed likelihood criterion using BIC approximation. Asterisks attached to LL values indicate a significant improvement in log-likelihood over the preceding model as judged by the test of Lo, Mendell, and Rubin (2001).

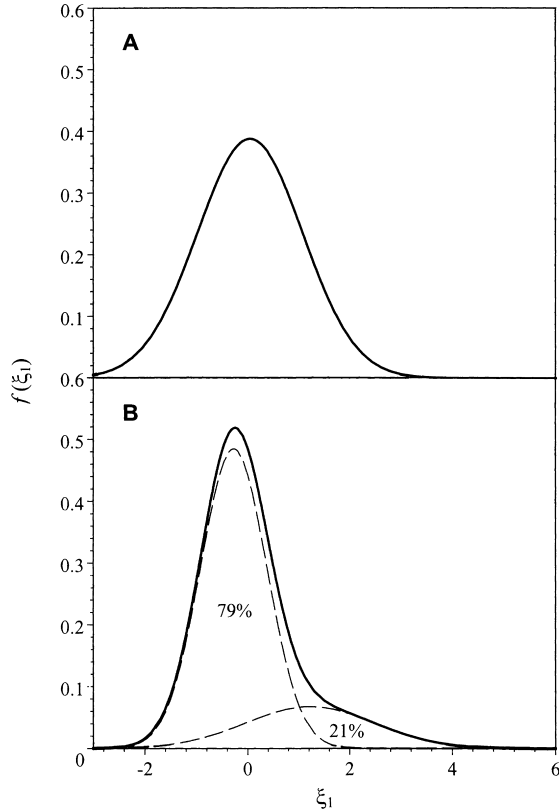


Figure 8. The distribution of the latent factor ξ_1 implied by a single-group confirmatory factor model (A) compared with a two-class model (B; the percentage of cases in each class is indicated).

were 1.81 and 3.91. Although the implied skew and kurtosis values of the two-class model underestimate the actual sample values (which in practice would not be known), they are clearly closer to the sample values than the values of zero assumed by the single-class model. In turn, the presence of the two classes better approximates the nonnormality of the observed variable distributions, and this leads to the superior fit of the two-class model. We thus see how latent classes can serve to capture nonnormality in the observed data, though only one group actually exists in the population.

This single artificial data example provides a simple illustration of the effect of nonnormality on latent class estimation, showing how the presence of latent classes can function to approximate a nonnormal multivariate distribution (i.e., capture the higher order moments) and thereby improve model fit, even when only one group truly exists in the population. If model fit is used as a guide for inferring the correct

number of classes, this may lead to the identification of classes that do not correspond to true groups in the population. This leads us to our second observation: Nonnormality can induce the estimation of latent classes even when the structural model is correctly specified and only one group exists in the population.

The Assumption of Linearity

In some ways, nonlinearity may be viewed as a special case of either model misspecification or non-normality. First, in the presence of nonlinearity, we may view a standard continuous latent variable model as misspecified. We may see the assumption of linearity as a misspecification of the functional form of the relationship, or we may view the misspecification as one of omitted variables (i.e., polynomial terms) that, if included, would serve to better capture the relationship (McDonald, 1967). It is rare, however, that polynomial terms are tested in latent variable models, in part because of the complexity this adds to the modeling process (Schumacker & Marcoulides, 1998). We also consider this case separately for the fundamental reason that, unlike the model misspecifications considered before, the population structure of a nonlinear model cannot be recovered by simply rearranging the linear relationships among the variables; at the least we must add new terms to the model to approximate the nonlinearities. Second, we may view nonlinearity as a special case of nonnormality. Multivariate normality implies linearity; hence, nonlinearity implies multivariate nonnormality. We consider nonlinearity separately, however, because, from the standpoint of the continuous latent variable model, it involves a simultaneous violation of both distributional assumptions and the functional form of the modeled relationships.

As Figure 3 illustrates, discrete latent variables can capture nonlinear relations among the observed measures, so we should expect the presence of (unmodeled) nonlinear relationships to induce the estimation of latent classes in an SEMM as well. To demonstrate this point, we generated a single sample of 500 cases of data conforming to the population structural equation model given in Figure 9.⁹ Three indicators (x_1-x_3) loaded on the exogenous factor ξ , and three indicators (y_1-y_3) loaded on the endogenous latent factor η . For simplicity (and without loss of generality), the

⁹ Consistent with Bollen (1995), we represent the nonlinear effect in the path model with a saw-toothed arrow.

