Multivariate Behavioral Research, 49:614–632, 2014

ISSN: 0027-3171 print / 1532-7906 online
DOI: 10.1080/00273171.2014.950719

Handling Missing Covariates in Conditional Mixture Models Under Missing at Random Assumptions

Sonya K. Sterba
Department of Psychology and Human Development, Vanderbilt University

Mixture modeling is a popular method that accounts for unobserved population heterogeneity using multiple latent classes that differ in response patterns. Psychologists use conditional mixture models to incorporate covariates into between-class and/or within-class regressions. Although psychologists often have missing covariate data, conditional mixtures are currently fit with a conditional likelihood, treating covariates as fixed and fully observed. Under this exogenous-*x* approach, missing covariates are handled primarily via listwise deletion. This sacrifices efficiency and does not allow missingness to depend on observed outcomes. Here we describe a modified joint likelihood approach that (a) allows inference about parameters of the exogenous-*x* conditional mixture even with nonnormal covariates, unlike a conventional multivariate mixture; (b) retains all cases under missing at random assumptions; (c) yields lower bias and higher efficiency than the exogenous-*x* approach under a variety of conditions with missing covariates; and (d) is straightforward to implement in available commercial software. The proposed approach is illustrated with an empirical analysis predicting membership in latent classes of conduct problems. Recommendations for practice are discussed.

Mixture models are frequently applied in the social and behavioral sciences. Recent reviews indicate hundreds of mixture modeling applications by psychologists in the last decade alone (e.g., Fontaine, Carbonneau, Vitaro, Barker, & Tremblay, 2009; Sterba, Baldasaro, & Bauer, 2012). Mixture models accommodate population heterogeneity in the form of discrete, unobserved groups or latent classes. A particular model is specified to hold within each class, and parameters may differ across classes. Beyond selecting a best-fitting number of classes, primary goals of mixture modeling studies involve incorporating covariates into between- and/or within-class regressions to explain between- and/or within-class variation. When exogenous covariates are included in such a manner, these models have been called *conditional mixture models.* Popular examples of conditional mixtures include mixtures of regression models, conditional latent class models, conditional mixtures of structural equation models, and conditional mixtures of growth trajectories (e.g., Collins & Lanza, 2010; Dayton & Macready, 1988; DeSarbo & Cron, 1988; Dolan, 2009; Jedidi, Ramaswamy, DeSarbo, & Wedel, 1996; Lubke & Muthén, 2005; Muthén, 2002; Nagin, 2005; Vermunt, Tran, & Magidson, 2008; Wedel, 2002).

Currently, conditional mixture models are fit using a conditional likelihood specification, treating *x*'s as fixed.[1] This approach is here termed an *exogenous-x approach* and is the default in mixture modeling software used by psychologists (Jones, Nagin, & Roeder, 2001; Lanza, Dziak, Huang, Xu, & Collins, 2011; Muthén & Muthén, 1998–2014; Vermunt & Magidson, 2005). Conditional *non*mixture models are also typically fit using an exogenous-*x* approach. Examples of conditional nonmixture models include single-level and multilevel regression models (Draper & Smith, 1998; Raudenbush & Bryk, 2002), generalized linear models (McCullagh & Nelder, 1989), conditional structural equation models (e.g., Muthén & Asparouhov, 2009), and conditional growth models (e.g., Bollen & Curran, 2006).

A defining characteristic of conditional (mixture or nonmixture) models is that they allow inferences about a conditional distribution of outcomes given covariates, and these inferences are unaffected by the shape of the covariate distribution (e.g., DeSarbo & Cron, 1988; Dolan, 2009; Jedidi et al., 1996; Lubke & Muthén, 2005; Muthén &

Correspondence concerning this article should be addressed to Sonya K. Sterba, Quantitative Methods Program, Department of Psychology and Human Development, Vanderbilt University, Peabody #552, 230 Appleton Place, Nashville, TN 37203. E-mail: Sonya.Sterba@Vanderbilt.edu

[1]Throughout we use *x* as shorthand for "covariate."

Shedden, 1999; Nagin, 2005; Vermunt, 2010b). In the mixture modeling context, this defining characteristic requires that conditional mixture models not allow nonnormality of covariates to determine class structure. Researchers consider this characteristic of conditional mixtures substantively desirable; it allows the use of external covariates to help assess the construct validity of a typology where, crucially, the typology is *not* defined based on distributions of the covariates themselves (see Bakk, Oberski, & Vermunt, 2014; Bauer & Curran, 2003; Collins & Lanza, 2010; Dayton & Macready, 1988; Muthén, 2003; Odgers et al. 2008; Petras & Masyn, 2010).[2]

One problem with which psychologists often are confronted when fitting conditional (mixture or nonmixture) models is missing covariate data. Because the exogenous-*x* approach assumes covariates are fully observed, cases with missing covariates are by default listwise deleted. This sacrifices efficiency and can induce bias if, for instance, missingness depends on observed outcomes (Little, 1992; Little & Zhang, 2011). For psychologists fitting conditional mixture models with missing covariates, listwise deletion is currently the status quo (e.g., Castelao & Kroner-Herwig, 2012; Chung, Flaherty, & Schafer, 2006; Greenbaum, Del Boca, Darkes, Wang, & Goldman, 2005; Hepworth, Law, Lawlor, & McKinney, 2010; Li & Hser, 2011; Lubke & Muthén, 2005; Petras & Masyn, 2010; Schaeffer et al., 2006; Scharoun-Lee et al., 2011; Sterba, Prinstein, & Cox, 2007; Van Horn, Jaki, Masyn, Ramey, & Smith, 2009; Walrath et al., 2004). For psychologists fitting conditional *non*mixture (e.g., multiple regression) models, two main alternatives to listwise deletion, described later, are used to retain cases with missing covariates under missing-at-random (MAR) assumptions. Unfortunately, there are obstacles to applying either alternative in the mixture context.

One alternative used in the nonmixture context is multiple imputation. However, for mixture models, prior research has shown that multiple imputation of covariates can result in considerable parameter bias, leading to recommendations that it should be "avoided altogether" (Enders & Gottschall, 2011, p. 50). This is because the number of latent classes, and thus the nature of class-varying predictor relations, is not known a priori. Hence, their correct specification in an imputation model in empirical practice would be highly unlikely (Enders & Gottschall, 2011).

Another alternative used in the nonmixture context is to employ a multivariate (joint likelihood) approach with random regressors in order to recover parameters of a conditional model. In this multivariate model, parameters of the conditional model of interest are estimated simultaneously with parameters of the covariates' marginal distribution, thus avoiding the need to listwise delete cases with missing covariates (e.g., Horton & Kleinman, 2007; Horton & Laird, 1998; Hox & Roberts, 2010; Ibrahim, 1990; Jöreskog & Sörbom, 1996; Lipsitz & Ibrahim, 1996; Little, 1992). Under assumptions that (a) parameters of the covariates' marginal distribution don't involve parameters of the conditional model and (b) residuals are independent of covariates (see Bollen, 1989; Johnston, 1984, pp. 281–285; Jöreskog, 1973, Jöreskog & Goldberger, 1975), multivariate nonmixture models recover parameters of conditional nonmixture models regardless of the shape of the covariate distribution. However, for mixture models, prior research with complete case data has shown that conventional multivariate mixtures do *not* recover parameters of a conditional mixture regardless of the shape of the covariate distribution (Arminger & Stein, 1997; Arminger, Stein, & Wittenberg, 1999). In fact, recovery of conditional mixture parameters was so poor using a conventional multivariate mixture with nonnormal covariates that Arminger and Stein (1997) exclusively recommended an exogenous-*x* approach for drawing inferences about conditional mixture parameters. They stated,

> The conditional modeling [here termed exogenous-*x*] approach is crucial to obtain consistent parameter estimates in this stage. One might consider an unconditional modeling [here termed conventional multivariate] approach in which the regressor variables $\mathbf{x}_i$ are stacked below the dependent variable $\mathbf{y}_i$ . . . however this approach works only if the regressor variables $\mathbf{x}_i$ are also multivariate normal. If the regressor variables are nonnormal, as is the case with dummy variables such as gender and occupation, then the estimates of the mixing probabilities and other parameters will be inconsistent. (p. 157)

In sum, alternatives to listwise deletion that are well known for conditional nonmixture (e.g., multiple regression) models encounter obstacles when applied to conditional mixtures with missing covariates. The objective of this article is to provide psychologists with an approach that (a) recovers parameters of a conditional mixture model and (b) accommodates missing covariates under MAR assumptions. This article proposes and evaluates a specification here termed an *endogenous-constrained-x* approach that fulfills this objective.

[2]For instance, Collins and Lanza (2010, Chapter 6) fit a conditional mixture with five outcomes serving as indicators of a discrete latent classification variable: eats green vegetables, gets more than 7 hr of sleep, eats breakfast, exercises vigorously, and eats fruit. The discrete latent construct was interpreted as a healthy behavior typology. They included two covariates predicting between-class differences (maternal education and gender). Collins and Lanza wanted to investigate whether the covariates predicted between-class differences in theoretically consistent ways—but did *not* want the covariate distributions to help define the discrete latent construct itself (this construct would then no longer be interpretable as a healthy behavior typology). Similarly, Nagin (2005, Chapter 7) fit a conditional mixture with repeated measures of delinquency serving as indicators of class. He included covariates predicting delinquency within class (time, grade retention, and IQ). Nagin fit a conditional mixture because he wanted to interpret the discrete latent classification variable as a typology of growth trajectories defined only by the conditional distribution of delinquency repeated measures, not also by the distributions of IQ, time, and grade retention.

The remainder of the article proceeds as follows. First, we provide some background details on conditional mixture models. Second, we contrast three likelihood specifications: the exogenous-x approach currently used for fitting conditional mixtures, a conventional multivariate mixture, and the proposed endogenous-constrained-x approach, all under complete-x data. We explain theoretically why the proposed approach should recover parameters of the conditional model just as well as the exogenous-x approach under complete-x data—something the conventional multivariate mixture cannot be counted on to do. In Study 1 we demonstrate these points via simulation, under complete-x data. In Study 2, we employ missing-x data and compare the exogenous-x and endogenous-constrained-x approaches across a variety of theoretically relevant simulated conditions. Finally, an empirical illustration involving prediction of membership in conduct problem classes shows how the simulation results generalize to other distributions of outcomes and covariates. We conclude with implications for practice.

BACKGROUND: CONDITIONAL MIXTURE MODELS

As background, we review different ways that covariate regressions can be incorporated into conditional mixtures. One option is to regress class membership on covariates to explain why certain persons are more likely to be members of one class than another (e.g., Collins & Lanza, 2010; Dayton & Macready, 1988; Lubke & Muthén, 2005; Muthén, 2002; Nagin, 2005; Vermunt et al., 2008; Wedel, 2002).[3] Let i index person, where $i = 1,\ldots,N$. $\mathbf{y}_i$ is a vector of outcomes for person i, and $\mathbf{x}_i$ is a vector of covariates for person i. Persons' data are assumed to be generated from a mixture of K latent classes, where classes are indexed $k = 1,\ldots,K$. Let c_i denote person i's unobserved class membership. Let $p(c_i = k)$ denote the probability that person i is a member of class k. Then, via a multinomial logistic regression specification, we allow $\mathbf{x}_i$ to predict class membership:

$$p(c_i = k|\mathbf{x}_i) = \exp(\alpha^{(k)} + \boldsymbol{\omega}'^{(k)}\mathbf{x}_i)/\sum_{k=1}^{K}\exp(\alpha^{(k)} + \boldsymbol{\omega}'^{(k)}\mathbf{x}_i). \tag{1}$$

In class k, $\alpha^{(k)}$ is a multinomial intercept and $\boldsymbol{\omega}^{(k)}$ is a vector of multinomial slopes. For identification, $\alpha^{(K)} = 0$ and $\boldsymbol{\omega}^{(K)} = \mathbf{0}$.

Alternatively, or additionally, we can allow outcomes $\mathbf{y}_i$ to be regressed on $\mathbf{x}_i$ within class (e.g., DeSarbo & Cron, 1988; Hosmer, 1974; Jansen, 1993; Jedidi et al., 1996; Lubke & Muthén, 2005; B. O. Muthén, 2002; Nagin, 2005; Vermunt et al., 2008; Wedel & DeSarbo, 1995). The within-class density is denoted $f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)$. As an example, in class k we could have

$$\mathbf{y}_i|\mathbf{x}_i, c_i = k \sim N(\boldsymbol{\mu}^{(k)} + \boldsymbol{\Gamma}^{(k)}\mathbf{x}_i, \sigma^{2(k)}\mathbf{I}) \tag{2}$$

In class k, $\boldsymbol{\mu}^{(k)}$ is a vector of intercepts and $\boldsymbol{\Gamma}^{(k)}$ is a matrix of regression slopes. $\sigma^{2(k)}$ is a residual variance (here held equal across outcomes, although this is not necessary).[4] If there were outcomes other than $\mathbf{y}_i$ within class, such as latent factors, then they could also or alternatively be regressed on $\mathbf{x}_i$ (e.g., Lubke & Muthén, 2005), although this is not explicitly represented here.

