M*plus*

Statistical Analysis With Latent Variables

Technical Appendices

Bengt O. Muthén

Following is the correct citation for this document:

Muthén, B.O. (1998-2004). Mplus Technical Appendices.
Los Angeles, CA: Muthén & Muthén

Muthén & Muthén
3463 Stoner Avenue
Los Angeles, CA 90066
Tel: (310) 391-9971
Fax: (310) 391-8971
Web: www.StatModel.com
Support@StatModel.com

TABLE OF CONTENTS

APPENDIX 1
REGRESSION WITH A CATEGORICAL DEPENDENT VARIABLE

LOGIT AND PROBIT REGRESSION. ODDS AND ODDS RATIOS

Consider as an example the logistic regression for a binary y variable scored 0 and 1 which is regressed on a binary x_1 variable scored 0 and 1 and a continuous x_2 variable. Define π as the probability of $y = 1$. The model expresses the probability by the logistic function (cf. Hosmer and Lemeshow, 1989),

$$P(y = 1|x) = \pi|x = \frac{1}{1 + e^{-(\beta_0 + \beta_1\ x_1 + \beta_2\ x_2)}} \tag{1}$$

where the βs are logistic regression coefficients. Alternatively, we may express the model as a linear logit equation

$$logit(\pi|x) = \beta_0 + \beta_1\ x_1 + \beta_2\ x_2 \tag{2}$$

where $logit(\pi) = log(\pi/(1 - \pi))$. The ratio $\pi/(1 - \pi)$ is referred to as the odds for $y = 1$ versus $y = 0$. In addition, odds ratios and adjusted odds ratios can be considered. The idea of an adjusted odds ratio is described next.

It is of interest to study the change in the probability of $y = 1$ as a function of the binary variable x_1. Let π_0 denote the probability of $y = 1$ for $x_1 = 0$ and let π_1 denote the probability of $y = 1$ for $x_1 = 1$. We have

$$logit(\pi_0|x) = \beta_0 + \beta_2\ x_2 \tag{3}$$

and

$$logit(\pi_1|x) = \beta_0 + \beta_1 + \beta_2\ x_2. \tag{4}$$

The log odds ratio for y and x_1 adjusted for x_2 is

$$log\ OR = log[\frac{\pi_1/(1 - \pi_1)}{\pi_0/(1 - \pi_0)}|x] = logit(\pi_1|x) - logit(\pi_0|x) = \beta_1 \tag{5}$$

so that $OR = exp(\beta_1)$. This OR value is therefore constant for all values of x_2 and can be interpreted as a measure of the effect of x_1 on y (this is analogous to ANCOVA for a continuous y, where the effect of x_1 is measured as a constant intercept difference between the regression lines for the two x_1 groups). If an interaction term for x_1 and x_2 is introduced, the constancy of the OR no longer holds.

Probit regression considers

$$P(y = 1|x) = \pi|x = \Phi(\beta_0 + \beta_1\ x_1 + \beta_2\ x_2) \tag{6}$$

1

APPENDIX 1

where Φ is the standard normal distribution function. Using the inverse normal function Φ^{-1}, the model can be written in line with (2) as a linear probit equation

$$\Phi^{-1}(\pi|x) = \beta_0 + \beta_1 \, x_1 + \beta_2 \, x_2. \tag{7}$$

In this case,

$$log \, OR = log[\frac{\Phi_1/(1 - \Phi_1)}{\Phi_0/(1 - \Phi_0)}, \tag{8}$$

where Φ_1 is the Φ expression of (6) evaluated at $x_1 = 1$ and Φ_0 is evaluated at $x_1 = 0$. Note that each Φ expression of (8) is an integral over a normal density. The ratios of these integral expressions cannot be expressed in simple terms of model parameters. Unlike (5), this $log \, OR$ function is not constant over the values of x_2 and is not a function of β_1 alone, but also depends on β_0 and β_2.

LATENT RESPONSE VARIABLE FORMULATION VERSUS PROBABILITY CURVE FORMULATION

As shown above, the probit and logit regression models are usually presented in terms of the conditional probability of y given x,

$$Prob(y = 1|x) = F(\alpha + \beta x) \tag{9}$$

where F is either the standard normal or logistic distribution function. This is a non-linear function of x which varies between 0 and 1.