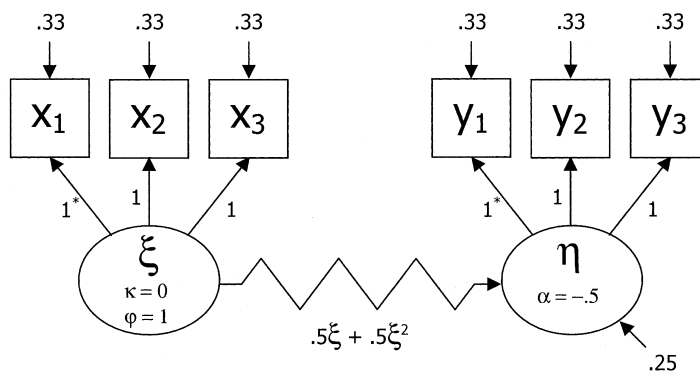


Figure 9. Path diagram displaying the population structural equation model used to generate data for the demonstration that nonlinearity may induce the estimation of latent classes. The saw-toothed arrow represents the nonlinear effect of latent factor ξ on latent factor η . In the estimated models the loadings with asterisks were fixed to scale the latent factors, and the effect of ξ on η was assumed to be linear. κ designates the mean of ξ , φ designates the variance of ξ , and α designates the intercept of η .

unconditional means of ξ and η were set to zero and the total variances of the factors were set to one. The indicators were assigned unit loadings and zero intercepts, and 75% of their variance was explained by the associated latent factor. The uniquenesses of the indicators, the factor ξ_i , and the disturbance of the endogenous factor (ζ_i) were independent and normally distributed. Most important, the relationship between the exogenous factor ξ and the endogenous factor η was curvilinear; specifically $\eta_i = -.5 + .5\xi_i + .5\xi_i^2 + \zeta_i$, where $\zeta_i \sim N(0, .25)$. This function implies that 75% of the variance in η_i is explained by ξ_i . If a linear approximation is used, only about 25% of the variance in η_i is explained by ξ_i . The sample statistics for the data are presented in Table 5. Note that the nonlinear relationship between ξ and η naturally induces skew and kurtosis in y_1 – y_3 .

Table 5
Intercorrelations, Means, Standard Deviations, Skew, and Kurtosis of Data Generated From the Nonlinear Two-Factor Structural Equation Model ($N = 500$)

Variable	1	2	3	4	5	6
1. x_1	—					
2. x_2	.751	—				
3. x_3	.770	.751	—			
4. y_1	.409	.375	.371	—		
5. y_2	.410	.382	.377	.698	—	
6. y_3	.419	.395	.397	.746	.736	—
M	−0.001	−0.007	0.033	0.075	0.109	0.093
SD	1.158	1.169	1.179	1.171	1.176	1.134
Skew	−0.002	−.131	−.173	.726	.820	.933
Kurtosis	−0.278	−0.245	−0.199	1.349	1.661	2.176

We first fit the data with a standard one-group structural equation model of the same form as in Figure 9 but with the key exception that the relation between ξ and η was modeled as linear. To scale the latent variables, we set the loadings of x_1 and y_1 to one and their intercepts to zero. All other parameters (including intercepts–means) were estimated. Using conventional criteria, we found that the overall fit of this model was quite good, despite failure to model the nonlinear relationship, $\chi^2(8) = 2.43$, $p = .97$; RMSEA = .00. This is not surprising, as conventional fit indices take no account of possible nonlinear relationships and the degree of nonnormality in the data was quite modest, so would not be expected to inflate the test statistic.

We then extended the model by estimating two and three latent classes. The form of each within-class structural model was identical to that of the model used in the prior analysis, but the parameter values were free to vary over classes. As in the prior example, the latent class variable thus served to moderate the parameter values within classes, consistent with the conceptualization of latent groups as being characterized by different distributions (i.e., mean levels and relationships between variables). The results of fitting these models are displayed in Table 6.¹⁰ A significant improvement in the log-likelihood

¹⁰ Although we present detailed results for these models only, we did again fit models with invariance constraints on the factor loadings and the item intercepts and residual variances. As with the previous nonnormal example, the AIC,

Table 6

Fit Statistics of a Linear Structural Equation Model With One to Three Classes Applied to Data Generated From a Homogeneous Population Model That Includes a Nonlinear Relationship Between Two Latent Variables

Model	LL	SRMR	AIC	BIC	CAIC	CLC	NEC	ICL-BIC
One class	-3,741.86	.009	7,521.72	7,601.80	7,620.80	7,483.72	1.00	7,601.80
Two classes	-3,623.92*	.005	7,325.83	7,490.20	7,529.20	7,578.46	1.40	7,820.83
Three classes	-3,580.29	.009	7,278.57	7,527.23	7,586.23	7,549.48	1.20	7,916.14

Note. For AIC, BIC, CAIC, CLC, NEC, and ICL-BIC, the minimum value obtained from the series of models is in boldface, to indicate the model with optimal fit. LL = log-likelihood; SRMR = standardized root-mean-square residual; AIC = Akaike's information criterion; BIC = Bayes's information criterion; CAIC = consistent AIC; CLC = classification likelihood criterion; NEC = normalized entropy criterion; ICL-BIC = integrated completed likelihood criterion using BIC approximation. Asterisks attached to LL values indicate a significant improvement in log-likelihood over the preceding model as judged by the test of Lo, Medell, and Rubin (2001).

of the model was obtained with the addition of the second latent class, but not with the addition of the third latent class. The AIC, BIC, and CAIC all also indicated that the two-class model was superior to the standard one-group structural equation model. Of these, only the AIC also supported the estimation of three classes. The three-class model, however, contained an improper (negative) residual variance estimate. As in the strictly nonnormal data example above, the SRMR showed high stability, again indicating that the improvement in fit associated with adding latent classes did not occur at the level of the covariance matrix (which captures only linear associations). None of the three criteria involving entropy penalties (i.e., the CLC, NEC, and ICL-BIC) supported more than one class, although in fact the mean-estimated probability of class membership of the individuals within each class was approximately .83 for the two-class solution and .80 for the three-class solution. From these results we concluded that two classes were optimal for the data but that the classes were likely not well separated.