The next sections cover three approaches for constructing a likelihood to fit a mixture model with covariates. Reasons are given for why two of these three approaches can recover parameters of a conditional mixture model, regardless of covariate distribution shape, and why one approach cannot. These approaches are introduced in the context of complete x's.

FITTING CONDITIONAL MIXTURES WITH A CONDITIONAL LIKELIHOOD: EXOGENOUS-*X* APPROACH

The exogenous-x approach is the currently used gold standard for fitting conditional mixtures (see, e.g., Bauer, 2007; Collins & Lanza, 2010; Dolan, 2009; Feldman, Masyn, & Conger, 2009; Lubke & Muthén, 2005; Muthén, 2002; Nagin, 2005; Pickles & Croudace, 2010; Reinecke, 2006; Vermunt, 2010a; Wedel & DeSarbo, 1995). Under this approach, the sample complete-case conditional likelihood can be constructed from the following (where, for simplicity, parameters are not shown):

$$L = f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} f(\mathbf{y}_i|\mathbf{x}_i) \tag{3}$$

When covariates are incorporated as in Equation (1) we would have

$$f(\mathbf{y}_i|\mathbf{x}_i) = \sum_{k=1}^{K} p(c_i = k|\mathbf{x}_i) f(\mathbf{y}_i|c_i = k). \tag{4}$$

Or, when covariates are incorporated as in Equation (2) we would have

$$f(\mathbf{y}_i|\mathbf{x}_i) = \sum_{k=1}^{K} p(c_i = k) f(\mathbf{y}_i|\mathbf{x}_i; c_i = k). \tag{5}$$

[3] This is sometimes termed a "concomitant variable mixture model" (e.g., Dayton & Macready, 1988).

[4] In cross-sectional applications, usually with a single y_i, this is sometimes called a "regression mixture model." In longitudinal applications where $\mathbf{y}_i$ are repeated measures and $\mathbf{x}_i$ also contain time scores, this is sometimes called a conditional "groups-based trajectory model."

When covariates are incorporated as in both Equations (1) and (2) we would have

$$f(\mathbf{y}_i|\mathbf{x}_i) = \sum_{k=1}^{K} p(c_i = k|\mathbf{x}_i) f(\mathbf{y}_i|\mathbf{x}_i; c_i = k). \tag{6}$$

In Equations (3)–(6), $\mathbf{x}_i$ are fixed and assumed fully observed. In this specification, distributional assumptions are imposed about outcomes *conditional on* covariates. For instance, in Equation (5) multivariate normality could be assumed within class for $\mathbf{y}_i|\mathbf{x}_i$. Note that no distributional assumptions are imposed on the covariates, so cases with missing covariates would need to be listwise deleted. Using Equation (4), (5), or (6) the conditional likelihood for person *i* is obtained as a weighted sum of class-specific densities. Figure 1 Panels A, B, and C provide heuristic diagrams of the Equations (4), (5), and (6) specifications, respectively, for a single outcome y_i and single covariate x_i. Squares represent measured variables and a circle represents the discrete latent classification variable. Directed arrows extending from x_i to c_i or x_i to y_i represent regression relationships (e.g., logistic or linear regression, depending on the nature of the outcome). A directed arrow extending from the circle to a square indicates that some parameter(s) associated with that measured variable are allowed to differ across class.

FITTING CONVENTIONAL MULTIVARIATE MIXTURES WITH A JOINT LIKELIHOOD

Conditional mixtures developed separately from multivariate mixtures in the mixture literature. Conventional multivariate mixtures subject to structural relations[5] were implemented by Blåfield (1980); Jedidi, Jagpal, and DeSarbo (1997a, 1997b); Dolan and van der Maas (1998); Hennig (2000); Wedel (2002); and Ingrassia, Minotti, and Vittadini (2012). This approach involves specifying a joint distribution of outcomes and covariates and then maximizing a joint likelihood in lieu of a conditional likelihood. This approach allows parameters of $\mathbf{x}'$s distribution (e.g., means, variances, and covariances of continuous $\mathbf{x}$) not only to be estimated but also to vary across class. The joint likelihood under this approach can be calculated from the following (where parameters are not shown, for simplicity):

$$L = f(\mathbf{y}, \mathbf{x}) = \prod_{i=1}^{N} f(\mathbf{y}_i, \mathbf{x}_i). \tag{7}$$

[5]This approach has also been called "saturated mixture regression" (Wedel, 2002) or "cluster-weighted modeling" (Ingrassia, Minotti, & Vittadini, 2012).

When covariates are incorporated as in Equation (2), we would have

$$f(\mathbf{y}_i, \mathbf{x}_i) = \sum_{k=1}^{K} f(\mathbf{x}_i|c_i = k) p(c_i = k) f(\mathbf{y}_i|\mathbf{x}_i; c_i = k). \tag{8}$$

The density of $\mathbf{x}_i$, $f(\mathbf{x}_i|c_i = k)$, is specific to class *k*. Conceptually, this specification amounts to treating the x's as additional indicators of a latent classification variable, as visually depicted in Figure 1 Panel D. Here, classes are now able to be formed on the basis of between-class differences in means, variances, and covariances of both $y's$ and $x's$. This specification requires making assumptions about the joint distribution of outcomes and covariates within class; often, multivariate normality is assumed. Consequently, unlike the exogenous-*x* approach, the conventional multivariate approach does not require listwise deletion for cases with missing covariates. However, this conventional multivariate mixture was not designed to recover parameters of a conditional mixture, and it poses two drawbacks for researchers with the latter objective.

Drawback 1

From the perspective of researchers desiring to recover parameters of a conditional mixture such as the regression mixture in Equation (5) (Figure 1 Panel B), a drawback of this conventional multivariate approach is that the conditional distribution of inferential interest is not marginally independent from the nuisance distribution $f(\mathbf{x}_i|c_i = k)$. Rather, both mutually depend on the same latent classification variable. This violates one of the earlier stated assumptions (see Bollen, 1989; Johnston, 1984; Jöreskog, 1973; Jöreskog & Goldberger, 1975) for using a joint likelihood approach to recover parameters of a conditional model regardless of the shape of the covariate distribution. As a consequence, using a conventional multivariate (e.g., normal) mixture, nonnormality of $\mathbf{x}$ can be reproduced, during model estimation, by an adjustment in class structure. For instance, parameters of inferential interest—including those of $f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)$ (e.g., $\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\Gamma}^{(k)}$, $\sigma^{2(k)}$) and those defining class proportions (e.g., $\alpha^{(k)}$)—could change based on the need to accommodate higher order moments of nonnormal *x*'s (e.g., if *x* is income or number of alcoholic drinks). This was demonstrated by Arminger et al. (1999) and Arminger and Stein (1997), which motivated their recommendation for exclusively using the exogenous-*x* (conditional likelihood) approach when the objective is to make inferences about parameters of the conditional mixture. This objective is typical of current mixture applications (see Dolan, 2009).

Drawback 2

From the perspective of researchers desiring to recover parameters of a conditional mixture such as Equation (4) or

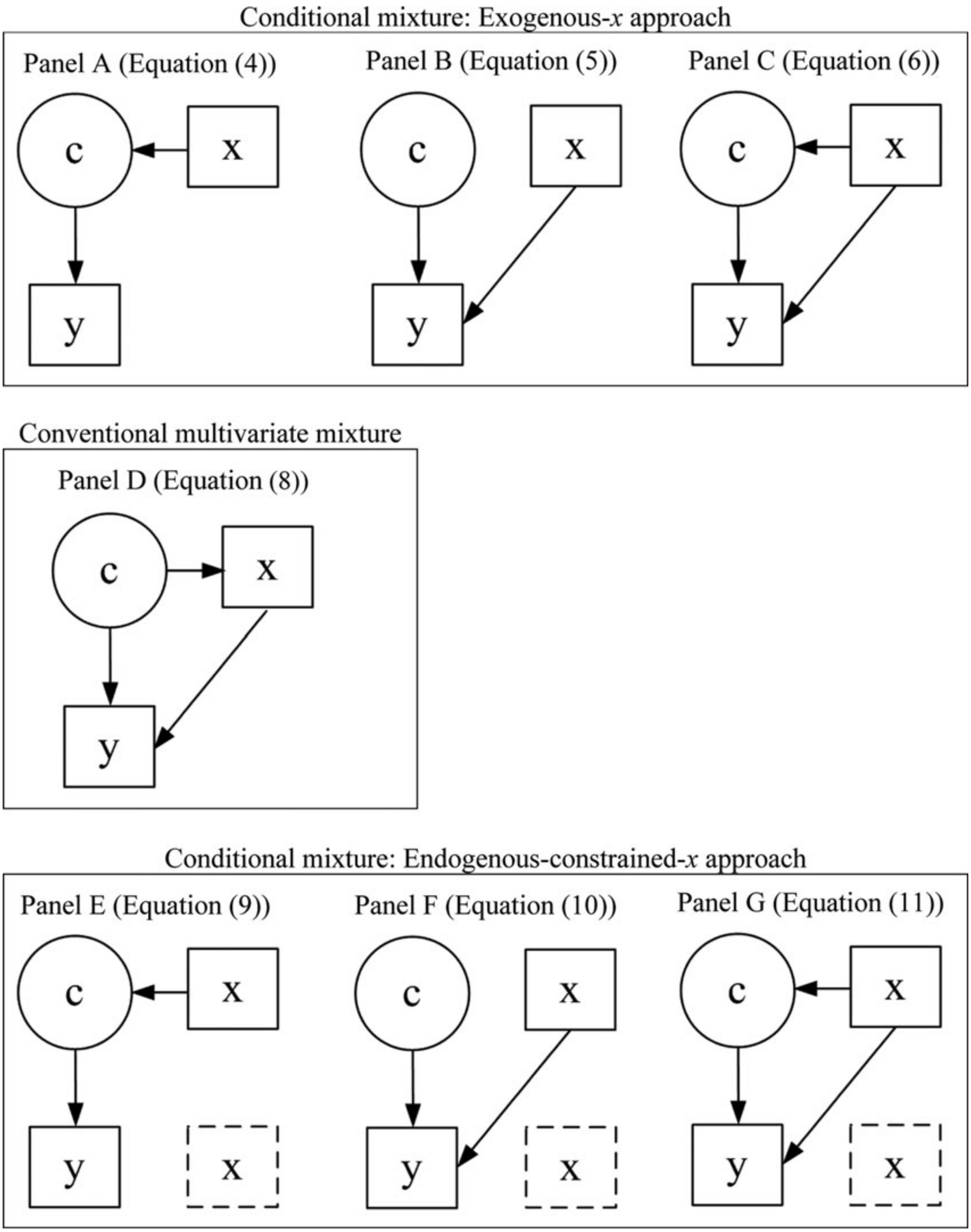

FIGURE 1 Heuristic path diagrams of three model fitting approaches involving one predictor (x). *Note.* These diagrams are heuristic only. c is the latent classification variable. When c points to a variable, this indicates that parameters associated with that variable (e.g., mean and variance) can vary across class. These diagrams depict a single outcome y and a single predictor x (the single predictor x may be incorporated in multiple ways). The dotted line portion heuristically depicts an alternative way to make the single predictor x endogenous.

(6) (Figure 1 Panel A or C, respectively), another drawback of the conventional multivariate mixture is that it cannot incorporate the multinomial regression of class membership on covariates, as in Equation (1). Such a specification would induce a nonrecursive loop and identification problems, as discussed in Asparouhov (2013; see also Maddala, 1983). In theory, multinomial regression coefficients from Equation (4) or (6) conditional mixtures could be derived from class mean differences on x's in particular conventional multivariate mixtures. However, this is not a practical

strategy for overcoming Drawback 2 because these multinomial coefficients can be solved for only under special case circumstances (see Ingrassia et al., 2012); an example is described in Appendix A. There are many situations in which these example special case criteria will not hold. For example, they will not hold for exogenous-x generating processes with x-nonnormality resembling a χ^2 distribution with few degrees of freedom (as in the Study 2 simulation) and with unbalanced classes. It is important to note that even if such special case criteria did hold, this conversion of parameter estimates is inconvenient—more so for increasingly complex within-class specifications. Furthermore, no conversion of standard errors has been developed. Finally, note that for researchers desiring to recover parameters of the Equation (6) conditional mixture, fitting the conventional multivariate mixture poses a combination of Drawbacks 1 and 2.

In sum, for researchers interested in recovering parameters of a conditional mixture with possibly-nonnormal incomplete covariates, we have reviewed drawbacks of using a conventional multivariate mixture (in this section) and drawbacks of using an exogenous-x approach (in the previous section). Hence, a third approach is necessary.