In a general modeling framework it is convenient to consider latent, continuous response variables y^*. Corresponding to the binary case in (9), the latent response variable formulation defines a threshold τ on y^* so that $y = 1$ is observed when y^* exceeds τ while otherwise $y = 0$ is observed. A linear regression equation is used to relate y^* to x,

$$y^* = \pi \, x + \delta \tag{10}$$

where π is a slope parameter and δ is a residual that is uncorrelated with x. An intercept term is not needed because of the threshold parameter τ.

Normality is assumed for the δ residual, $\delta \sim N(0, V(\delta))$. This latent response variable formulation results in the same model as (9), with F taken to be the standard normal distribution function Φ,

$$Prob(y = 1|x) = Prob(y^* > \tau|x) = 1 - Prob(y^* \le \tau|x) = 1 - \Phi[(\tau - \pi x)V(\delta)^{-1/2}]. \tag{11}$$

Standardizing to $V(\delta) = 1$ this defines a probit model with $\alpha = -\tau$ and $\beta = \pi$.

Alternatively, a logistic density may be assumed for δ,

$$f[\delta; 0, \pi^2/3] = dF/d\delta = F(1 - F), \tag{12}$$

2

where in this case F is the logistic distribution function $1/(1 + e^{-\delta})$. This also results in (9).

The latent response variable formulation focuses on the linear relation between y^* and x instead of the non-linear relationship between y and x. This is in agreement with considering logit and probit values as linear functions of x.

The use of latent response variables allows a convenient way of describing more general models than regression. It is possible to maintain linear relationships among all variables in the Mplus framework $(y^*, x, \eta, \epsilon, \zeta)$, whereas (9) presents a non-linear relationship between y and x.

While the conventional formulation focuses on changes in probabilities of the binary y, the y^* formulation focuses on changes in the values of the continuous variable y^*. Nevertheless, the y^* formulation can also be used to derive the resulting changes in the probabilities of the binary y.

To describe changes in the latent response variable y^* as a function of x it is convenient to consider the standardized form of (10), where

$$\pi_{standardized} = \pi \times s.d.(x)/s.d(y^*), \tag{13}$$

where $s.d.(.)$ denotes the standard deviation and

$$s.d.(y^*) = \sqrt{\pi^2 V(x) + V(\delta)}. \tag{14}$$

One may follow the convention of not standardizing with respect to the variance of an x variable which is binary. This means that in this case the coefficient refers to a standard deviation change in y^* for a change in x from 0 to 1.

An R^2 value for y^* can also be defined,

$$R^2 = [\pi^2 V(x)]/[\pi^2 V(x) + V(\delta)]. \tag{15}$$

This R^2 value is different from an R^2 of regression with a continuous y because $V(\delta)$ is not a free parameter. Nevertheless, this is a useful summary. For other R^2-like quantities proposed for binary response variables, e.g., see Amemiya (1981, pp. 1503-1507).

PROBIT AND LOGIT WITH AN ORDERED POLYTOMOUS DEPENDENT VARIABLE

The concept of a latent response variable y^* is useful for defining a categorical variable y with C ordered categories,

$$y = c, \; if \quad \tau_c < y^* \leq \tau_{c+1} \tag{16}$$

for categories $c = 0, 1, 2, \ldots, C - 1$ and $\tau_0 = -\infty$, $\tau_C = \infty$. The probit and logit regression can be generalized as follows for the ordered polytomous case. For simplicity, an example

APPENDIX 1

with a single x variable and a y variable with three categories will be considered. Here, there are two threshold parameters to be estimated, τ_1 and τ_2.

In probit regression, the probabilities for the three outcomes can be derived from the latent response variable regression

$$y^* = \pi\, x + \delta, \tag{17}$$

where $y^*|x \sim N(\pi\, x, V(\delta))$, where $V(\delta)$ is standardized to one. Taken together with (16), this gives the conditional probabilities for the three categories of y,

$$P(y = 0|x) = \Phi(\tau_1 - \pi\, x), \tag{18}$$
$$P(y = 1|x) = \Phi(\tau_2 - \pi\, x) - \Phi(\tau_1 - \pi\, x), \tag{19}$$
$$P(y = 2|x) = 1 - \Phi(\tau_2 - \pi\, x) = \Phi(-\tau_2 + \pi\, x). \tag{20}$$