To further clarify the results, in Figure 10 we present a graph of the 95% confidence ellipses from the two-class model overlaid on a scatter plot of the true scores for the two latent factors. Note that although the relationship between ξ and η is linear within classes, together the two ellipses provide an almost piecewise approximation of the true nonlinear relationship. The approximation is visually somewhat crude but is clearly better than what a single ellipse would allow (e.g., Figure 3A), and this is more formally reflected in the significantly better log-

likelihood and superior AIC, BIC, and CAIC of the two-class model. Figure 10 also reveals why the two-class model is not supported by the entropy-based criteria. Despite the fact that the relationship between the latent variables is significantly negative in one class and significantly positive in the other (i.e., moderated by the latent class variable), the classes are not considered well separated because the densest area of points is found in the middle of the graph where the

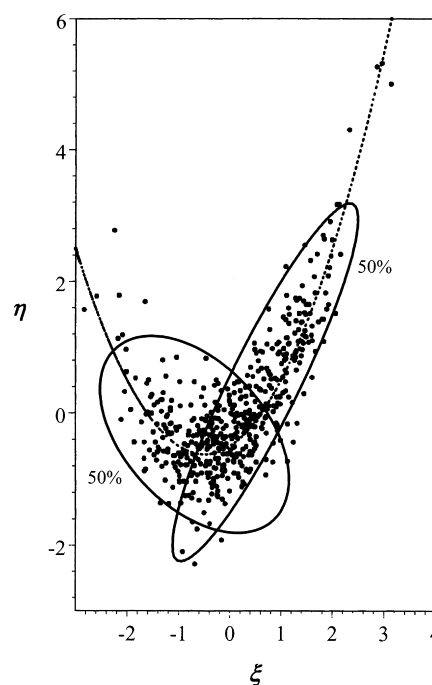


Figure 10. Scatter plot of the true scores for latent factors ξ and η . The dashed line plots the nonlinear effect of latent factor ξ on latent factor η in the population, and the 95% confidence ellipses (estimated from the sample) show how the nonlinear effect can be captured via the discrete latent class variable (the percentage of cases in each class is indicated).

BIC, and CAIC of the two- and three-class models all improved as more invariance constraints were imposed, but in this case the entropy-based CLC and NEC changed little.

two ellipses overlap. This specific pattern is a consequence of our choice of a normal distribution for ξ and is certainly not a necessary feature of nonlinear relationships. Hence, we should not expect entropy-based criteria to offer general protection against the estimation of spurious latent classes when linearity assumptions are violated.

In summary, the nonlinear relationship between ξ and η necessitated estimation of two latent classes to achieve optimal fit to the data (by most measures) even though the population consisted of just one group. This was true because the standard structural equation model, assuming linearity, could not fully capture the relationship between ξ and η . The nonlinearity of this relationship could, however, be approximated with the estimation of a second latent class. Our third and final observation is thus that unmodeled nonlinearity can induce the estimation of latent classes even when no latent groups are present.

Potential Problems

In this section we consider some of the potentially problematic implications of the observations explicated above, particularly when these models are used to make inferences about the presence of qualitatively distinct unobserved groups in the population. Specifically, there are several alternative and competing explanations for the estimation of multiple latent classes, including (a) that the structural model is misspecified, (b) that the data is nonnormal, and (c) that the relationships among the latent and/or observed variables are nonlinear (or some combination of the three). Each of these alternative explanations implies that a core assumption of traditional structural equation modeling has been violated and is compensated for by latent classes that do not reflect true groups in the population. We now discuss how the complications these issues raise can be mitigated, at least to some degree.

Model Specification

The fact that misspecification of the within-class structural model may induce the estimation of spurious latent classes is especially disconcerting if we consider that, in practice, we never know the true generating model. Indeed, Cudeck and Browne (1983) and MacCallum, Browne, and Sugawara (1996) have remarked that we may reasonably regard all structural equation models as misspecified because all models are approximations of reality (see also Meehl, 1967).

Given this, how can we have confidence that the latent classes in any given SEMM are not due simply to a poorly specified structural model?