FITTING CONDITIONAL MIXTURES WITH A MODIFIED JOINT LIKELIHOOD: ENDOGENOUS-CONSTRAINED-*X* APPROACH

Our alternative is here called an *endogenous-constrained-x* approach. For researchers interested in interpreting a conditional mixture, it avoids limitations of the two previously considered approaches. It (a) avoids the limitations of the exogenous-x approach by using a joint likelihood to retain cases with missing covariates, and (b) avoids the limitations of the conventional multivariate mixture by modifying the joint likelihood so that the conditional distribution of inferential interest is independent from the nuisance distribution of the x's. Specifically, we employ a joint likelihood but constrain all parameters of the marginal distribution of the x's to be equal across class. Then the joint likelihood can be based on Equation (9) when covariates predict class, Equation (10) when covariates directly predict outcomes within class, or Equation (11) for the combination.

$$f(\mathbf{y}_i, \mathbf{x}_i) = f(\mathbf{x}_i) f(\mathbf{y}_i|\mathbf{x}_i) = f(\mathbf{x}_i) \sum_{k=1}^{K} p(c_i = k|\mathbf{x}_i) f(\mathbf{y}_i|c_i = k) \qquad (9)$$

$$f(\mathbf{y}_i, \mathbf{x}_i) = f(\mathbf{x}_i) f(\mathbf{y}_i|\mathbf{x}_i) = f(\mathbf{x}_i) \sum_{k=1}^{K} p(c_i = k) f(\mathbf{y}_i|\mathbf{x}_i; c_i = k) \qquad (10)$$

$$f(\mathbf{y}_i, \mathbf{x}_i) = f(\mathbf{x}_i) f(\mathbf{y}_i|\mathbf{x}_i) = f(\mathbf{x}_i) \sum_{k=1}^{K} p(c_i = k|\mathbf{x}_i) f(\mathbf{y}_i|\mathbf{x}_i; c_i = k). \qquad (11)$$

Equation (9), (10), or (11) can then be substituted into Equation (7). These extensions are heuristically depicted in the dashed-line portion of Figure 1 Panels E, F, and G. Now, the density of $\mathbf{x}_i$ is not specific to class k. Because the conditional distribution of inferential interest is marginally independent from the nuisance distribution, $f(\mathbf{x}_i)$, the latter is brought out of the across-class summation in Equations (9)–(11) for simplicity. The constraint imposed in Equations (9)–(11) fulfills the earlier stated assumption (e.g., Bollen, 1989; Johnston, 1984) for consistent conditional model inference from a joint likelihood—an assumption violated by the conventional multivariate mixture. The assumption is that parameters of the covariates' marginal distribution are distinct from parameters of the conditional model. In particular, constraining *all* estimated distributional parameters of the x's (e.g., their means and (co)variances) to be invariant across class effectively limits the ability of the model to adjust within-class structure to reproduce nonnormality of x's.

For a different purpose—simply improving parsimony—researchers intending to fit and interpret a conventional multivariate mixture might impose equality constraints on *some* parameters of a variable's distribution (e.g., holding its variance equal across class while letting its mean vary across class; McLachlan & Peel, 2000). However, there is no motivation to impose the particular constraint in Equations (9)–(11)—that *all* parameters of the $\mathbf{x}$ distribution be equal across class—for objectives other than those considered in this article. The reason is that the constraint is unique in converting a conventional multivariate mixture to instead have the *interpretation* of a different model—a conditional mixture. Only researchers who both (a) have the desire to interpret a conditional mixture and (b) have incomplete x's would be motivated to adopt the proposed approach in practice. In sum, this constraint lies at the interface of two separate literatures—the conditional mixture literature and the conventional multivariate mixture literature—borrowing the interpretation of the former and the missing data properties of the latter.

STUDY 1: RECOVERING PARAMETERS OF A CONDITIONAL MIXTURE WITH COMPLETE X′S

In the previous sections we explained why the proposed endogenous-constrained-x approach should have an advantage over the conventional multivariate mixture for recovering parameters of a conditional mixture. Study 1 compares these approaches in the context of complete-x data. The results of Study 1 are used to inform a later Study 2 comparison involving missing-x data.

Hypothesis

Given nonnormal **x**, only the endogenous-constrained-x approach will recover the generating parameters of the conditional mixture as well as does the exogenous-x approach.

Methods

To investigate this hypothesis, we generated 500 complete case data sets of $N = 500$ from the population model of Arminger et al. (1999)—the only previous source to compare exogenous-x and conventional multivariate approaches for recovery of population parameters of a conditional mixture. This population model was a two-class regression mixture model (e.g., DeSarbo & Cron, 1988) using the within-class model of Equation (2) with one outcome and three covariates and using the between-class model: $p(c_i = k) = \exp(\omega_0^{(k)}) / \sum_{k=1}^{K} \exp(\omega_0^{(k)})$ where $\omega_0^{(K)} = 0$. Population parameters from Arminger et al. are given in Table 1. Following Arminger et al., the three covariates were generated to be highly nonnormal—$\chi^2(1)$. Also consistent with Arminger et al., the three covariates were uncorrelated. However, the Study 1 simulation was rerun with the nonnormal covariates correlated at .60 (see Ruscio & Kaczetow, 2008, for generation details) and the same pattern of results was obtained (see the Online Appendix at www.vanderbilt.edu/peabody/sterba/appxs.htm). The covariates were all transformed to have means of 0 and variances of 1.[6] Two-class regression mixture models were fit to the 500 data sets using the three approaches: exogenous-x, conventional multivariate, and endogenous-constrained-x. SAS 9.2 was used for generation and Mplus 7.11 (Muthén & Muthén, 1998–2014) was used for fitting, with 200 sets of random starting values per sample. Syntax used is provided in Appendix B.

[6]Arminger et al. (1999) also considered *normal* x's. Trivial bias for their regression mixture parameters was found when fitting the conventional multivariate mixture with normal x's. The reason is as follows. In the context of their example, normal x's are a special case in which fitting Equation (8) (conventional multivariate) closely approximates the results of fitting Equation (10) (endogenous-constrained-x) in finite samples. In this special case, the freely estimated class-specific means and (co)variances of the x's are estimated at nearly identical values across class because they are not obliged to approximate any x-nonnormality using between-class differences in their values. This is consistent with Ingrassia et al.'s (2012) Proposition 2, which was derived in a different way. It should be emphasized, however, that even in the special case of normal x's in the population, the conventional multivariate mixture does not *exactly* match the exogenous-x approach within a finite sample of, say, $N = 500$ (or on average across such finite samples, as seen in Arminger et al.'s [1999] results). Only the endogenous-constrained-x approach would provide an exact match.

TABLE 1
Comparison of the Conventional Multivariate, Endogenous-Constrained-x, and Exogenous-x Approaches Using Complete-Case Data and Parameters From Arminger et al. (1999): Results Averaged Across 498† Repeated Samples of $N = 500$

		Estimate (*SE*) Averaged Across Samples		
Parameter	Gen. Value	Conventional Multivariate Mixture	Conditional Mixture: Endogenous-Constrained-x	Conditional Mixture: Exogenous-x
			Class 1	
Intercept ($\mu^{(1)}$)	−.5	−.029 *(.150)*	−.499 *(.056)*	−.499 *(.056)*
Slope of y on x_1 ($\gamma_1^{(1)}$)	−1	−.046 *(.182)*	−.999 *(.050)*	−.999 *(.050)*
Slope of y on x_2 ($\gamma_2^{(1)}$)	−1.5	−.078 *(.208)*	−1.501 *(.049)*	−1.501 *(.049)*
Slope of y on x_3 ($\gamma_3^{(1)}$)	.5	.043 *(.182)*	.498 *(.052)*	.498 *(.052)*
Resid. var. ($\sigma^{2(1)}$)	.5	1.164 *(.104)*	.488 *(.059)*	.488 *(.059)*
			Class 2	
Intercept ($\mu^{(2)}$)	.5	.000 *(.325)*	.505 *(.076)*	.505 *(.076)*
Slope of y on x_1 ($\gamma_1^{(2)}$)	1	−.003 *(.183)*	1.002 *(.070)*	1.002 *(.070)*
Slope of y on x_2 ($\gamma_2^{(2)}$)	1.5	.004 *(.186)*	1.502 *(.069)*	1.502 *(.069)*
Slope of y on x_3 ($\gamma_3^{(2)}$)	−.5	−.001 *(.182)*	−.505 *(.073)*	−.505 *(.073)*
Resid. var. ($\sigma^{2(2)}$)	1	9.263 *(1.014)*	.975 *(.107)*	.975 *(.107)*
Multinomial intercept ($\alpha^{(1)}$)*	0	.466 *(.119)*	.007 *(.127)*	.007 *(.127)*

Note. † = convergence achieved for 498/500 samples. * = Generating multinomial intercept of 0 in Class 1 implies class proportions of .50/.50. The estimated Class 1 multinomial intercept of .466 implies class proportions of .61/.39. The estimated Class 1 multinomial intercept of .007 implies proportions .50/.50. Resid. var. = Residual variance. Gen. = generating.

Results and Discussion

Results in Table 1 show that, averaging across converged repeated samples (99.6% converged), there is accurate recovery for the parameters of the conditional mixture under the exogenous-x approach (comparable to Arminger et al.'s [1999] results) but poor recovery using the conventional multivariate mixture (also comparable to Arminger et al.'s results). It is important to note that there was also accurate recovery for the proposed endogenous-constrained-x approach, despite nonnormality of the x's. Furthermore, Table 2 illustrates that, within a given *single sample*, results from the endogenous-constrained-x approach exactly match the exogenous-x results. This is important because in practice researchers have only a single sample. In sum, only with the conventional multivariate normal mixture do estimated class

TABLE 2
Comparison of the Conventional Multivariate, Endogenous-Constrained-*x*, and Exogenous-*x* Approaches Using Complete-Case Data and Parameters From Arminger et al. (1999): Results Shown for a Single Sample of $N = 500$

Parameter	Gen. Value	Estimate (*SE*) for a Single Sample: Conventional Multivariate Mixture	Conditional Mixture: Endogenous-Constrained-*x*	Conditional Mixture: Exogenous-*x*
Class 1				
Intercept ($\mu^{(1)}$)	−.5	.066 *(.136)*	−.591 *(.064)*	−.591 *(.064)*
Slope of *y* on x_1 ($\gamma_1^{(1)}$)	−1	.012 *(.139)*	−.966 *(.064)*	−.966 *(.064)*
Slope of *y* on x_2 ($\gamma_2^{(1)}$)	−1.5	.037 *(.204)*	−1.493 *(.049)*	−1.493 *(.049)*
Slope of *y* on x_3 ($\gamma_3^{(1)}$)	.5	.060 *(.186)*	.601 *(.067)*	.601 *(.067)*
Resid. var. ($\sigma^{2(1)}$)	.5	1.248 *(.110)*	.560 *(.072)*	.560 *(.072)*
Class 2				
Intercept ($\mu^{(2)}$)	.5	.126 *(.380)*	.580 *(.077)*	.580 *(.077)*
Slope of *y* on x_1 ($\gamma_1^{(2)}$)	1	.042 *(.228)*	.830 *(.075)*	.830 *(.075)*
Slope of *y* on x_2 ($\gamma_2^{(2)}$)	1.5	.216 *(.173)*	1.400 *(.055)*	1.400 *(.055)*
Slope of *y* on x_3 ($\gamma_3^{(2)}$)	−.5	.040 *(.273)*	−.481 *(.096)*	−.481 *(.096)*
Resid. var. ($\sigma^{2(2)}$)	1	11.504 *(1.361)*	1.066 *(.112)*	1.066 *(.112)*
Multinomial intercept ($\alpha^{(1)}$)*	0	.668 *(.129)*	−.144 *(.132)*	−.144 *(.132)*

locations and proportions change based on the need to reproduce higher order moments of nonnormal covariates from the regression mixture. In contrast, the endogenous-constrained-*x* approach renders the covariate distribution distinct from the conditional distribution in the modified joint likelihood. Note that this inability of the conventional multivariate approach to recover parameters of this conditional likelihood for nonnormal *x*'s is not a sample size issue. To illustrate this point, the Online Appendix shows that its poor performance persists at $N = 50{,}000$.