This implies that

$$P(y = 1 \text{ or } 2|x) = P(y = 1|x) + P(y = 2|x) \tag{21}$$
$$= 1 - \Phi(\tau_1 - \pi\, x) \tag{22}$$
$$= \Phi(-\tau_1 + \pi\, x) \tag{23}$$
$$= 1 - P(y = 0|x), \tag{24}$$

resulting in a probit expression for the highest category and for the two highest categories,

$$P(y = 2|x) = \Phi(-\tau_2 + \pi\, x), \tag{25}$$
$$P(y = 1 \text{ or } 2|x) = \Phi(-\tau_1 + \pi\, x). \tag{26}$$

The characteristic feature of modeling ordered categories is that these two expressions share the same slope π. This means that the corresponding conditional probability curves expressed as functions of x are parallell and only differ due to the thresholds.

Logit regression with an ordered polytomous y variable uses a model analogous to (25) and (26), but bases it on the logistic function,

$$P(y = 2|x) = \frac{1}{1 + e^{-(\beta_2 + \beta\, x)}}, \tag{27}$$

$$P(y = 1 \text{ or } 2|x) = \frac{1}{1 + e^{-(\beta_1 + \beta\, x)}}. \tag{28}$$

This is the proportional odds model (see, e.g., Agresti, 1990, pp. 322-324). As in the binary case, the logit, or log odds, for each of these two events is a linear expression,

$$logit[P(y = 2|x)] = log[P(y = 2|x)/(1 - P(y = 2|x)] = \beta_2 + \beta\, x, \tag{29}$$
$$logit[P(y = 1 \text{ or } 2|x)] = log[P(y = 1 \text{ or } 2|x)/(1 - P(y = 1 \text{ or } 2|x)] = \beta_1 + \beta\, x. \tag{30}$$

When x is a 0/1 variable,

$$logit[P(y = 2|x = 1)] - logit[P(y = 2|x = 0)] = \beta, \qquad (31)$$
$$logit[P(y = 1 \; or \; 2|x = 1)] - logit[P(y = 1 \; or \; 2|x = 0)] = \beta, \qquad (32)$$

showing that the ordered polytomous logistic regression model has constant odds ratios for these different outcomes.

APPENDIX 1

APPENDIX 2
THE GENERAL MODELING FRAMEWORK

VARIABLES

In the general modeling framework it is useful to distinguish between two types of variables. In the body of the User's Guide these are referred to as dependent and independent variables, while in the Appendix they are referred to by their technical terms, y variables and x variables.

Several observed y measurement scales are allowed in the Mplus framework. To describe these it is convenient to relate each of the p observed dependent variables y to a corresponding continuous latent response variable y^*. The use of a latent response variables to describe the relationship between a categorical variables y_j and other variables in the model is shown in Appendix 1. Let the subscript $i(i = 1, 2, \ldots, n)$ refer to the observational unit (the individual) and the subscript $j(j = 1, 2, \ldots, p)$ refer to the observed dependent variable.

A continuous variable $y_j(j = 1, 2, \ldots, p)$ is defined as

$$y_{ij} = y_{ij}^*. \tag{33}$$

A binary variable $y_j(j = 1, 2, \ldots, p)$ is defined as

$$y_{ij} = \begin{cases} 1 & \text{if } y_{ij}^* > \tau_j \\ 0 & otherwise. \end{cases} \tag{34}$$

A categorical variable $y_j(j = 1, 2, \ldots, p)$ with C ordered categories is defined as

$$y_{ij} = c, \; if \quad \tau_{j,c} < y_{ij}^* \leq \tau_{j,c+1} \tag{35}$$

for categories $c = 0, 1, 2, \ldots, C - 1$ and $\tau_0 = -\infty$, $\tau_C = \infty$.

STATISTICAL MODEL

The continuous latent variable structural equation model used in the Mplus framework is expressed in two parts: a measurement part and a structural part (cf. Bollen, 1989). Multiple populations, or groups, are allowed for so that parameters can be defined for each group. The following description of the model draws on Muthén (1979, 1983, 1984, 1989b).