At issue is the fundamental interdependency between the within-class and between-class components of the model that we identified in Equation 10. Although we focused on the possibility that misfit of the within-class model would induce the estimation of spurious latent classes, the interdependence of these two components of the model works both ways. As Meredith and Horn (2001) recently noted, failure to model heterogeneity due to the mixture of unobserved groups may lead to model misfit even if the within-class model structure is identical for all groups and properly specified, a phenomenon known more generally as *Simpson's paradox* (Pearl, 2000). In this case, the aggregate covariance matrix is augmented by apparent "common variance" due to the unmodeled between-class mean differences. Post hoc model modifications to include additional latent factors might then absorb this additional common variance and lead to good fit, though the added factors would be an entirely spurious consequence of the failure to model the true heterogeneity of the population.

This interdependency between the within-class and between-class components of the model was dealt with explicitly in the latent profile model through use of a clear decision rule for the partitioning of covariances—namely, that the observed covariances should be reproduced (that is, explained) exclusively by the between-class component of the model (the discrete latent variable). We propose consideration of a similar decision rule that would prevent estimating latent classes solely to compensate for misspecifications of the structural model. The rule we propose is to determine the optimal number of latent classes for an SEMM by using a model that is, by definition, not structurally misspecified: an unrestricted (or saturated) model.

Using this decision rule, the modeling process would consist of two steps. First, a series of unrestricted multivariate finite normal mixture models (or saturated SEMMs) would be estimated, corresponding to the model in Equation 6.¹¹ The number of latent

¹¹ If exogenous predictors were included in the model, an unrestricted conditional finite normal mixture model would be required instead. In this case, the unrestricted model would include all possible regression paths from the exogenous predictors to the endogenous variables, and the mix-

classes would then be determined from the comparative fit of these unrestricted models. Because no restrictions are imposed on the class means and covariances, we know a priori that the latent classes are not estimated solely to compensate for model misspecifications. In essence, Step 1 answers the question, How many covariance matrices and/or mean vectors are needed to adequately reproduce the observed data? Consistent with the conceptualization of a homogeneous latent group as being defined by a shared distribution, this question may be translated to mean, How many latent groups do the data suggest are present?

The second step then addresses the question, Assuming the number of latent classes identified in Step 1 is correct, does the hypothesized model structure adequately reproduce the within-class means and covariances? If we set the number of classes to be equal to the optimal number identified in Step 1, the hypothesized SEMM would then be estimated with the primary goal of recovering the class means and covariances from the unrestricted model. This corresponds to the usual objective of more traditional SEM analyses to recover the observed sample means and covariances. Of importance, just as in traditional structural equation modeling, we can formally test how well the SEMM in Step 2 succeeds in this task. As Yung (1997) pointed out, the unrestricted (or saturated) model chosen in Step 1 and the fitted model in Step 2 are formally nested, allowing for the computation of a standard chi-square difference test of overall model fit, much as is done in traditional multiple-groups structural equation modeling. In fact, on the basis of this result, Arminger and Stein (1997) advocated general use of this two-step approach to estimating SEMMs for the express purpose of testing the fit of the structural model (assuming the number of latent classes is correct). The current results suggest that the two-step approach may be equally important for avoiding the estimation of spurious latent classes due to misfit of the structural model.

Although theoretically compelling, this two-step strategy has several practical drawbacks. First, the unrestricted models estimated in Step 1 will generally use far more parameters than the eventual SEMM. As such, statistics such as the AIC, BIC, and ICL-BIC that penalize for the number of estimated parameters

may tend to be overly conservative (i.e., identify fewer latent classes than a correctly specified, or good-fitting, SEMM with fewer parameters). A related problem concerns estimation. Unrestricted multivariate finite normal mixture models are known to have very poorly behaved likelihood functions, especially when variance-covariance parameters are permitted to differ between classes (McLachlan & Peel, 2000, pp. 94–96). One advantage of SEMM is that the restrictions imposed in the within-class structural model may make the model more stable (Blåfield, 1980; Dolan & van der Maas, 1998; Jedidi et al., 1997b). Beginning SEMM analyses with an unrestricted model would then potentially counteract this important advantage. Given these complexities, it may be fruitful to consider alternative approaches to the same issue, such as the development and use of sensitivity indices that would signal latent class estimation due to misspecification of the structural model, or the use of residual diagnostics rather than model fit indices to identify the optimal number of latent classes (e.g., Lindsay & Roeder, 1992).

Nonnormality and Nonlinearity

The fact that latent classes may be estimated to accommodate nonnormality in the observed data or nonlinear relationships between variables is a less tractable problem. The reason is that nonnormality and/or nonlinearity in the aggregate data is a natural outcome of the mixture of unobserved groups, even when the data are normally distributed and the variables are linearly related within groups (Teicher, 1960; see Bauer & Curran, 2003a, for a more detailed discussion of this point). This would be the case if, for instance, the two class distributions in Figure 8B, or the two ellipses in Figure 10, corresponded to true groups. Of course, there are countless other reasons that nonnormality or nonlinearities could arise within a truly homogeneous population. Nonnormality could be a natural feature of the distribution of the phenomenon of interest (e.g., alcohol use in adolescence), could be a product of the sampling design (e.g., uneven selection from a normal population distribution), or could be due to measurement properties (e.g., floor or ceiling effects, analysis of counts or proportions, etc.), among other possibilities. Similarly, nonlinear bivariate and trivariate relationships are ubiquitous in psychological theory, such as the case in which two variables interact in the prediction of a third.