Other equivalencies can be noted between the exogenous-*x* and endogenous-constrained-*x* approaches for complete-*x* data. With complete-*x* data, these two approaches can be shown to yield the same result (e.g., same degrees of freedom, likelihood ratio test [LRT] statistic, Bayesian information criterion difference [BIC], etc.) for a model comparison evaluating the within-class and/or between-class effect of **x**. To illustrate this, we must explicitly designate the parameters of the conditional distribution of inferential interest here: $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. $\boldsymbol{\theta}$ contains all parameters of the within-class model (e.g., $\boldsymbol{\theta} = [\boldsymbol{\mu}^{(1)},.., \boldsymbol{\mu}^{(K)}, \boldsymbol{\Gamma}^{(1)},.., \boldsymbol{\Gamma}^{(K)}, \sigma^{2(1)},.., \sigma^{2(K)}]$). $\boldsymbol{\varphi}$ contains all parameters of the between-class model (e.g., $\boldsymbol{\varphi} = [\alpha^{(1)}, ..., \alpha^{(K)}, \boldsymbol{\omega}^{(1)}, ..., \boldsymbol{\omega}^{(K)}]$). The parameters of the nuisance marginal distribution of the *x*'s are denoted $\boldsymbol{\delta}$. Further, suppose $\boldsymbol{\theta}_B$ and $\boldsymbol{\varphi}_B$ contain parameters of the conditional distribution when **x** is allowed to predict **y** in the manner of interest. Also suppose $\boldsymbol{\theta}_A$ and $\boldsymbol{\varphi}_A$ contain parameters of the marginal distribution of **y** when **x** is *not* allowed to predict **y** in this manner. In Equation (12), the LRT statistic under the endogenous-constrained-*x* approach is depicted on the left and the LRT statistic under the exogenous-*x* approach is to the right.

$$-2\ln\left(\frac{L(\boldsymbol{\delta}|\mathbf{x})\,L(\boldsymbol{\theta}_A, \boldsymbol{\varphi}_A|\mathbf{y})}{L(\boldsymbol{\delta}|\mathbf{x})\,L(\boldsymbol{\theta}_B, \boldsymbol{\varphi}_B|(\mathbf{y}|\mathbf{x}))}\right) = -2\ln\left(\frac{L(\boldsymbol{\theta}_A, \boldsymbol{\varphi}_A|\mathbf{y})}{L(\boldsymbol{\theta}_B, \boldsymbol{\varphi}_B|(\mathbf{y}|\mathbf{x}))}\right) \quad (12)$$

Within the brackets on the left versus right side, the numerator likelihoods differ, and the denominator likelihoods differ, but the ratio of likelihoods is equivalent. First consider the left side of the equality. The denominator likelihood corresponds with the endogenous-constrained-*x* specification in which **x** directly predicts **y** and/or indirectly predicts **y** through class (e.g., Equation (9)). In the numerator likelihood, these slope(s) of **x** predicting **y** have been fixed to 0 (e.g., fixing multinomial slopes $\omega_1^{(k)}...\omega_p^{(k)} = 0$ for all *k* in Equation (9)). However, the constrained marginal distribution of **x** is still included, unchanged. Now consider the right side of the equality. The denominator likelihood corresponds with the exogenous-*x* specification in which **x** directly predicts **y** and/or indirectly predicts **y** through class (e.g., Equation (4)). In the numerator likelihood, **x** has been removed from the model entirely. In Equation (12), parameters describing the marginal distribution of **x** (denoted $\boldsymbol{\delta}$) are distinct from those involved in the model comparison (denoted $\boldsymbol{\theta}_A$,$\boldsymbol{\varphi}_A$, $\boldsymbol{\theta}_B$,$\boldsymbol{\varphi}_B$), and $L(\boldsymbol{\delta}|\mathbf{x})$ effectively cancels on the left side.

Here we illustrate the equivalence in Equation (12) in the context of evaluating whether **x** predicts *y* directly within class. Both LRT statistics in Equation (12) were computed for the single sample from Table 2 that was simulated after Arminger et al. (1999).[7] The LRT statistic (H_0:$\gamma_1^{(1)}$=$\gamma_2^{(1)}$=$\gamma_3^{(1)}$=$\gamma_1^{(2)}$=$\gamma_2^{(2)}$=$\gamma_3^{(2)}$=0, i.e., inclusion of **x** does not

[7]In this illustration of testing the within-class effect of **x**, the likelihood ratios in Equation (12), considering parameters unknown, are based on the following probability density function (PDF) ratios: $\frac{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_A^{(k)})}{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_B^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_B^{(k)})} = \frac{\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_A^{(k)})}{\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_B^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_B^{(k)})}$. If instead we had wanted to test the between-class effect of **x**, the likelihood ratios in Equation (12), considering parameters as unknown, would be based on these PDF ratios: $\frac{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_A^{(k)})}{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\mathbf{x}_i,\boldsymbol{\varphi}_B^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_B^{(k)})} = \frac{\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_A^{(k)})}{\sum_{k=1}^{K} p(c_i=k|\mathbf{x}_i,\boldsymbol{\varphi}_B^{(k)})f(y_i|c_i=k,\boldsymbol{\theta}_B^{(k)})}$. Other possibilities include, for instance, testing the between-class effect of **x** using a model that already

improve fit) for the endogenous-constrained-x approach in Equation (12): $\chi^2(6) = 444.92 = (-2\times(-3069.09 + 2846.63))$ matches the LRT statistic for the exogenous-x approach in Equation (12): $\chi^2(6) = 444.92 = (-2\times(-1035.50 + 813.04)$. Neither matches a corresponding LRT statistic for the conventional multivariate mixture ($\chi^2(6) = 1.79$). Mplus syntax for both models employed in this LRT comparison, using the endogenous-constrained-x approach, is provided in Appendix B. It also would have been possible to conduct LRTs for each x instead of the set of three x's. This illustration shows that an LRT for investigating effects of covariates within class provides the same results using endogenous-constrained-x versus exogenous-x approaches.

Also, information criteria differences (e.g., ΔBIC) for model comparisons involving class enumeration will correspond for both approaches under complete-x data.[8] For instance, in the simulated example from Table 2, comparing $K = 1$ versus $K = 2$, $\Delta\text{BIC} = 532.22 = 2226.66–1694.44$, $\Delta df = 6$ (for exogenous-x) and $\Delta\text{BIC} = 532.22 = 6349.77-5817.55$, $\Delta df = 6$ (for endogenous-constrained-x).

Another equivalency between the endogenous-constrained-x versus exogenous-x approaches for complete-x data concerns posterior probabilities of class membership. Posterior probabilities are useful in assessing classification accuracy and in calculating modal class assignments. The left side of Equation (13) depicts posterior probabilities obtained after fitting the modified joint likelihood in Equation (10) (endogenous-constrained-x). The right side depicts posterior probabilities obtained after fitting the conditional likelihood in Equation (5) (exogenous-x). $f(\mathbf{x}_i)$ cancels on the left side.

$$
\begin{aligned}
p(c_i = k|\mathbf{y}_i, \mathbf{x}_i) &= \frac{f(\mathbf{x}_i)p(c_i = k)f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)}{f(\mathbf{x}_i)\sum_{k=1}^{K} p(c_i = k)f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)} \\
&= \frac{p(c_i = k)f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)}{\sum_{k=1}^{K} p(c_i = k)f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)} \qquad (13)
\end{aligned}
$$

For instance, a person from the Table 2 sample with $y_i = .247$, $x_{1i} = .833$, $x_{2i} = -.547$, $x_{3i} = -.401$ has posterior probabilities of .322 for $k = 1$ and .678 for $k = 2$ using either the exogenous-x or endogenous-constrained-x approach.

In sum, we contrasted two existing approaches and one new approach for specifying mixture models with covariates. We discussed that the conventional multivariate approach, although it would accommodate missing x's, is not applicable for researchers interested in interpreting a conditional mixture in that it need not recover generating conditional distribution parameters of interest when x's are nonnormal. We discussed that the exogenous-x approach is also limited in that it does not retain cases with missing x's. The endogenous-constrained-x approach recovers parameters of the conditional distribution, LRT and ΔBIC results, and posterior probabilities equivalently to the exogenous-x approach, with complete-x data. Moreover, the endogenous-constrained-x approach does not require complete-x data. Next, Study 2 addresses this approach's performance under missing x's. Due to the poor performance of the conventional multivariate mixture at recovering parameters of a conditional mixture—the objective of not only this article but also of many psychology applications—the conventional multivariate mixture is not considered further under missing-x in Study 2.

STUDY 2: RECOVERING PARAMETERS OF A CONDITIONAL MIXTURE WITH MISSING X'S

As mentioned earlier, cases with missing covariates are typically listwise deleted in conditional mixture applications in order to fit the model by maximizing a conditional likelihood (exogenous-x approach) using the Expectation-Maximization algorithm (EM; McLachlan & Peel, 2000). Under this approach, the EM algorithm accommodates missing data on measured outcomes (see Lee & Song, 2003) and on the latent classification variable (i.e., unobserved class memberships) but not on measured covariates. This approach is suboptimal for two reasons. First, it results in efficiency loss. Second, when covariates are MCAR (wherein the probability of missingness depends on neither unobserved nor observed variables in the model; Rubin, 1976), it will not result in parameter bias. But if covariates are a kind of MAR (wherein the probability of missingness can depend on observed outcomes in the model) listwise deletion can introduce parameter bias (e.g., Little, 1992; Little & Zhang, 2011; Rabe-Hesketh & Skrondal, 2014).

In contrast, using EM with the endogenous-constrained-x approach in Equation (9), (10), or (11) to fit conditional mixtures allows cases with missing x to be retained under MAR assumptions for x-missingness. Denoting $\mathbf{x}_i = (\mathbf{x}_i^{obs}, \mathbf{x}_i^{mis})$, where $\mathbf{x}_i^{obs}$ is the observed portion of $\mathbf{x}_i$ and $\mathbf{x}_i^{mis}$ is the missing portion, the casewise observed data likelihood can be expressed by integrating the joint likelihood over the missing values:

$$
L_i = \int f(\mathbf{y}_i|\mathbf{x}_i^{obs}, \mathbf{x}_i^{mis}) f(\mathbf{x}_i^{obs}, \mathbf{x}_i^{mis}) d\mathbf{x}_i^{mis} \qquad (14)
$$

(Ghahramani & Jordan, 1997; Little & Rubin, 1983; see also Marlin, 2008). Here we do not consider missing *y's* because this has been previously discussed by Lee and Song (2003) and Chung, Park, and Lanza (2005) and because it is already widely accommodated in EM implementations for fitting mixtures.

In contrast to the complete-x context, in the missing-x context nonnormal covariates have the potential to induce some bias in estimates of conditional distribution parameters, even

---

contained the within-class effect of $\mathbf{x}$; corresponding PDF ratios are $\frac{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_A^{(k)})}{f(\mathbf{x}_i|\boldsymbol{\delta})\sum_{k=1}^{K} p(c_i=k|\mathbf{x}_i,\boldsymbol{\varphi}_B^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_B^{(k)})} = \frac{\sum_{k=1}^{K} p(c_i=k|\boldsymbol{\varphi}_A^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_A^{(k)})}{\sum_{k=1}^{K} p(c_i=k|\mathbf{x}_i,\boldsymbol{\varphi}_B^{(k)})f(y_i|\mathbf{x}_i;c_i=k,\boldsymbol{\theta}_B^{(k)})}$.

[8]We assume no local solutions are obtained for either model under comparison using either approach.

under the endogenous-constrained-x approach. This is because, in the missing-x context, factorization of the joint likelihood in Equation (14) is approximate and no longer holds precisely (see Schafer, 1997). Hence, a simulation study is needed to compare the (typical software default) exogenous-x approach with the endogenous-constrained-x approach for estimating parameters of a conditional mixture with missing x's.

Hypotheses

In Study 2 we compare the performance of the endogenous-constrained-x approach (which retains all cases and assumes MAR) with the exogenous-x approach (which involves listwise deletion of cases[9] with missing x[10]). We compare these approaches under two missingness proportions, two mechanisms for x-missingness (MCAR vs. MAR), and two distributions of **x** (normal vs. nonnormal). This study asks, when x-missingness is MAR versus MCAR and when x's are normal versus nonnormal, is it preferable to treat x's as exogenous or use the endogenous-constrained-x approach?

Hypothesis 1. The exogenous-x approach will show more bias and less efficiency under MAR missingness (here, dependent on observed outcomes) and larger missingness proportions; it will be insensitive to the distribution of x's.

Hypothesis 2. The endogenous-constrained-x approach will be insensitive to the missingness mechanism (MAR vs. MCAR) and the missingness proportion. Whereas the endogenous-constrained-x approach has the potential to be sensitive to the distribution of the x's, we hypothesize that bias incurred under nonnormal x's will be relatively small.

Methods

The generating model was a conditional groups-based trajectory model (Nagin, 2005), which is a particularly widely used conditional mixture (Bauer, 2007; Nagin & Odgers, 2010). Population growth coefficients (see later) were chosen to resemble empirical results from Nagin and Tremblay's (2001) oppositional behavior application. To be consistent with this application, our generating model also had four classes and seven repeated measures. Note that the Study 2 simulation was rerun with a different number of repeated measures (five) and the same pattern of results was obtained. In class k, at occasion $t, y_{ti}|time_{ti}, c_i = k \sim N(\mu^{(k)} + \gamma_1^{(k)} time_{ti} + \gamma_2^{(k)} time_{ti}^2, \sigma^{2(k)})$. Time scores were 0, 1, 2, 3, 4, 5, 6. Two uncorrelated predictors of class membership were included as in Equation (1): $p(c_i = k|x_{1i}, x_{2i}) = \exp(\alpha^{(k)} + \omega_1^{(k)} x_{1i} + \omega_2^{(k)} x_{2i}) / \sum_{k=1}^{K} \exp(\alpha^{(k)} + \omega_1^{(k)} x_{1i} + \omega_2^{(k)} x_{2i})$. In fitted models, multinomial coefficients in the last class were fixed to 0 for identification: $\alpha^{(4)} = \omega_1^{(4)} = \omega_2^{(4)} = 0$.

Residual variances were time and class invariant ($\sigma^{2(k)} = \sigma^2 = 1$). Class-specific parameters were as follows:

$$\begin{array}{l} \mu^{(1)} = 4.0; \gamma_1^{(1)} = .10; \gamma_2^{(1)} = -.02; \alpha^{(1)} = -1; \omega_1^{(1)} = 1; \omega_2^{(1)} = -.75 \\ \mu^{(2)} = 2.7; \gamma_1^{(2)} = -.3; \gamma_2^{(2)} = -.03; \alpha^{(2)} = .1; \omega_1^{(2)} = .5; \omega_2^{(2)} = -.25 \\ \mu^{(3)} = 1.0; \gamma_1^{(3)} = .25; \gamma_2^{(3)} = .05; \alpha^{(3)} = -.8; \omega_1^{(3)} = .75; \omega_2^{(3)} = -.5 \\ \mu^{(4)} = .03; \gamma_1^{(4)} = .01; \gamma_2^{(4)} = .00; \alpha^{(4)} = 0; \omega_1^{(4)} = 0; \omega_2^{(4)} = 0 \end{array}.$$

The multinomial coefficients implied that the marginal class probabilities were .13, .38, .15, .34.