The measurement part of the model is defined in terms of the p-dimensional latent response variable vector $\boldsymbol{y}^*$,

$$\boldsymbol{y}^*{}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\, \boldsymbol{\eta}_i + \mathbf{K}\, \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{36}$$

7

where $\boldsymbol{\eta}$ is an m-dimensional vector of latent variables (constructs or factors), $\mathbf{x}$ is a q-dimensional vector of independent (background) variables, $\boldsymbol{\epsilon}$ is a p-dimensional vector of residual or measurement errors which is uncorrelated with other variables, $\boldsymbol{\nu}$ is a p-dimensional parameter vector of measurement intercepts, $\boldsymbol{\Lambda}$ is a $p \times m$ parameter matrix of measurement slopes or factor loadings, and $\mathbf{K}$ is a $p \times q$ parameter matrix of regression slopes. Usually, only a few of the rows of $\mathbf{K}$ are nonzero, where a non-zero row corresponds to a y variable that is directly influenced by one or more x variables. The covariance matrix of $\boldsymbol{\epsilon}$ is denoted $\boldsymbol{\Theta}$.

The structural part of the model is defined in terms of the latent variables regressed on each other and the q-dimensional vector $\mathbf{x}$ of independent variables ,

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\,\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\,\mathbf{x}_i + \boldsymbol{\zeta}_i. \tag{37}$$

Here, $\boldsymbol{\alpha}$ is an m-dimensional parameter vector, $\mathbf{B}$ is an $m \times m$ parameter matrix of slopes for regressions of latent variables on other latent variables. $\mathbf{B}$ has zero diagonal elements and it is assumed that $\mathbf{I} - \mathbf{B}$ is non-singular. Furthermore, $\boldsymbol{\Gamma}$ is an $m \times q$ slope parameter matrix for regressions of the latent variables on the independent variables, and $\boldsymbol{\zeta}$ is an m-dimensional vector of residuals. The covariance matrix of $\boldsymbol{\zeta}$ is denoted $\boldsymbol{\Psi}$.

The parameter arrays $\boldsymbol{\nu}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{\Theta}$ correspond to the program's BY statements. The parameter arrays $\mathbf{K}$, $\boldsymbol{\alpha}$, $\mathbf{B}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\Psi}$ correspond to the program's ON statements.

The Mplus framework assumes conditional normality for $\boldsymbol{y}^*$ given $\mathbf{x}$, so that it suffices to consider the conditional expectation and conditional variance,

$$E(\boldsymbol{y}^*{}_{si}|\mathbf{x}_i) = \boldsymbol{\Delta}[\boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\,\mathbf{x}_i + \mathbf{K}\,\mathbf{x}_i], \tag{38}$$

$$V(\boldsymbol{y}^*{}_{si}|\mathbf{x}_i) = \boldsymbol{\Delta}[\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]\boldsymbol{\Delta}. \tag{39}$$

where the latent response variables have been scaled as

$$\boldsymbol{y}^*{}_{si} = \boldsymbol{\Delta}\,\boldsymbol{y}^*{}_i, \tag{40}$$

defining p additional scaling parameters in the diagonal scaling matrix $\boldsymbol{\Delta}$. The use of the $\boldsymbol{\Delta}$ is discussed below.

When $\mathbf{y}$ contains categorical variables, the conditional normality assumption is an important feature. This feature avoids the more restrictive assumption of full multivariate normality for $(\mathbf{y}', \mathbf{x}')$ which is used for polyserial and polychoric correlations. Full multivariate normality implies an assumption of multivariate normality for $\mathbf{x}$ which is often not realistic. When x variables are present, the conditional normality assumption allows non-normality for $\mathbf{y}^*$ as a function of non-normal x variables.

THE SCALING PARAMETERS OF $\boldsymbol{\Delta}$

In conventional analyses, the scaling matrix $\boldsymbol{\Delta}$ is set to $\mathbf{I}$ and has no impact. The scaling matrix makes it possible to relate covariance and correlation structures. It can be used in two major ways, to analyze sample correlations with a covariance structure and to analyze sample covariances with a correlation structure. When all variables of $\boldsymbol{y^*}$ correspond to categorical y variables, $\boldsymbol{\Delta}$ is used to fit the correlation structure $V(\boldsymbol{y^*}_{si}|\mathbf{x}_i)$ to sample correlations whereas

$$\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} \tag{41}$$

represents the corresponding covariance structure. In this case, the diagonal of $\boldsymbol{\Delta}$ contains the inverted standard deviations of the conditional $\boldsymbol{y^*}$ variances given $\mathbf{x}$,

$$diag[\boldsymbol{\Delta}] = diag[V(\boldsymbol{y^*}|\mathbf{x})]^{-1/2} = diag[\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]^{-1/2}. \tag{42}$$