The two-step procedure advocated above would be ineffective in discriminating between these alternative

ture would be based on the residual covariance matrices and mean vectors of the latent classes (see Arminger & Stein, 1997, for further detail).

explanations, as it would incorrectly assume that normality and linearity hold within the latent groups. Another possibility is to apply additional consistency checks to the model. For instance, whereas some non-normal distributions may be consistent with a mixture of normals, others may not be. On the basis of this idea, Rindskopf (2003) noted that, from a Bayesian approach, one might consider applying posterior predictive model assessment techniques to gauge the ability of the model to reproduce the data. Similarly, B. Muthén (2003) formulated a test designed to assess the fit of an SEMM to the higher order moments of the observed data. If the implied multivariate skew and kurtosis of the SEMM do not correspond to the observed values, then this would be cause for rejection of the model. One interpretation of this result would be that the mixture hypothesis is incorrect and that the nonnormality in the data likely originates through some other mechanism than the pooling of unobserved groups. Another interpretation, however, is that too few latent classes have been estimated, and hence the approximation that the mixture distribution provides to the observed distribution is not yet adequate (Bauer & Curran, 2003b). Clearly, further research on this issue is needed.

Promising Opportunities

Up to this point, we have considered the chief goal of SEMM analyses to be the identification of distinctive subgroups within the population for whom different models may hold. This presentation is consistent with the introduction of SEMMs by Jedidi et al. (1997a, 1997b) and Arminger and colleagues (Armingier & Stein, 1997; Arminger, Stein, & Wittenberg, 1999) and was made most explicit by Dolan and van der Maas (1998), who made the following statement:

In so-called *direct* applications of finite mixture modeling, . . . the aim is to determine the number and the type of components in the mixture, to estimate the unknown parameters, and to assign the cases to their respective components. In *indirect* applications the finite mixture model is employed as a mathematical device, for example, to approximate intractable heavy-tailed distributions. In such applications, the underlying components do not necessarily have a physical interpretation. Here we are concerned with fitting multivariate normal finite mixtures in direct applications subject to structural equation modeling. . . . (pp. 227–228)

What we have shown thus far is that it is possible that the SEMM is really serving an indirect function when a direct interpretation is desired. However, it

may be argued that in empirical applications we can never really know which function the model is serving. From this perspective, attempting to recover true latent subgroups is a fool's errand and should not be the primary focus of the investigation. Rather, the merit of the model should be based on its ability to represent the data in a theoretically meaningful way (Cudeck & Henly, 2003). As we have expressed elsewhere (Bauer & Curran, 2003b), we do not fully share this view; however, we do feel that, in some cases, indirect applications of the SEMM can yield scientifically useful information. It is in this spirit that we now reinterpret our observations on SEMM analyses in terms of the opportunities they may provide to overcome some of the traditional limiting assumptions of standard continuous latent variable models, particularly when one is willing to relinquish direct interpretations of the latent classes as true population subgroups.

Model Specification

There may be some situations in which one would prefer to summarize an admittedly continuous latent variable using discrete latent classes. A salient example is the LCGA model. A one-class LCGA model holds that there is a single fixed trajectory in the sample and that there are no systematic individual differences around that trajectory over time. For most growth modeling applications this would be an unrealistic assumption, as there is often reason to believe that individuals differ both in overall level and rate of change over time (e.g., Curran, 2000; Curran, Stice, & Chassin, 1997). However, by imposing this assumption, we require additional latent classes in order to recover the observed covariances (as in the latent profile model). By requiring that the between-class portion of the model account for the relations among the repeated measures over time, we can summarize the possibly continuous distribution of individual trajectories using a small set of class mean trajectories.

The value of this approach was emphasized by Land and Nagin (1996) and Nagin (1999), who suggested that the class trajectories be interpreted as points of support for the continuous distribution of individual differences, analogous to the bars of a sample histogram. From this perspective, Figure 5B yields much important summary information. It correctly suggests that the individual trajectories are symmetrically distributed and that initial levels and rates of change are positively correlated. On the basis of these observations, we might feel more confident

estimating a standard one-group latent curve model, which assumes the individual intercepts and slopes are normally distributed and linearly related. Of course, it is possible that such a plot would reveal asymmetries in the distribution of the trajectories, or that the individual intercepts and slopes of the trajectories are nonlinearly related, calling into question the usefulness of a standard latent curve model and suggesting the need for a semiparametric or nonparametric analysis (Nagin, 1999; Vermunt & van Dijk, 2001). In our opinion, it is under conditions such as these that SEMMs offer the most exciting (and potentially controversial) new modeling opportunities.¹²

Nonnormality and Nonlinearity

Traditional structural equation models (and arguably all techniques under the umbrella of the general linear model) are often applied under the tacit assumptions that the data are multivariate normal and that the variables of the model are linearly related. As we have shown, SEMMs do not make these assumptions. Through the inclusion of latent classes, SEMMs may accommodate both nonnormally distributed and nonlinearly related continuous variables. Given this (and assuming sufficient power), if only one latent class is needed to optimally summarize the data (and here we assume a properly specified or unrestricted model), then it is likely that the data are approximately normally distributed and that the relationships among variables are close to linear. Alternatively, if more than one class is required (again assuming a properly specified or unrestricted model), then these assumptions of traditional structural equation modeling must be rejected.