Recall that the manipulated conditions were two missingness percentages (15% vs. 35%), two missingness mechanisms (MCAR vs. MAR, described later), and two distributions of x's (normal or nonnormal, described later). For the normal-x condition, x_1 and x_2 were standard normally distributed; for the nonnormal-x condition, they were distributed as $\chi^2(1)$—transformed to have mean 0 and variance 1. Five hundred samples of $N = 500$ were generated for these two x-distribution conditions using SAS 9.2. This N was found to be a typical sample size used in social science mixture applications by Sterba et al. (2012). For each sample, missingness was generated according to each of the four combinations of missingness mechanism and percentage. The two missingness percentages used were previously considered realistic for practice (Enders & Bandalos, 2001; Merkle, 2011; Wothke, 2000). The probability that x's were missing, denoted $p(m_i = 1)$, depended on the score of the first repeated measure. In the MAR 35% missing condition, $p(m_i = 1) = \exp(-1.55 + .5y_{1i})/(1 + \exp(-1.55 + .5y_{1i}))$ and in the MAR 15% missing condition, $p(m_i = 1) = \exp(-2.8 + .5y_{1i})/(1 + \exp(-2.8 + .5y_{1i}))$.

For each sample, four-class conditional groups-based trajectory models were fit in Mplus 7 with the EM algorithm—using either the exogenous-x or endogenous-constrained-x approach. Mixture models can be identified up to an arbitrary ordering of the class labels. The potential for label switching across samples within cell was addressed using procedures described in, for instance, Asparouhov and Muthén (2010) and Cho, Cohen, and Kim, (2011). According to McLachlan and Peel (2000) and Stephens (2000), these procedures are most common. The same procedures were used in each cell. Class separation in the population values of the intercept parameters ($\mu^{(k)}$) was sufficient to allow consistent label ordering when an inequality on the order of class-specific intercepts was imposed. It could be satisfied by only one permutation of the class labels. The order was initialized with three out of four intercept values in conjunction with repeated random starts. As a subsequent check, distributions

[9]When longitudinal conditional mixtures are specified in multivariate (wide) format, listwise deletion of *cases* with missing covariate(s) is required under the exogenous-x approach. When these models are specified in multilevel (long) format, listwise deletion of *observations* with missing time-varying covariates but *cases* with missing time-invariant covariates is required. To simplify presentation, examples used here do not include time-varying covariates.

[10]Hereafter, *exogenous-x approach* refers to its typical implementation: a conditional likelihood specification using listwise deletion for cases with missing x's.

TABLE 3
Comparing the Exogenous-*x* Approach Versus Endogenous-Constrained-*x* Approach, With Missing *x*'s: Percentage Absolute Relative Bias (%ARB) for Each Parameter Type

	Normal *x*'s				Nonnormal *x*'s			
	MAR		MCAR		MAR		MCAR	
	Endog-constr	Exog	Endog-constr	Exog	Endog-constr	Exog	Endog-constr	Exog
	35% missingness							
Growth coefficient*	5.61	93.13	5.75	7.99	8.06	82.73	7.99	7.69
Multinomial coefficient*	7.13	59.68	7.57	8.98	18.33	59.49	3.87	5.80
Residual variance	0.27	1.01	0.27	0.39	0.27	1.05	0.27	0.48
	15% missingness							
Growth coefficient*	5.57	40.97	5.57	6.17	8.19	28.95	8.08	7.05
Multinomial coefficient*	7.58	26.35	8.09	8.48	6.19	25.87	3.52	4.02
Residual variance	0.27	0.71	0.27	0.32	0.27	0.74	0.27	0.33

Note. * = %ARB was calculated separately for each parameter, and then the %ARB's for parameters of a particular type were averaged. One parameter type was growth coefficients: classes' intercept, linear, and quadratic growth coefficients. Another parameter type was multinomial coefficients: multinomial intercepts and multinomial slopes of *x*'s that predict class. Endog-constr = endogenous-constrained-*x* approach; Exog = exogenous-*x* approach; MCAR = missing completely at random; MAR = missing at random (here, where missingness depended on observed outcomes). One growth coefficient's %ARB was undefined because its population parameter was 0; it was not included in the average %ARB. An Online Appendix provides %ARB for each parameter individually.

of estimates from fitted solutions were inspected for label switching. We employed 200 sets of random starting values per sample to decrease the risk of local solutions.

The outcomes of interest for parameter estimates were percentage absolute relative bias (%ARB) and relative efficiency (RE). For a generic parameter ϑ, %ARB $= |((\bar{\hat{\vartheta}} - \vartheta)/\vartheta) \times 100|$. RE is defined as a parameter's empirical repeated sampling standard deviation (*SD*) under the exogenous-*x* approach divided by its standard deviation under the endogenous-constrained-*x* approach.

Results and Discussion

Convergence was 99–100%. %ARB and RE were calculated for each parameter separately. Subsequently, averages of these %ARB and RE values were calculated and reported for each type of parameter, in Tables 3 and 4, respectively. This averaging was done for simplicity and because results were similar for a parameter type (i.e., for growth coefficients ($\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\Gamma}^{(k)}$), for multinomial coefficients ($\alpha^{(k)}$, $\boldsymbol{\omega}'^{(k)}$), and for residual variance (σ^2)).[11] Tables containing%ARB results for each individual parameter separately are given in the Online Appendix. The Online Appendix also contains, for each parameter separately, the across-samples average of the estimate, the standard deviation, and the average analytic standard error of the estimate (*SE*).

[11]Similar reporting procedures were used in Lu, Zhang, and Lubke (2011); Forero, Maydeu-Olivares, and Gallardo-Pujol, (2009); and Reinartz, Echambadi, and Chin, (2002).

Table 3 shows that, for nearly every parameter type, the endogenous-constrained-*x* approach yielded smaller %ARB than the popular exogenous-*x* approach within each cell of the simulation.

More specifically, in support of Hypothesis 1, Table 3 shows that using the exogenous-*x* approach, there could be much greater bias when missingness was MAR (vs. MCAR). These patterns were more prominent for higher missingness (35%) and for growth coefficients (i.e., $\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\Gamma}^{(k)}$). Whether the distribution of **x** was normal or nonnormal did not matter, as expected. Under the anticipated unfavorable condition for the exogenous-*x* approach (i.e., MAR missingness dependent on observed outcomes) at 35% missingness,%ARB was for growth coefficients ($\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\Gamma}^{(k)}$) 82.73–93.13%, for multinomial coefficients ($\alpha^{(k)}, \boldsymbol{\omega}'^{(k)}$) 59.49–59.68%, and for residual variance (σ^2) 1.01–1.05%. Under the same MAR condition at 15% missingness,%ARB was for growth coefficients 28.95–40.97%, for multinomial coefficients 25.87–26.35%, and for residual variance 0.71–0.74%.

In support of Hypothesis 2, Table 3 also shows that, when using the endogenous-constrained-*x* approach, there could be slightly more bias when **x** was extremely nonnormal ($\chi^2(1)$). Whether missingness was MAR versus MCAR and whether the missingness proportion was 15% or 35% mattered little, as expected. Under the anticipated unfavorable condition for the endogenous-constrained-*x* approach (i.e., extreme nonnormality)%ARB was for growth coefficients ($\boldsymbol{\mu}^{(k)}$, $\boldsymbol{\Gamma}^{(k)}$) 7.99–8.19%, for multinomial coefficients ($\alpha^{(k)}, \boldsymbol{\omega}'^{(k)}$) 3.52–18.33%, and for residual variance (σ^2) 0.27%.

Additionally, Table 4 shows that the exogenous-*x* approach generally led to worse efficiency than the

TABLE 4
Comparing the Exogenous-x Approach Versus Endogenous-Constrained-x Approach, With Missing x's: Relative Efficiency (RE) for Each Parameter Type

	Normal x's		Nonnormal x's	
	MAR	MCAR	MAR	MCAR
	35% missingness			
Growth coefficient*	1.38	1.30	1.29	1.41
Multinomial coefficient*	1.03	1.03	0.95	1.07
Residual variance	1.27	1.28	1.28	1.24
	15% missingness			
Growth coefficient*	1.12	1.10	1.11	1.14
Multinomial coefficient*	1.01	1.01	0.98	1.02
Residual variance	1.10	1.07	1.12	1.08

Note. * = RE was calculated separately for each parameter, and then the RE's for parameters of a particular type were averaged. One parameter type was growth coefficients: classes' intercept, linear, and quadratic growth coefficients. Another parameter type was multinomial coefficients: multinomial intercepts and multinomial slopes of x's that predict class. MCAR = missing completely at random; MAR = missing at random (here, where missingness depended on observed outcomes).

endogenous-constrained-x approach. That is, RE ratios were greater than 1 in 22 out of 24 cells of Table 4. The exogenous-x approach had particularly worse efficiency under larger missingness proportions and for growth coefficients (where it was 29–41% less efficient than the endogenous-constrained-x approach).

In sum, for conditions examined here, the endogenous-constrained-x approach emerged as a more attractive option for handling x-missingness in conditional mixtures. This approach is particularly attractive when the x-missingness mechanism is unknown as is the case in real-world settings (and thus possibly MAR depending on observed outcomes) and when the proportion of missingness is sizable.

Generalizability of Study 2 Results

The Study 2 simulation involved one chosen MAR mechanism and one chosen degree of x-nonnormality. Here we address how the results would be expected to generalize to other levels of MAR and nonnormality. Our x-nonnormality was intentionally chosen to be extreme to provide a conservative evaluation of the proposed endogenous-constrained-x approach, whereas the MAR condition was chosen to be moderate to provide a more liberal evaluation of the exogenous-x approach for comparison purposes, as follows. The x-nonnormality in Study 2 (and Study 1), $\chi^2(1)$, implied kurtosis of approximately 12 and skewness of approximately 3. Such skew and kurtosis can be considered extreme according to Micceri's (1989) survey, which reported that only 8% of psychological variables had kurtosis > 3 and only 10% had skewness > 2. The degree of MAR in Study 2, on the other hand, could be considered moderate rather than extreme because it implied an average correlation between missingness and the observed variable on which missingness depends (i.e., y_1) of $r = .38$ for the 35% MAR condition and $r = .29$ for the 15% MAR condition. This is a medium effect size according to Cohen (1988, p. 80). Other simulations have considered more extreme MAR mechanisms as potentially reflective of practice (e.g., $r = .60$; Yuncel, He, & Zaslavsky, 2011; $r = .45$; Allison, 2006; $r = .67$ and .45; Graham, 2012). Thus, based on previous studies, we could have chosen a stronger MAR mechanism (more departure from MCAR) and/or weaker nonnormality. Either could result in a greater bias advantage for endogenous-constrained-x compared with exogenous-x. If the MAR mechanism were instead chosen to be very weak (regardless of whether x-nonnormality remained extreme or was reduced), both approaches should have quite low bias at that point. However, the exogenous-x approach would *still* be less efficient. Next, an empirical example is used to illustrate how the performance of the endogenous-constrained-x approach generalizes to other conditions not considered in the simulations.

EMPIRICAL EXAMPLE

The simulations described earlier used continuous y's and x's. Here we use an empirical example to show how those simulation results generalize to alternative outcome and covariate distributions. Specifically, this empirical illustration involves six binary y's and a combination of two binary x's and two nonnormal (continuous) x's. The empirical illustration sample consists of 17- to 18-year-olds exiting Midwestern state-run or foster care facilities (Courtney & Cusick, 2007). These adolescents are at high risk of conduct problems (McMillen et al., 2005). The goal of this analysis is to use a conditional latent class analysis model (e.g., Collins & Lanza, 2010) to predict membership in latent classes of conduct disorder symptoms. The six binary conduct disorder outcomes serve as indicators of class membership. These outcomes are property damage, stealing > \$50, breaking/entering, making threats, group fighting, and weapon use. Four x's serve as predictors of class membership. They are gender (binary), social support (a continuous standardized scale score from Courtney & Cusick, 2007), employment status (binary), and physical abuse history (continuous). The two continuous covariates were nonnormally distributed. Physical abuse had skew = 1.73 and kurtosis = 2.14. Social support had skew = −.84 and kurtosis = .06.

Our interest is in comparing the results of the exogenous-x and endogenous-constrained-x approaches within sample, given complete-x versus missing-x data. Hence, we needed access to a complete-x data set from which we could construct a missing-x data set. Our complete-x data set had $N = 698$. Our missing-x data set was created by inducing missingness on two of the four covariates at a rate (39%) that matched the empirical rate of missingness on the covariates at a follow-up survey (ages 19–20).[12] Specifically, the missing-x data set was created by generating missing values for two sensitive questions—physical abuse history and employment status—in the age 17–18 data set under an MAR process in which adolescents were more likely to respond on the sensitive questions if they had higher social support: $p(m_i = 1) = \exp(-.5 - .7support_i)/(1 + \exp(-.5 - .7support_i))$.[13] There was no missingness on **y** or on the other two covariates (social support and gender).