Here, $\boldsymbol{y^*}_i$ contains variables in their original metric and $\boldsymbol{y^*}_{si}$ is standardized to unit variances. The diagonal of $V(\boldsymbol{y^*}_{si}|\mathbf{x}_i)$ does not enter into the analysis given that correlations are analyzed. Because of this, diagonal elements of $\boldsymbol{\Theta}$ are not separately identified parameters. They can therefore be chosen so that $\boldsymbol{\Delta} = \mathbf{I}$, which is the standardization used in probit regression. The probit standardization is further discussed in Appendix 1. In this case, diagonal $\boldsymbol{\Theta}$ elements are obtained as the remainder

$$diag[\boldsymbol{\Theta}] = \mathbf{I} - diag[\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}']. \tag{43}$$

The diagonal elements of $\boldsymbol{\Delta}$ are useful when comparing the same y variables over time or across groups. In such cases, the $\boldsymbol{\Delta}$ element for the first time point or the first group can be standardized to one, whereas $\boldsymbol{\Delta}$ elements can be estimated for other time points or groups to capture differences in $\boldsymbol{y^*}$ variances over time or across groups. As opposed to the diagonal elements of $\boldsymbol{\Theta}$, the diagonal elements of $\boldsymbol{\Delta}$ do contribute to the off-diagonal elements of $V(\boldsymbol{y^*}_{si}|\mathbf{x}_i)$ and therefore do enter into the analysis. When $\boldsymbol{\Delta}$ is not set to the identity matrix, (43) changes to the general expression

$$diag[\boldsymbol{\Theta}] = \boldsymbol{\Delta}^{-2} - diag[\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}'], \tag{44}$$

where the diagonal of $\boldsymbol{\Delta}^{-2}$ represents the variances of $V(\boldsymbol{y^*}|\mathbf{x})$. The *unc*onditional variances of $\boldsymbol{y^*}$ are given by

$$diag[V(\boldsymbol{y^*})] = diag[\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\,\boldsymbol{\Gamma}\,\boldsymbol{\Sigma}_{xx}\,\boldsymbol{\Gamma}'\,(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Lambda}(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}-\mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]. \tag{45}$$

When some of the variables of $\boldsymbol{y^*}$ correspond to categorical y variables and some to continuous y variables, diagonal elements of $\boldsymbol{\Delta}$ are as described above for the categorical y variables, whereas diagonal elements corresponding to continuous y variables are typically set at one.

APPENDIX 2

When all variables of $\boldsymbol{y}^*$ correspond to continuous y variables, $\boldsymbol{\Delta}$ can be used to analyze a sample covariance matrix with a correlation structure using a reversal of the approach given above. The correlation structure is

$$\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \tag{46}$$

whereas $V(\boldsymbol{y}^*{}_{si}|\mathbf{x}_i)$ represents the corresponding covariance matrix fitted to the sample covariance matrix. In this case, the diagonal of $\boldsymbol{\Delta}$ contains the standard deviations of $\mathbf{y}$ so that $\boldsymbol{y}^*{}_{si}$ contains variables in their original metric, $\boldsymbol{y}^*{}_{si} = \mathbf{y}_i$, and $\boldsymbol{y}^*{}_i$ is standardized to unit variances.

THE THRESHOLD PARAMETERS OF $\boldsymbol{\tau}$

With categorical y variables, threshold parameters enter into the model. While many models can be fitted by correlation structures alone, threshold parameters are needed in models where measurement characteristics of variables are compared over time as in longitudinal models, or across groups.