We are not proposing the use of SEMM as simply another tool for testing the assumptions of traditional structural equation modeling. Other tools for this task are readily available (e.g., Mardia, 1985). Further, techniques exist aside from SEMM for explicitly accommodating violation of these assumptions. For instance, Browne (1984), Satorra and Bentler (1994), and Bollen (1996) have discussed robust or distribution-free methods for estimating structural equation models in the presence of nonnormal data, and Schumacker and Marcoulides (1998) presented several approaches to modeling nonlinear relationships in structural equation modeling. However, we submit that SEMM offers new and unique opportunities to more fully explore these issues. For instance, current methods for estimating structural equation models with nonnormal outcomes typically only provide a

means of obviating or correcting for the problem of nonnormality. SEMM instead capitalizes on the additional information available with nonnormal outcomes and provides a means to explicitly map out the distributions of the latent variables (as in Figure 8B). Multimodality of these distributions, although not a necessary requirement of a normal mixture, would strongly suggest the presence of more than one population (McDonald, 1967, p. 60; Waller & Meehl, 1998, chap. 5). Similarly, the techniques reviewed in Schumacker and Marcoulides for modeling nonlinear relationships among latent variables are currently cumbersome to implement and require strong theory about the functional form of the relationships. In contrast, SEMM can be used in a more exploratory mode. For instance, plotting the results of the SEMM (as in Figure 10) may suggest the presence of a nonlinear relationship between the continuous latent variables. In both of these cases, SEMM yields new information on the phenomenon of interest that could not be obtained easily from a traditional one-group structural equation model.

Recommendations

The observations that we have explored in this article have important implications for the practical use of SEMMs in applied research. We conclude by drawing these implications together into an explicit list of recommendations for using SEMMs in practice. Of importance, these recommendations should be regarded as tentative and subject to debate and later revision in light of additional research on these models.

Recommendation 1: Determine what function the latent classes are intended to serve. By this we mean that applied researchers should decide a priori whether a direct interpretation of the latent classes, in terms of qualitatively distinct latent subgroups, is desired or if the latent classes are to serve one of the indirect functions noted above. In indirect applications, we personally believe inferences should be reserved for the aggregate population (rather than within-class parameter estimates), though others may disagree. In direct applications, the latent classes are assumed to represent

¹² The potential controversy over these uses of the model was reflected in two reviews of this article with equally strong but opposing reactions (with one reviewer feeling we had gone too far and the other feeling we had not gone far enough).

true groups in the population, a stronger inference requiring a greater burden of proof. Our remaining recommendations are given primarily for direct applications of the model.

Recommendation 2: Attempt to ascertain what function the latent classes are actually serving. To avoid estimating latent classes to accommodate misspecifications of the structural model, consider using the two-step procedure outlined above to determine the optimal number of latent classes. After selecting the optimal number of latent classes, plot the model-implied distributions of the latent and/or observed variables (as in Figures 8 and 10) to better understand how the latent classes are capturing nonnormality or nonlinearity in the data and apply formal tests to assess the fit between the model and data. Consider whether, aside from the mixture of unobserved groups, there are other possible sources contributing to the nonnormality or nonlinearity of the measured variables. Attempt to rule out these alternative sources. For instance, examine univariate distributions to see if they exhibit floor or ceiling effects. Evaluate the tenability of the assumption that normality represents homogeneity (e.g., infrequent count variables would be expected to show nonnormality even in the absence of a mixture, so a finite normal mixture would be inappropriate). Also consider the nature of the sampling design and whether it may have distorted the shape of the sample distribution relative to the population distribution (e.g., individuals self-selecting into the study with probability related to their level on the construct of interest).

Recommendation 3: Embark on a program of construct validation. Accumulate evidence in favor of the hypothesis that the population indeed consists of a mixture of unobserved latent groups. Are the characteristics of the latent groups consistent with a priori theoretical predictions? Are the latent groups distinguished in multiple samples and/or by multiple measures? Can they be distinguished by differential relations to precursor variables, or do they show distinctly different sequelae in subsequent assessments that are consistent with the motivating theoretical model? See Bauer and Curran (2003b) for further discussion on the problem of construct validation in SEMMs, with a particular emphasis on growth mixture models.