First we select K. Under the assumption that covariates only predict class membership and do not enter the within-class model, class enumeration can employ an unconditional mixture, as done here (Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Collins & Lanza, 2010; Lubke & Muthén, 2007). The best-fitting number of classes was chosen as $K = 2$ based on BIC comparisons of unconditional latent class models with $K = 1$ (BIC = 3600.85), $K = 2$ (BIC = 3336.529), and $K = 3$ (BIC = 3340.996). One class, containing 17.6% of adolescents, had higher endorsement probabilities of all conduct disorder symptoms, particularly property damage and fighting. The other, a low-problems class, consisted of 82.4% of adolescents. Next, conditional latent class models with x's predicting class were fit with $K = 2$ using both approaches for the complete-x and missing-x data sets.

Results in Table 5 show that, for complete-x data, the exogenous-x and endogenous-constrained-x results[14] match—as was the case in earlier simulations with continuous y's and x's. Table 5 also shows that, for missing-x data, standard errors under the exogenous-x approach are largest due to the efficiency loss incurred in listwise deletion, which reduced the sample from 698 to 425.

To interpret the effects of predictors on class membership, we focus on the endogenous-constrained-x approach results for the missing-x data set and exponentiate multinomial slopes ($\omega_1^{(k)}, \omega_2^{(k)}, \omega_3^{(k)}, \omega_4^{(k)}$) for the four predictors. Doing so indicates that the odds of being in the high conduct problems class versus the low problems class significantly increase by a factor of 1.30 for each physical abuse experience and by a factor of 4.72 for being male, where each effect controls for other predictors. On the other hand, the odds of being in the high versus low problems class are significantly reduced by a factor of .41 (or 59%) for being employed and by a factor of .71 (or 29%) for every 1 *SD* increase in social support, controlling for other predictors. The finding that having a prior abuse history and being male serve as risk factors for membership in the conduct problems class but having social support and employment serve as protective factors is consistent with prior theories of risk and resilience (e.g., Rutter, 1987). The latter finding is also consistent with social control theories (e.g., Brame, Bushway, Paternoster, & Apel, 2004; Crutchfield, 2014; Hirschi, 1969). Nevertheless, causal inferences are not warranted particularly because social support level and employment status were measured at the same time as the conduct problems (Shadish, Cook, & Campbell, 2002). One way to somewhat strengthen the grounds for inferring causation would be to assess the effect of employment status at time t-1 on conduct problem profile at time t, controlling for conduct problem profile at time t-1.

TABLE 5
Empirical Example Results

Parameter	Complete-x Endogenous-Constrained $N = 698$	Complete-x Exogenous $N = 698$	Missing-x Endogenous-Constrained $N = 698$	Missing-x Exogenous (Listwise Deletion) $N = 425$
	Multinomial coefficients			
$\alpha^{(1)}$	−2.53*(.37)	−2.53*(.37)	−2.58*(.43)	−2.88*(.71)
$\omega_1^{(1)}$ (male)	1.47*(.31)	1.47*(.31)	1.55*(.33)	1.83*(.56)
$\omega_2^{(1)}$ (employed)	−.67*(.30)	−.67*(.30)	−.90*(.40)	−1.00*(.43)
$\omega_3^{(1)}$ (support)	−.37*(.13)	−.37*(.13)	−.36*(.13)	−.53*(.20)
$\omega_4^{(1)}$ (abuse)	.27*(.07)	.27*(.07)	.26*(.09)	.28*(.10)
	Class 1 thresholds†			
Prop dmg	−.76*(.27)	−.76*(.27)	−.77*(.28)	−.63(.33)
Steal	.43(.23)	.43(.23)	.39(.24)	.14(.34)
Break/enter	.13(.26)	.13(.26)	.06(.28)	−.37(.47)
Threat	.22(.23)	.22(.23)	.18(.23)	−.04(.32)
Group fight	−.37(.22)	−.37(.22)	−.42(.23)	−.27(.30)
Weapon	.28(.22)	.28(.22)	.24(.22)	.21(.31)
	Class 2 thresholds†			
Prop dmg	2.15*(.19)	2.15*(.19)	2.11*(.18)	1.87*(.21)
Steal	3.30*(.29)	3.30*(.29)	3.28*(.28)	3.25*(.37)
Break/enter	3.43*(.29)	3.43*(.29)	3.43*(.29)	3.46*(.38)
Threat	3.72*(.46)	3.72*(.46)	3.66*(.46)	3.26*(.46)
Group fight	.85*(.10)	.85*(.10)	.85*(.10)	.80*(.12)
Weapon	2.80*(.24)	2.80*(.24)	2.79*(.24)	2.76*(.29)

Note. * = $p < .05$. †Thresholds can be converted to probabilities; probability = 1/(1 + exp(threshold)). Prop dmg = property damage.

DISCUSSION

Applications of conditional mixture models are common because researchers want to predict unobserved individual heterogeneity in behavior or in change over time using external covariates as part of assessing the construct validity of a ty-

[12]The follow-up survey was not used in the current analysis of 17- and 18-year-olds; it was just used to gauge an externally valid missing x rate for the participants when constructing the missing-x data set.

[13]The endogenous-constrained-x approach accommodates MAR missingness depending on any observed variables in the model (y's or x's). Previous research outside the mixture context has shown that the exogenous-x approach risks bias when x-missingness depends on y's rather than x's (see Little, 1992, p. 1229). We also generated x-missingness on abuse history and employment that depended on y_1 (the property damage symptom): $p(m_i = 1) = \exp(-.5 + .7propdmg_i)/(1 + \exp(-.5 + .7propdmg_i))$. The same overall pattern of results was found with respect to the comparison of the endogenous-constrained-x and exogenous-x approaches, though only a single sample was used here.

[14]When comparing parameter estimates across alternative mixture models fit to the same sample, it is important to ensure (e.g., using starting values) that the classes retain the same order.

pology (e.g., Bandeen-Roche, Miglioretti, Zeger, & Rathouz, 1997; Chung et al., 2006; Goodman, Crouter, Lanza, Cox, & Vernon-Feagans, 2011; Jedidi et al., 1996; Muthén, 2002; Nagin, 2005; Schaeffer et al., 2006; Sterba & Bauer, 2014). Psychologists currently fit conditional mixtures using a conditional likelihood specification that treats covariates as fixed and fully observed. Using this exogenous-x approach, missingness on covariates is most commonly handled with listwise deletion, which is inefficient and requires that covariate missingness not depend on outcomes.

For researchers with incomplete covariates and the objective of interpreting a conditional mixture, this article considered a modified joint likelihood approach that could retain all available cases under more realistic MAR assumptions. Previous research with complete-x data had shown that one joint likelihood approach, a conventional multivariate normal mixture, could not meet the objective of recovering conditional mixture parameters under $\mathbf{x}$ nonnormality of arbitrary origin. This article provided the first explanation of this finding, showing that the conventional multivariate mixture violates a necessary assumption for drawing conditional model inference from a multivariate model with random regressors (e.g., Bollen, 1989; Johnston, 1984). A modified joint likelihood approach—an endogenous-constrained-x approach—avoids this violation.

After demonstrating that the proposed approach performed well at recovering parameters of the conditional mixture under complete-x, we evaluated its performance more thoroughly for missing x's in comparison with the typical exogenous-x approach. Simulation results with missing x's evidenced that treating x's as endogenous-constrained when fitting a mixture model was preferable (in terms of lower bias and better efficiency) to the typical exogenous-x approach across a variety of conditions. Although this simulation with missing x's involved only one kind of predictor effect (i.e., x's predicting class membership), the same pattern of results was obtained in several simulation checks with other kinds of predictor effects (e.g., x's predicting y's within class—exclusively or in addition to x's predicting class membership). Results are available from the author upon request. Furthermore, the pattern of results from the simulations was also obtained in our empirical example on conduct problems, which involved categorical x's as well as continuous, nonnormal x's. Future research can evaluate the performance of the endogenous-constrained-x approach across other conditions and other kinds of mixture models.

Other Advantages of the Endogenous-Constrained-*x* Approach

Other important advantages of the endogenous-constrained-x approach are as follows. First, it is straightforward for applied researchers to implement this approach in existing commercial software for mixture modeling (see Appendix B for an example). Next, the applicability of this approach is not limited to situations in which researchers desire to incorporate covariates directly into the mixture model (called a one-step or simultaneous estimation strategy). Rather, this approach can also be used in conjunction with "three-step" estimation strategies[15] for incorporating predictors of class membership (e.g., Asparouhov & Muthén, 2013; Vermunt, 2010b). These three-step strategies currently incorporate predictors in Step 3 using an exogenous-x approach. The incorporation of predictors in Step 3 could instead be changed to an endogenous-constrained-x approach, under MAR assumptions, to prevent listwise deletion of cases with missing covariates.

The endogenous-constrained-x approach was demonstrated here with three illustrative conditional mixtures—a regression mixture (in Study 1), a conditional groups-based trajectory model (in Study 2), and a conditional latent class model (in the empirical example). An important strength of this approach is that it can also be applied to many other different kinds of conditional mixtures.

Another advantage is that the number of nuisance parameters that need to be estimated under the endogenous-constrained-x approach is relatively small in that the required number does not grow as K increases. A final advantage over the exogenous-x approach can be noted. When incomplete x's are treated as exogenous, a problem maintaining the same N arises when comparing mixtures with versus without x's in the model using information criteria or an LRT (i.e., Equation (12), right side). Fitting the latter model allows the full sample, whereas fitting the former model uses a reduced N due to listwise deletion. LRTs or information criteria require using same N across competing models (e.g., Burnham & Anderson, 2002; Enders, 2010). Researchers often address this issue by using the reduced N for all fitted models, even unconditional models, sacrificing efficiency. In contrast, when using the endogenous-constrained-x approach (e.g., LRT on the left side of Equation (12)), the full sample could always be used. Similarly, the full sample can be used in comparisons any time x's are endogenous.

When Is the Endogenous-Constrained-*x* Approach Useful?

In nonmixture (e.g., multiple regression) contexts, researchers are accustomed to generally obtaining the same conditional model inference using a conditional or joint likelihood. In mixture contexts, researchers encounter a choice because conditional mixtures and conventional multivariate mixtures generally lead to *different* interpretations and

[15]An example three-step strategy is as follows. First an unconditional mixture model may be fit, and posterior probabilities as well as modal class assignments are saved. Second, these, along with estimated model parameters, are used to calculate classification error probabilities (see Asparouhov & Muthén, 2013; Vermunt, 2010b). Third, a conditional mixture model is fit with covariates predicting class membership. Modal class assignments serve as a nominal indicator of class, and the indicator's K-1 thresholds per class are fixed to a function of the classification error probabilities.

inferences. As stated previously, a fitted conventional multivariate normal mixture typically cannot recover parameters of a conditional mixture (see our Study 1's results and see Arminger et al., 1999). Conversely, a fitted conditional mixture typically cannot recover parameters of a conventional multivariate normal mixture (see simulation results of Ingrassia et al., 2012).[16] Hence, the researcher needs to decide which model's parameters are of substantive and inferential interest and specify their fitted model accordingly.

This article assumed interest in recovering conditional mixture parameters. Hence, both the Study 1 and 2 simulations used exogenous-*x* conditional mixtures as population-generating models (similar to Arminger et al., 1999, who shared the same goal). This means that, first, *x*′s are generated. Next, class memberships are generated potentially as a function of *x*′s. Finally, *y*′s are generated as a function of class membership and potentially *x*′s. The reason for our focus on conditional mixtures is that, currently, empirical and methodological articles in psychology concerning mixtures with covariates largely choose to fit conditional mixtures (rather than conventional multivariate mixtures) for substantive reasons (e.g., Bauer, 2007; Collins & Lanza, 2010; Dayton & Macready, 1988; Dolan, 2009, Feldman et al., 2009; Goodman et al., 2011; Lubke & Muthén, 2005; Muthén & Shedden, 1999; Nagin, 2005; Pickles & Croudace, 2010; Reinecke, 2006; Vermunt, 2010a; Wedel & DeSarbo, 1995). Researchers also make this choice so they can distinguish between outcomes whose (conditional) distributions are used to define the typology and covariates that can be used to validate the typology but whose distributions do not define it. According to Shadish et al. (2002), part of construct validation is deciding what variables are *not* definitional of a construct so as to prevent what they call "construct confounding" (p. 75). Similarly, in our empirical example we did not want the distribution of covariates (e.g., gender and social support) to help define class structure—as the categorical latent construct would then no longer be interpretable as a conduct behavior symptom typology.

Should our interest have been in recovery of parameters of a conventional multivariate mixture, that model could instead have served as a generating model (e.g., Ingrassia et al., 2012). Doing so would have required first generating class membership, then generating *x*-distributions as a function of class membership, and then generating *y*'s as a function of class and potentially *x*'s. If researchers are interested in interpreting a conventional multivariate mixture (e.g., Equation (8)) or a related multivariate mixture such as a parallel process mixture (e.g., Sterba, 2013), there is no need to employ the endogenous-constrained-*x* approach.

[16]Although Ingrassia et al. (2012) fit conditional mixtures using one approach (exogenous-*x*), their findings should hold using the other approach (endogenous-constrained-*x*).