As an example of how threshold parameters enter into the model, consider a model with two binary y variables and a set of x variables. Using the general model of (36) and (37), the regression of $\mathbf{y}^*$ on $\mathbf{x}$ may be written as

$$\mathbf{y}_i^* = \boldsymbol{\nu} + \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\alpha} + \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\Gamma}\,\mathbf{x}_i + \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\zeta}_i + \mathbf{K}\,\mathbf{x}_i + \boldsymbol{\epsilon}_i \tag{47}$$

$$= \boldsymbol{\pi}_0 + \boldsymbol{\Pi}\,\mathbf{x}_i + \boldsymbol{\delta}_i. \tag{48}$$

The term $\boldsymbol{\nu}$ is not separately identified from the threshold parameters and is typically fixed at zero. Here,

$$V(\boldsymbol{\delta}_i) = V(\boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\zeta}_i + \boldsymbol{\epsilon}_i) = \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta} = \boldsymbol{\Omega}. \tag{49}$$

In line with (38) and (39), the modeling considers the conditional mean vector and covariance matrix of the scaled latent response variables,

$$E(\boldsymbol{y}^*{}_s|\mathbf{x}) = \boldsymbol{\Delta}[\boldsymbol{\pi}_0 + \boldsymbol{\Pi}\,\mathbf{x}] = \boldsymbol{\mu}^*(\mathbf{x}), \tag{50}$$

$$V(\boldsymbol{y}^*{}_s|\mathbf{x}) = \boldsymbol{\Delta}\,\boldsymbol{\Omega}\,\boldsymbol{\Delta} = \boldsymbol{\Sigma}^*, \tag{51}$$

where $\boldsymbol{\Sigma}^*$ has unit diagonal elements. Given the conditional normality assumption this leads to the univariate and bivariate probability expressions

$$P(y_j = 1|\mathbf{x}) = \int_{\tau_j^* - \mu_j^*(\mathbf{x})}^{\infty} \phi_1(y_j^*|\mathbf{x})\,dy_j^*, \tag{52}$$

and

$$P(y_j = 1, y_k = 1|\mathbf{x}) = \int_{\tau_j^* - \mu_j^*(\mathbf{x})}^{\infty} \int_{\tau_k^* - \mu_k^*(\mathbf{x})}^{\infty} \phi_2(y_j^*, y_k^*|\mathbf{x})\,dy_k^*\,dy_j^*, \tag{53}$$

10

where τ_j^* denotes the threshold parameter for y_j^* multiplied by the jth diagonal element of Δ, ϕ_1 denotes a univariate standard normal density, ϕ_2 denotes a bivariate normal density with unit variances, zero means, and correlation coefficient σ_{jk}^*, where σ_{jk}^* is an off-diagonal element of Σ^*. The off-diagonal elements of Σ^* are referred to as probit residual correlations. The elements of $\Delta\Pi$ are referred to as probit slopes.

THE EXPLORATORY FACTOR ANALYSIS MODEL

The model used for exploratory factor analysis is a special case of the general model of (36) and (37),

$$\boldsymbol{y^*}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i, \tag{54}$$

with

$$V(\boldsymbol{y^*}) = \boldsymbol{\Lambda}\,\boldsymbol{\Psi}\,\boldsymbol{\Lambda}' + \boldsymbol{\Theta}. \tag{55}$$

In the exploratory factor analysis model there are m^2 indeterminacies in $V(\boldsymbol{y^*})$ so that m^2 restrictions need to be imposed on the elements of $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ to make the model identified. These restrictions are imposed to make it computationally simple to reach a solution, and this solution is then rotated by VARIMAX and PROMAX methods (see, e.g., Lawley & Maxwell, 1971) to give a more easily interpretable solution.

COMPUTATIONAL IMPLEMENTATION

In terms of computational implementation of the general model, a distinction is made between two different cases: the case where not all y variables are continuous and the case where all y variables are continuous.

NOT ALL y VARIABLES CONTINUOUS

In the case where not all y variables are continuous, it is convenient to divide the parametric structure of the model into three parts. This general structure encompasses a combination of categorical and continuous y variables.

Part 1 contains a mean, or threshold, or intercept structure. It represents the part of the integral limits in the probability expressions such as (52) and (53) that are not related to $\mathbf{x}$. The Part 1 structure is

$$\boldsymbol{\Delta}^* \left[\mathbf{K}_\tau\,\boldsymbol{\tau} - \mathbf{K}_\nu[\boldsymbol{\nu} + \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\alpha}]\right]. \tag{56}$$

Here, $\mathbf{K}_\nu$ is a selection matrix that selects rows from the vector it premultiplies. The premultiplied vector is recognized as $\boldsymbol{\pi}_0$ in (48). For categorical y variables with more than one threshold (more than two categories), $\mathbf{K}_\nu$ repeats $\boldsymbol{\pi}_0$. $\boldsymbol{\tau}$ contains the threshold

parameters for all categorical y variables. $\mathbf{K}_\tau$ is a matrix that selects elements from $\boldsymbol{\tau}$ for categorical y variables and does not select an element for a continuous y variable. $\boldsymbol{\Delta}^*$ contains an expanded $\boldsymbol{\Delta}$ matrix so that with a categorical y variable having more than one threshold, the same $\boldsymbol{\Delta}$ scaling takes place for each component of the vector it premultiplies, corresponding to each threshold.