Recommendation 4: Most important, take care in drawing conclusions from the fit of a single model. As Recommendation 3 implies, the construct validation of the latent classes will often require a programmatic series of research studies. Determining that the opti-

mal model for a given sample includes several latent classes on the basis of model fit statistics is not sufficient evidence for concluding that the population is a heterogeneous mixture of latent groups. The latent classes may be serving an indirect function even when a direct interpretation is desired; thus, the conclusions drawn from a single model should be guarded. Alternative hypotheses for the latent classes should be noted and, where possible, evaluated.

Concluding Remarks

We have attempted to elucidate some of the implications that may arise from integrating continuous and discrete latent variable models. Although these models have historically been viewed as alternative ways of representing individual differences, the more recent SEMM allows for the simultaneous estimation of both continuous and discrete latent variables. It is the inherent flexibility of the SEMM that leads both to potential problems and to promising new opportunities. This flexibility allows for violation of assumptions in one part of the model (i.e., the structural model) to be accommodated by changes to the other (i.e., the latent class model). This is problematic if the goal of the analysis is the identification of qualitatively distinct groups, as it suggests that the latent classes may instead serve simply to accommodate assumption violations of the structural model. Alternatively, we may eschew conclusions about the true structure of individual differences (i.e., whether latent groups exist in the population or not) and knowingly capitalize on the flexibility of the model to capture phenomena that are not readily tractable otherwise. Using the models in this way avoids reification of the latent classes yet retains the ability to explore complex features of the data in ways that up to now have not been easily accessible.

References

- Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors. *Sociological Methods and Research, 26*, 148–182.
- Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika, 64*, 475–494.
- Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. London: Griffin.
- Bauer, D. J., & Curran P. J. (2003a). Distributional assumptions of growth mixture models: Implications for overex-

- traction of latent trajectory classes. *Psychological Methods*, 8, 338–363.
- Bauer, D. J., & Curran, P. J. (2003b). Overextraction of latent trajectory classes: Much ado about nothing? Reply to Rindskopf (2003), Muthén (2003), and Cudeck and Henly (2003). *Psychological Methods*, 8, 384–383.
- Biernacki, C., Celeux, G., & Govaert, G. (1999). An improvement of the NEC criterion for assessing the number of clusters in a mixture model. *Pattern Recognition Letters*, 20, 267–272.
- Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 719–725.
- Blåfield, E. (1980). *Jyvaskyla studies in computer science, economics & statistics 2: Clustering of observations from finite mixtures with structural information*. Jyvaskyla, Finland: Jyvaskyla University.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. (1995). Structural equation models that are nonlinear in latent variables: A least-squares estimator. In P. M. Marsden (Ed.), *Sociological methodology* (pp. 223–252). Oxford, England: Blackwell.
- Bollen, K. A. (1996). An alternative two stage least squares (2SLS) estimator for latent variable equations. *Psychometrika*, 61, 109–121.
- Browne, M. W. (1984). Asymptotic distribution free methods in analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, 37, 62–83.
- Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, 13, 195–212.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 141–167.
- Cudeck, R., & Henly, S. J. (2003). A realistic perspective on pattern representation in growth data: Comment on Bauer and Curran (2003). *Psychological Methods*, 8, 378–383.
- Curran, P. J. (2000). A latent curve framework for studying developmental trajectories of adolescent substance use. In J. Rose, L. Chassin, C. Presson, & J. Sherman (Eds.), *Multivariate applications in substance use research* (pp. 1–42). Hillsdale, NJ: Erlbaum.
- Curran, P. J., Stice, E., & Chassin, L. (1997). The relation between adolescent and peer alcohol use: A longitudinal random coefficients model. *Journal of Consulting and Clinical Psychology*, 65, 130–140.
- Dolan, C. V., & van der Maas, H. L. J. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, 63, 227–253.
- Escobar, M. D., & West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90, 577–588.
- Everitt, B. S., & Hand, D. J. (1981). *Finite mixture distributions*. London: Chapman & Hall.
- Ferguson, T. S. (1983). Bayesian density estimation via mixtures of normal distributions. In M. H. Rizvi, J. S. Rustagi, & D. Siegmund (Eds.), *Recent advances in statistics* (pp. 287–302). New York: Academic Press.
- Galton, F. R. S. (1888). Co-relations and their measurement, chiefly from anthropometric data. *Proceedings of the Royal Society of London*, 45, 135–145.
- Gibson, W. A. (1959). Three multivariate models: Factor analysis, latent structure analysis, and latent profile analysis. *Psychometrika*, 24, 229–252.
- Hu, L., & Bentler, P. M. (1999). Cutoff criterion for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997a). Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39–59.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification*, 14, 23–50.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Academic Press.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Land, K. C., & Nagin, D. S. (1996). Micro-models of criminal careers: A synthesis of the criminal careers and life course approaches via semiparametric mixed Poisson models with empirical applications. *Journal of Quantitative Criminology*, 12, 163–191.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Lindsay, B. G., & Roeder, K. (1992). Residual diagnostics for mixture models. *Journal of the American Statistical Association*, 87, 785–794.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767–778.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size