Limitations

Several limitations deserve mention. First, the simulation held the sample constant at 500. Smaller sample sizes together with substantial proportions of missing data could lead to high rates of estimation problems and empirical underidentification for small classes (McLachlan & Peel, 2000), which were not a focus here. Second, only one illustrative MAR mechanism was used here in Study 2; generalizability of the pattern of results under alternative levels of MAR was discussed earlier. Third, there could be other ways to include a saturated, class-invariant $f(\mathbf{x}_i)$ in a modified joint likelihood that were not explored here. Nevertheless, the method considered here is simple to implement, generally applicable, and was effective in the simulations conducted. Fourth, Appendix A provided one kind of parameter transformation; different kinds of model transformations are used in other literatures for a variety of purposes (e.g., see von Oertzen [2010] and von Oertzen & Brandmaier [2013] in the context of power equivalence in structural equation modeling). Fifth, although we considered only MAR mechanisms, covariates could be missing-not-at-random (MNAR). MNAR covariates are addressed by Cai, Song, and Hser (2010) within a Bayesian framework. Finally, if missingness proportions were trivial—not considered here—the choice between exogenous-*x* and endogenous-constrained-*x* approaches may be of little empirical consequence.

Conclusions

We recommend the endogenous-constrained-*x* approach over the typical exogenous-*x* approach for handling covariate missingness when parameters of a conditional mixture model are of inferential interest. Even for highly nonnormal *x*′s, this approach can outperform an exogenous-*x* approach. The advantage of the endogenous-constrained-*x* approach in terms of bias and efficiency was most notable for larger missing data proportions and MAR mechanisms depending on observed outcomes.

ACKNOWLEDGMENTS

I thank Daniel J. Bauer and Kristopher J. Preacher for helpful comments.

REFERENCES

Allison, P. D. (2006). *Multiple imputation of categorical variables under the multivariate normal model*. Paper presented at the annual meeting of the American Sociological Association, Montreal, Canada.

Arminger, G., & Stein, P. (1997). Finite mixtures of covariance structure models with regressors: Log-likelihood function, minimum distance estimation, fit indices, and a complex sample. *Sociological Methods & Research*, *26*, 148–182.

Arminger, G., Stein, P., & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika*, *64*, 475–494.

Asparouhov, T. (2013). *Bidirectional association*. Technical appendix retrievable from www.statmodel.com

Asparouhov, T., & Muthén, B. O. (2010). *Bayesian analysis using Mplus: Technical implementation*. Retrievable from www.statmodel.com

Asparouhov, T., & Muthén, B. O. (2013). *Auxiliary variables in mixture modeling: 3-step approaches using Mplus* (Mplus Web Note No. 15). Retrievable from www.statmodel.com

Bakk, Z., Oberski, D. L., & Vermunt, J. K. (2014). Relating latent class assignments to external variables: Standard errors for correct inference. *Political Analysis*. doi:10.1093/pan/mpu003

Bandeen-Roche, K., Miglioretti, D., Zeger, S., & Rathouz, P. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*, 1375–1386.

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786.

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*, 338–363.

Blåfield, E. (1980). *Clustering of observations from finite mixtures with structural information* (Unpublished doctoral dissertation). Jyvaskyla University, Jyvaskyla, Finland.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation approach*. Hoboken, NJ: Wiley.

Brame, R., Bushway, S., Paternoster, R., & Apel, R. (2004). Assessing the effect of adolescent employment on involvement in criminal activity. *Journal of Contemporary Criminal Justice*, *20*, 236–256.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). New York, NY: Springer-Verlag.

Cai, J.-H., Song, X.-Y., & Hser, Y.-I. (2010). A Bayesian analysis of mixture structural equation models with non-ignorable missing responses and covariates. *Statistics in Medicine*, *29*, 1861–1874.

Castelao, C., & Kroner-Herwig, B. (2012). Different trajectories of depressive symptoms in children and adolescents: Predictors and differences in girls and boys. *Journal of Youth and Adolescence*, *42*, 1169–1182.

Cho, S.-J., Cohen, A. S., & Kim, S.-H. (2011). Markov chain Monte Carlo estimation of a mixture item response theory model. *Journal of Statistical Computation and Simulation*, *83*, 278–306.

Chung, H., Flaherty, B., & Schafer, J. (2006). Latent class logistic regression: Application to marijuana use and attitudes among high school seniors. *Journal of the Royal Statistical Society, Series A*, *169*, 723–743.

Chung, H., Park, Y., & Lanza, S. (2005). Latent transition analysis with covariates: Pubertal timing and substance abuse behaviors in adolescent females. *Statistics in Medicine*, *24*, 2895–2910.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences.* Hoboken, NJ: Wiley.

Courtney, M., & Cusick, G. (2007). *Crime during the transition to adulthood: How youth fare as they leave out-of-home care in Illinois, Iowa and Wisconsin* [Computer file].

Crutchfield, R. D. (2014). *Labor markets, economic opportunity, and crime*. New York, NY: New York University Press.

Dayton, C., & Macready, G. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, *83*, 173–178.

DeSarbo, W. S., & Cron, W. L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*, 249–282.

Dolan, C. V. (2009). Structural equation mixture modeling. In R. Millsap & A. Maydeu-Olivares, (Eds.), (*The Sage handbook of quantitative methods in psychology* (pp. 568–591). London, UK: Sage.

Dolan, C., & van der Maas, H. (1998). Fitting multivariate normal finite mixtures subject to structural equation modeling. *Psychometrika*, *63*, 227–253.

Draper, N. R., & Smith, H. (1998). *Applied regression analysis* (3rd ed.). New York, NY: Wiley.

Enders, C. K. (2010). *Applied missing data analysis.* New York, NY: Guilford Press.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457.

Enders, C. K., & Gottschall, A. (2011). Multiple imputation strategies for multiple group structural equation models. *Structural Equation Modeling*, *18*, 35–54.

Feldman, B. J., Masyn, K. E., & Conger, R. D. (2009). New approaches to studying problem behaviors: A comparison of methods for modeling longitudinal, categorical adolescent drinking data. *Developmental Psychology*, *45*, 652–676.

Fontaine, N., Carbonneau, R., Vitaro, F., Barker, E. D., & Tremblay, R. E. (2009). Research review: A critical review of studies on the developmental trajectories of antisocial behavior in females. *Journal of Child Psychology and Psychiatry*, *50*, 363–385.

Forero, C., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling*, *16*, 625–641.

Ghahramani, Z., & Jordan, M. (1997). Mixture models for learning from incomplete data. In R. Greiner, T. Petsche, & S. Hanson (Eds.), *Computational learning theory and natural learning systems* (pp. 67–86). Boston, MA: Bradford.

Goodman, W. B., Crouter, A., Lanza, S., Cox, M., & Vernon-Feagans, L. (2011). Paternal work stress and latent profiles of father-infant parenting quality. *Journal of Marriage & Family*, *73*, 588–604.

Graham, J. W. (2012). *Missing data: Analysis and design*. New York, NY: Springer.

Greenbaum, P., Del Boca, F., Darkes, J., Wang, C., & Goldman, M. (2005). Variation in the drinking trajectories of freshman college students. *Journal of Consulting and Clinical Psychology*, *73*, 229–238.

Hennig, C. (2000). Identifiability of models for clusterwise linear regression. *Journal of Classification*, *17*, 273–296.

Hepworth, S., Law, G., Lawlor, D., & McKinney, P. (2010). Early life patterns of common infection: A latent class analysis. *European Journal of Epidemiology*, *25*, 875–883.

Hirschi, T. (1969). *Causes of delinquency*. Berkeley: University of California Press.

Horton, N., & Kleinman, K. (2007). Much ado about nothing. *The American Statistician*, *61*, 79–90.

Horton, N., & Laird, N. (1998). Maximum likelihood analysis of generalized linear models with missing covariates. *Statistical Methods in Medical Research*, *8*, 37–50.

Hosmer, D. (1974). Maximum likelihood estimates of the parameters of a mixture of two regression lines. *Communications in Statistics*, *3*, 995–1006.

Hox, J., & Roberts, K. (2010). *Handbook of advanced multilevel analysis*. New York, NY: Routledge.

Ibrahim, J. G. (1990). Incomplete data in generalized linear models. *Journal of the American Statistical Association*, *85*, 765–769.

Ingrassia, S., Minotti, S., & Vittadini, G. (2012). Local statistical modeling via a cluster-weighted approach with elliptical distributions. *Journal of Classification*, *29*, 363–401.

Jansen, R. C. (1993). Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics*, *49*, 227–231.

Jedidi, K., Jagpal, H., & DeSarbo, W. (1997a). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, *16*, 39–59.

Jedidi, K., Jagpal, H., & DeSarbo, W. (1997b). STEMM: A general finite mixture structural equation model. *Journal of Classification*, *14*, 23–50.

Jedidi, K., Ramaswamy, V., DeSarbo, W., & Wedel, M. (1996). On estimating finite mixtures of multivariate regression and simultaneous equation models. *Structural Equation Modeling*, *3*, 266–289.

Johnston, J. (1984). *Econometric methods* (3rd ed.). New York, NY: McGraw-Hill.

Jones, B., Nagin, D., & Roeder, K. (2001). An SAS procedure based on mixture models for estimating developmental trajectories. *Sociological Methods & Research*, *29*, 374–393.

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, NY: Academia Press.

Jöreskog, K., & Goldberger, A. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, *70*, 631–639.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8 user's reference guide*. Uppsala, Sweden: Scientific Software International.

Lanza, S., Dziak, J., Huang, L., Xu, S., & Collins, L. (2011). *Proc LCA & Proc LTA users' guide* (Version 1.2.7). University Park, PA: The Methodology Center, Penn State.

Lee, S., & Song, X.-Y. (2003). Maximum likelihood estimation and model comparison for mixtures of structural equation models with ignorable missing data. *Journal of Classification*, *20*, 221–255.

Li, L., & Hser, Y. (2011). On inclusion of covariates for class enumeration of growth mixture models. *Multivariate Behavioral Research*, *46*, 266–302.

Lipsitz, S. R., & Ibrahim, J. G. (1996). A conditional model for incomplete covariates in parametric regression models. *Biometrika*, *83*, 916–922.

Little, R. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association*, *87*, 1227–1237.

Little, R., & Rubin, D. (1983). On jointly estimating parameters and missing data by maximizing the complete-data likelihood. *The American Statistician*, *37*, 218–220.

Little, R., & Zhang, N. (2011). Subsample ignorable likelihood for regression analysis with missing data. *Applied Statistics*, *60*, 591–605.

Lu, Z., Zhang, Z., & Lubke, G. (2011). Bayesian inferences for growth mixture models with latent class dependent missing data. *Multivariate Behavioral Research*, *46*, 567–597.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*, 21–39.

Lubke, G., & Muthén, B. (2007). Performance of factor mixture models as a function of model size, covariate effects, and class-specific parameters. *Structural Equation Modeling*, *14*, 26–47.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: University of Cambridge Press.

Marlin, B. M. (2008). *Missing data problems in machine learning* (Unpublished doctoral dissertation). University of Toronto, Canada.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.

McMillen, J., Zima, B., Scott, L., Auslander, W., Munson, M., Ollie, M., & Spitznagel, E. (2005). Prevalence of psychiatric disorders among older youths in the foster care system. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*, 88–95.

Merkle, E. (2011). A comparison of imputation methods for Bayesian factor analysis models. *Journal of Educational and Behavioral Statistics*, *36*, 257–276.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*, 156–166.

Muthén, B. O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika*, *29*, 81–117.

Muthén, B. O. (2003). Statistical and substantive checking in growth mixture modeling: Comment on Bauer (2003). *Psychological Methods*, *8*, 369–377.

Muthén, B. O., & Asparouhov, T. (2009). Growth mixture modeling: Analysis with non-Gaussian random effects. In F. G. Davidian, M. G. Verbeke, & G. Molenberghs (Eds.), *Longitudinal data analysis* (pp. 143–165). Boca Raton, FL: Chapman & Hall/CRC Press.

Muthén, B. O., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*, 463–469.

Muthén, L. K., & Muthén, B. O. (1998–2014). *Mplus users guide* (7th ed.). Los Angeles, CA: Author.

Nagin, D. S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Nagin, D. S., & Odgers, C. L. (2010). Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, *6*, 109–138.

Nagin, D. S., & Tremblay, R. E. (2001). Analyzing developmental trajectories of distinct but related behaviors: A group-based method. *Psychological Methods*, *6*, 18–34.

Odgers, C., Moffitt, T., Broadbent, J., Dickson, N., Hancox, R., Harrington, H., . . ., Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology*, *20*, 673–716.

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. Piquero & D. Weisburd (Eds.), *Handbook of quantitative criminology* (pp. 69–100). New York, NY: Springer.

Pickles, A., & Croudace, T. (2010). Latent mixture models for multivariate and longitudinal outcomes. *Statistical Methods in Medical Research*, *19*, 271–289.

Rabe-Hesketh, S., & Skrondal, A. (2014, May). *Simple methods for handling non-randomly missing data*. Paper presented at the Modern Modeling Methods Conference, Storrs, CT.

Raudenbush, S. W., & Bryk, A. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Reinartz, W., Echambadi, R., & Chin, W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioral Research*, *37*, 227–244.