Part 2 contains a slope structure. It represents the part of the integral limits in probability expressions such as (52) and (53) that are related to $\mathbf{x}$. Part 2 is only used if at least one of the y variables is categorical. The Part 2 structure is

$$vec[\boldsymbol{\Delta} \, \boldsymbol{\Lambda} \, (\mathbf{I} - \mathbf{B})^{-1} \, \boldsymbol{\Gamma}], \tag{57}$$

where the $vec[\mathbf{A}]$ operator arranges elements of the matrix $\mathbf{A}$ in a vector, taking the elements of $\mathbf{A}$ row wise and only including the lower-triangular elements including the diagonal if $\mathbf{A}$ is symmetric. The $\mathbf{K}$ parameter matrix is not involved, but is handled as described below for continuous y variables.

Part 3 contains a covariance, correlation, or residual correlation structure. With categorical y variables, the residual correlation matrix is given by $\boldsymbol{\Sigma}_s^*$ in (51). The Part 3 structure is

$$\mathbf{K}_\sigma \, vec[\boldsymbol{\Delta} \, [\boldsymbol{\Lambda} \, (\mathbf{I} - \mathbf{B})^{-1} \, \boldsymbol{\Psi} \, (\mathbf{I} - \mathbf{B})'^{-1} \boldsymbol{\Lambda}' + \boldsymbol{\Theta}] \, \boldsymbol{\Delta}], \tag{58}$$

where $\mathbf{K}_\sigma$ selects elements from the vector it premultiplies so that diagonal elements are not included if the corresponding y variable is categorical.

The full set of parameter arrays is as follows.

- $\boldsymbol{\tau}$ is a threshold vector of length equal to the number of thresholds in the model, where a categorical y variable with C categories contributes $C - 1$ thresholds.

- $\boldsymbol{\nu}$ is a $p \times 1$ vector of measurement intercepts.

- $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of measurement slopes or loadings.

- $\boldsymbol{\Theta}$ is a $p \times p$ covariance matrix for the residuals in the measurement relations.

- $\boldsymbol{\alpha}$ is an $m \times 1$ vector of latent variable means and intercepts.

- $\mathbf{B}$ is an $m \times m$ matrix of slopes for the regressions among the m latent variables. It has zero diagonal elements.

- $\boldsymbol{\Gamma}$ is an $m \times q$ matrix of slopes for the regressions of the m latent variables on the q x variables.

- $\boldsymbol{\Psi}$ is an $m \times m$ covariance matrix for the latent variables and the residuals in the latent variable relations.

- $\boldsymbol{\Delta}$ is a diagonal $p \times p$ matrix of scaling factors. Only the p diagonal elements are parameters.

ALL y VARIABLES CONTINUOUS

In the case where each variable of $\boldsymbol{y^*}$ corresponds to a continuous y variable, the computational implementation differs somewhat from what is expressed above. Here, Part 1 captures a mean structure that does not involve $\boldsymbol{\tau}$. Part 2 is deleted because probit slopes are not considered. Part 3 captures a covariance or correlation structure. In addition, the computations do not include the $\mathbf{K}$ or $\boldsymbol{\Gamma}$ matrices in (36) and (37) but use fewer parameter arrays based on the measurement and structural equations,

$$\mathbf{v}_i = \boldsymbol{\nu}_v + \boldsymbol{\Lambda}_v \, \boldsymbol{\eta}_{vi} + \boldsymbol{\epsilon}_{vi} \tag{59}$$

and

$$\boldsymbol{\eta}_{vi} = \boldsymbol{\alpha}_v + \mathbf{B}_v \, \boldsymbol{\eta}_{vi} + \boldsymbol{\zeta}_{vi}. \tag{60}$$