- for covariance structure modeling. *Psychological Methods*, *1*, 130–149.
- Mardia, K. V. (1985). Mardia's test of multinormality. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 5, pp. 217–221). New York: Wiley.
- McDonald, R. P. (1967). *Nonlinear factor analysis* (Psychometric Monograph No. 15). Richmond, VA: Byrd Press.
- McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meehl, P. E. (1965). *Detecting latent clinical taxa by fallible quantitative indicators lacking an accepted criterion* (Report No. PR-65-2). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.
- Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, *34*, 103–115.
- Meehl, P. E. (1968). *Detecting latent clinical taxa: II. A simplified procedure, some additional hitmax cut locators, a single-indicator method, and miscellaneous theorems* (Report No. PR-68-4). Minneapolis: University of Minnesota, Research Laboratories of the Department of Psychiatry.
- Meehl, P. E. (1995). Bootstrap taxometrics: Solving the classification problem in psychopathology. *American Psychologist*, *50*, 266–275.
- Meehl, P. E., & Golden, R. R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–182). New York: Wiley.
- Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
- Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.
- Molenaar, P. C. M., & von Eye, A. (1994). On the arbitrary nature of latent variables. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 226–242). Thousand Oaks, CA: Sage.
- Muthén, B. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 369–377.
- Muthén, B. O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.
- Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81–117.
- Muthén, B. O., & Muthén, L. K. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research*, *24*, 882–891.
- Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.
- Muthén, L. K., & Muthén, B. O. (1998). Mplus (Version 2) [Computer software and manual]. Los Angeles: Muthen & Muthen.
- Nagin, D. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, *4*, 139–157.
- Nagin, D., & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: Specification and estimation of a nonparametric, mixed Poisson model. *Criminology*, *31*, 327–362.
- Pearl, J. (2000). *Causality*. Cambridge, England: Cambridge University Press.
- Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London, A*, *185*, 71–110.
- Ramaswamy, V., DeSarbo, W. S., Reibstein, D. J., & Robinson, W. T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, *12*, 103–124.
- Rindskopf, D. (2003). Mixture or homogeneous? Comment on Bauer and Curran (2003). *Psychological Methods*, *8*, 364–368.
- Roeder, K., & Wasserman, L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, *92*, 894–902.
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Schumacker, R. E., & Marcoulides, G. A. (1998). *Interaction and nonlinear effects in structural equation modeling*. Mahwah, NJ: Erlbaum.
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.
- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: Harvard University Press.

Teicher, H. (1960). On the mixture of distributions. *Annals of Mathematical Statistics*, 31, 55-73.

Thurstone, L. L. (1935). *The vectors of the mind*. Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: University of Chicago Press.

Titterton, D. M., Smith, A. F. M., & Makov, U. E. (1985). *Statistical analysis of finite mixture distributions*. Chichester, England: Wiley.

Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate nonnormal distributions. *Psychometrika*, 48, 465-471.

Verbeke, G., & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91, 217-221.

Vermunt, J. K., & van Dijk, L. A. (2001). A nonparametric random coefficient approach: The latent class regression model. *Multilevel Modelling Newsletter*, 13, 6-13.

Waller, N. G., & Meehl, P. E. (1998). *Multivariate taxometric procedures: Distinguishing types from continua*. Thousand Oaks, CA: Sage.

Wherry, R. J., & Gaylord, R. H. (1944). Factor pattern of test items and tests as a function of the correlation coefficient: Content, difficulty, and constant error factors. *Psychometrika*, 9, 237-244.

White, H. R., Johnson, V., & Buyske, S. (2000). Parental modeling and parental behavior effects on offspring alcohol and cigarette use: A growth curve analysis. *Journal of Substance Abuse*, 12, 287-310.

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

Wolfe, J. H. (1971). *A Monte Carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions* (Tech. Bulletin No. STB 72-2). San Diego, CA: U.S. Naval Personnel and Training Research Laboratory.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297-330.

Zhu, H.-T., & Lee, S.-Y. (2001). A Bayesian analysis of finite mixtures in the LISREL model. *Psychometrika*, 66, 133-152.

Received June 13, 2002

Revision received May 21, 2003

Accepted July 21, 2003 ■

ORDER FORM

Start my 2004 subscription to *Psychological Methods!*

ISSN: 1082-989X

\$45.00, APA MEMBER/AFFILIATE _____
 \$70.00, INDIVIDUAL NONMEMBER _____
 \$187.00, INSTITUTION _____
In DC add 5.75% / In MD add 5% sales tax
TOTAL AMOUNT ENCLOSED \$ _____

Subscription orders must be prepaid. (Subscriptions are on a calendar year basis only.) Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO:

American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Or call (800) 374-2721, fax (202) 336-5568.
TDD/TTY (202) 336-6123.

For subscription information, e-mail:
subscriptions@apa.org

Send me a FREE Sample Issue

Check enclosed (make payable to APA)

Charge my: VISA MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

BILLING ADDRESS: _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

SHIP TO:

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____ METAI4