Reinecke, J. (2006). Longitudinal analysis of adolescents' deviant and delinquent behavior: Applications of latent class growth curves and growth mixture models. *Methodology*, *2*, 100–112.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*, 581–592.

Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research*, *43*, 355–381.

Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, *57*, 316–331.

Schaeffer, C., Petras, H., Ialongo, N., Masyn, K., Hubbard, S., Poduska, J., & Kellam, S. (2006). A comparison of girls' and boys' aggressive-disruptive behavior trajectories across elementary school: Prediction to young adult antisocial outcomes. *Journal of Consulting and Clinical Psychology*, *74*, 500–510.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. London, UK: Chapman & Hall.

Scharoun-Lee, M., Gordon-Larsen, P., Adair, L., Popkin, B., Kaufman, J., & Suchindran, C. (2011). Intergenerational profiles of socioeconomic (dis)advantage and obesity during the transition to adulthood. *Demography*, *48*, 625–651.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of the Royal Statistical Society, Series B*, *62*, 795–809.

Sterba, S. K. (2013). Understanding linkages among mixture models. *Multivariate Behavioral Research*, *48*, 775–815.

Sterba, S. K., Baldasaro, R. E., & Bauer, D. J. (2012). Factors affecting the adequacy and preferability of semiparametric groups-based approximations of continuous growth trajectories. *Multivariate Behavioral Research*, *40*, 590–634.

Sterba, S. K., & Bauer, D. J. (2014). Predictions of individual change recovered with latent class or random coefficient growth models. *Structural Equation Modeling*, *21*, 342–360.

Sterba, S. K., Prinstein, M. J., & Cox, M. J. (2007). Trajectories of internalizing problems across childhood: Heterogeneity, external validity, and gender differences. *Development and Psychopathology*, *19*, 345–366.

Van Horn, M., Jaki, T., Masyn, K., Ramey, S., & Smith, J. (2009). Assessing differential effects: Applying regression mixture models to identify variations in the influence of family resources on academic achievement. *Developmental Psychology*, *45*, 1298–1313.

Vermunt, J. (2010a). Longitudinal research using mixture models. In K. Van Monfort, J. Oud, & A. Sattora (Eds.), *Longitudinal research with latent variables* (pp. 119–152). Heidelberg, Germany: Springer.

Vermunt, J. (2010b). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*, 450–469.

Vermunt, J., & Magidson, J. (2005). *Technical guide for Latent Gold 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Vermunt, J., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research design, measurement, and analysis* (pp. 373–385). Burlington, MA: Elsevier.

von Oertzen, T. (2010). Power equivalence in structural equation modelling. *British Journal of Mathematical and Statistical Psychology*, *63*, 257–272.

von Oertzen, T., & Brandmaier, A. (2013). Optimal study design with identical power: An application of power equivalence to latent growth curve models. *Psychology & Aging*, *28*, 414–428.

Walrath, C., Petras, H., Mandell, D., Stephens, R., Holden, E., & Leaf, P. (2004). Gender differences in patterns of risk factors among children receiving mental health services: Latent class analyses. *Journal of Behavioral Health Services & Research*, *31*, 297–311.

Wedel, M. (2002). Concomitant variables in finite mixture models. *Statistica Neerlandica*, *56*, 362–375.

Wedel, M., & DeSarbo, W. (1995). A mixture likelihood approach for generalized linear models. *Journal of Classification*, *12*, 21–55.

Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. Little, K. Schnabel, & J. Baumert (Eds), *Testing structural equation models* (pp. 256–293). Mahwah, NJ: Erlbaum.

Yuncel, R., He, Y., & Zaslavsky, A. (2011). Gaussian-based routines to impute categorical variables in heath surveys. *Statistics in Medicine*, *30*, 3447–3460.

APPENDIX A

Here an example set of special case criteria (Ingrassia, Minotti, & Vittadini, 2012) is reviewed under which parameters of a particular conditional mixture (e.g., Figure 1 Panel C) could be solved for from parameters of a particular fitted conventional multivariate (here, *normal*) mixture, assuming both models have the same within-class specification of $f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)$ and the same number of predictors, p. These criteria are as follows:

1. Within class, $\mathbf{y}_i$ *and* $\mathbf{x}_i$ are multivariate normally distributed.
2. The covariance matrix of $\mathbf{x}_i$ is class invariant (here denoted $\mathbf{\Sigma}^{(k)} = \mathbf{\Sigma}$). Note this is an assumption about the data, *not* a parameter constraint.
3. Class proportions are balanced (i.e., $p(c_i = k) = (1/K)$).

It is important to note that these criteria would have to hold exactly in a given sample data set for this conversion to be accurate using actual parameter estimates. Hence, the conversion in Appendix A is of theoretical rather than practical interest. To see why the aforementioned three criteria are required for this conversion, we may start with Equation (8):

$$
\begin{aligned}
f(\mathbf{y}_i, \mathbf{x}_i) &= \sum_{k=1}^{K} f(\mathbf{y}_i|\mathbf{x}_i; c_i = k) f(\mathbf{x}_i|c_i = k) p(c_i = k) \\
&= f(\mathbf{x}_i) \sum_{k=1}^{K} f(\mathbf{y}_i|\mathbf{x}_i; c_i = k) \frac{f(\mathbf{x}_i|c_i = k) p(c_i = k)}{f(\mathbf{x}_i)}
\end{aligned} \tag{15}
$$

The remaining steps do not concern $f(\mathbf{y}_i|\mathbf{x}_i; c_i = k)$; under the assumption of within-class multivariate normality, its parameters in Equation (2) would be unchanged.

Using Bayes's rule (see also Li & Hser, 2011),

$$
\begin{aligned}
p(c_i = k|\mathbf{x}_i) &= \frac{f(\mathbf{x}_i|c_i = k) p(c_i = k)}{f(\mathbf{x}_i)} \\
&= \frac{f(\mathbf{x}_i|c_i = k) p(c_i = k)}{\sum_{k=1}^{K} f(\mathbf{x}_i|c_i = k) p(c_i = k)}.
\end{aligned}
$$

Elaborating a sketched proof from Ingrassia et al. (2012, Proposition 4), under the assumptions of $p(c_i = k) = (1/K)$ and within-class normality, we have

$$
\begin{aligned}
p(c_i = k|\mathbf{x}_i) &= \frac{f(\mathbf{x}_i|c_i = k)(1/K)}{(1/K)\sum_{k=1}^{K} f(\mathbf{x}_i|c_i = k)} = \frac{f(\mathbf{x}_i|c_i = k)}{\sum_{k=1}^{K} f(\mathbf{x}_i|c_i = k)} \\
&= \frac{\frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}^{(k)}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})'(\mathbf{\Sigma}^{(k)})^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right]\right\}}{\sum_{k=1}^{K} \frac{1}{(2\pi)^{p/2}|\mathbf{\Sigma}^{(k)}|^{1/2}} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})'(\mathbf{\Sigma}^{(k)})^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right]\right\}}
\end{aligned}
$$

Under the assumption $\mathbf{\Sigma}^{(k)} = \mathbf{\Sigma}$ we may simplify as follows:

$$
p(c_i = k|\mathbf{x}_i) = \frac{\exp\left\{-\frac{1}{2}\left[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right]\right\}}{\sum_{k=1}^{K} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})\right]\right\}}.
$$

For shorthand, let $\nu^{(k)} = -\frac{1}{2}[(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})'\mathbf{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^{(k)})]$. Consider the illustrative situation of $K = 2$. For the second class, the conditional probability of class membership is

$$
\begin{aligned}
p(c_i = 2|\mathbf{x}_i) &= \frac{\exp(\nu^{(2)})}{\exp(\nu^{(1)}) + \exp(\nu^{(2)})} = \frac{1}{1 + \frac{\exp(\nu^{(1)})}{\exp(\nu^{(2)})}} \\
&= \frac{1}{1 + \exp(\nu^{(1)} - \nu^{(2)})},
\end{aligned} \tag{16}
$$

where $\nu^{(1)} - \nu^{(2)}$ is equal to

$$-\frac{1}{2}\left(\left[(\mathbf{x}_i-\boldsymbol{\mu}^1)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}^1)\right]-\left[(\mathbf{x}_i-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i-\boldsymbol{\mu}^2)\right]\right)$$

$$=-\frac{1}{2}\big(\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\mathbf{x}_i-\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\mathbf{x}_i+\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2$$
$$-\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i+\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2+\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\big)$$

$$=-\frac{1}{2}\big(\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2-\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i$$
$$+\mathbf{x}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2+\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\big)$$

$$=-\frac{1}{2}\big(\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2+(\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2-\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1)$$
$$-\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i-\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i+\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i+\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\big)$$

$$=-\frac{1}{2}\big((\boldsymbol{\mu}^{1\prime}-\boldsymbol{\mu}^{2\prime})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1-(\boldsymbol{\mu}^{2\prime}-\boldsymbol{\mu}^{1\prime})\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2$$
$$-2(\boldsymbol{\mu}^{1\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i-\boldsymbol{\mu}^{2\prime}\boldsymbol{\Sigma}^{-1}\mathbf{x}_i)\big)$$

$$=-\frac{1}{2}\left((\boldsymbol{\mu}^{1\prime}-\boldsymbol{\mu}^{2\prime})(\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^1+\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}^2)-2(\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\right)$$

$$=-\frac{1}{2}\left((\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^1+\boldsymbol{\mu}^2)-2(\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}\mathbf{x}_i\right)$$

$$=-\frac{1}{2}\left((\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^1+\boldsymbol{\mu}^2)\right)+\left((\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}\right)\mathbf{x}_i.$$

Letting $\alpha_0^{(1)}=-\frac{1}{2}\left((\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}^1+\boldsymbol{\mu}^2)\right)$ and $\boldsymbol{\omega}^{(1)\prime}=\left((\boldsymbol{\mu}^1-\boldsymbol{\mu}^2)'\boldsymbol{\Sigma}^{-1}\right)$ substituting back in to Equation (15) yields $p(c_i=2|\mathbf{x}_i)=\frac{1}{1+\exp(\alpha_0^{(1)}+\boldsymbol{\omega}^{(1)\prime}\mathbf{x}_i)}$ and $p(c_i=1|\mathbf{x}_i)=\frac{\exp(\alpha_0^{(1)}+\boldsymbol{\omega}^{(1)\prime}\mathbf{x}_i)}{1+\exp(\alpha_0^{(1)}+\boldsymbol{\omega}^{(1)\prime}\mathbf{x}_i)}$ and thus we have solved for the multinomial regression coefficients from Equation (1) under the three special case criteria mentioned earlier.

Thus far, Appendix A has described three special case criteria under which multinomial coefficients of the conditional mixture in Figure 1 Panel C could be computed from parameters of a conventional multivariate normal mixture. Suppose that a researcher had instead wanted to solve for multinomial coefficients of the conditional mixture in Figure 1 *Panel A*. In the aforementioned proof, $f(\mathbf{y}_i|\mathbf{x}_i;c_i=k)$ would need to be replaced by $f(\mathbf{y}_i;c_i=k)$ and the three special case criteria mentioned earlier would again need to hold.

APPENDIX B

This appendix provides the Mplus syntax for the exogenous-*x,* conventional multivariate, and endogenous-constrained-*x* mixture approaches in the Table 1 illustration.

```
!Syntax for Exogenous-x Approach
MODEL:
%overall%
y on x1 x2 x3;
%class#1%
[y]; y; y on x1 x2 x3;
%class#2%
[y]; y; y on x1 x2 x3;
!Syntax for Conventional multivariate Approach
MODEL:
%overall%
y on x1 x2 x3;
%class#1%
[y]; y; y on x1 x2 x3;
x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6);
x1 with x2 (7); x1 with x3 (8); x2 with x3 (9);
%class#2%
[y]; y; y on x1 x2 x3;
x1 (11); x2 (12); x3 (13); [x1] (14); [x2] (15); [x3] (16);
x1 with x2 (17); x1 with x3 (18); x2 with x3 (19);
!Syntax for Endogenous-constrained-x Approach
MODEL:
%overall%
y on x1 x2 x3;
%class#1%
[y]; y; y on x1 x2 x3;
x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6);
x1 with x2 (7); x1 with x3 (8); x2 with x3 (9);
%class#2%
[y]; y; y on x1 x2 x3;
x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6);
x1 with x2 (7); x1 with x3 (8); x2 with x3 (9);
```

Mplus syntax for the likelihood ratio test (LRT) illustration evaluating the effect of **x** on *y* within class using the endogenous-constrained-*x* approach.

!Syntax for Model A (*no* effect of x on y within class)	!Syntax for Model B (effect of x on y within class allowed)
MODEL: %overall% %class#1% [y]; y; x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6); x1 with x2 (7); x1 with x3 (8); x2 with x3 (9); %class#2% [y]; y; x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6); x1 with x2 (7); x1 with x3 (8); x2 with x3 (9);	MODEL: %overall% y on x1 x2 x3; %class#1% [y]; y; y on x1 x2 x3; x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6); x1 with x2 (7); x1 with x3 (8); x2 with x3 (9); %class#2% [y]; y; y on x1 x2 x3; x1 (1); x2 (2); x3 (3); [x1] (4); [x2] (5); [x3] (6); x1 with x2 (7); x1 with x3 (8); x2 with x3 (9);