Here, $\mathbf{v}_i = (\mathbf{y}_i', \, \mathbf{x}_i')'$ and $\boldsymbol{\eta}_{vi} = (\boldsymbol{\eta}_i', \, \boldsymbol{\eta}_{yi}', \, \boldsymbol{\eta}_{xi}')'$, where $\boldsymbol{\eta}_y$ contains a new latent variable η_y for each y variable that is regressed on another variable for which y is not an indicator, and $\boldsymbol{\eta}_x$ contains a new latent variable η_x for each x variable. The two new types of latent variables are defined to be identical to their observed counterparts. With this variable definition, the parameter arrays used in (59) and (60) can be expressed as

$$\boldsymbol{\nu}_v = \begin{pmatrix} \boldsymbol{\nu}_{nd} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix},$$

$$\boldsymbol{\Lambda}_v = \begin{pmatrix} \boldsymbol{\Lambda}_{nd} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix},$$

$$\boldsymbol{\Theta}_v = \begin{pmatrix} \boldsymbol{\Theta}_{nd} & & symm. \\ \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$$\boldsymbol{\alpha}_v = \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\nu}_d \\ \boldsymbol{\mu}_x \end{pmatrix},$$

$$\mathbf{B}_v = \begin{pmatrix} \mathbf{B} & \mathbf{B}_{\eta d} & \boldsymbol{\Gamma} \\ \boldsymbol{\Lambda}_d & \mathbf{B}_d & \mathbf{K}_d \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$$\boldsymbol{\Psi}_v = \begin{pmatrix} \boldsymbol{\Psi} & & symm. \\ \boldsymbol{\Psi}_{d\eta} & \boldsymbol{\Theta}_d & \\ \boldsymbol{\Psi}_{x\eta} & \boldsymbol{\Psi}_{xd} & \boldsymbol{\Sigma}_{xx} \end{pmatrix},$$

APPENDIX 2

where $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ are the mean vector and covariance matrix of the x variables. In typical applications, $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ are unrestricted because the model concerns the conditional moments (38) and (39), given $\mathbf{x}$. In the optimization, the $\boldsymbol{\mu}_x$ and $\boldsymbol{\Sigma}_{xx}$ parameter arrays are held fixed at the corresponding sample statistics and the degree of freedom calculation is adjusted accordingly. The $\boldsymbol{\Delta}$ matrix can be used. In the formulas above the order of the y variables has been rearranged to simplify the expressions. In the parameter specification printout of parameter arrays, however, Mplus maintains the original order of the y variables.

APPENDIX 3
STANDARDIZED PARAMETERS

To facilitate the parameter interpretation, parameters and their estimates may be standardized with respect to the variances of $\boldsymbol{\eta}$ as well as the variances of $\mathbf{y}$ and $\mathbf{x}$.

The modeling considers

$$\boldsymbol{y^*}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_i + \mathbf{K}\,\mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{61}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\,\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\,\mathbf{x}_i + \boldsymbol{\zeta}_i, \tag{62}$$

and

$$E(\boldsymbol{y^*}_{si}|\mathbf{x}_i) = \boldsymbol{\Delta}[\boldsymbol{\nu} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\,\mathbf{x}_i + \mathbf{K}\,\mathbf{x}_i], \tag{63}$$

$$V(\boldsymbol{y^*}_{si}|\mathbf{x}_i) = \boldsymbol{\Delta}[\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}]\boldsymbol{\Delta}. \tag{64}$$

Consider first the standardization to unit $\boldsymbol{\eta}$ variances. Let the diagonal matrix $\mathbf{D}_\eta$ represent $\boldsymbol{\eta}$ standard deviations,

$$\mathbf{D}_\eta = diag[(\mathbf{I} - \mathbf{B})^{-1}\,\boldsymbol{\Gamma}\,\boldsymbol{\Sigma}_{xx}\,\boldsymbol{\Gamma}'\,(\mathbf{I} - \mathbf{B})'^{-1} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}]^{1/2}, \tag{65}$$

so that $\boldsymbol{\eta}_i^* = \mathbf{D}_\eta^{-1}\,\boldsymbol{\eta}_i$ has unit variances. Noting that

$$\mathbf{D}_\eta^{-1}(\mathbf{I} - \mathbf{B})^{-1} = [(\mathbf{I} - \mathbf{B})\,\mathbf{D}_\eta]^{-1} = [\mathbf{D}_\eta\,(\mathbf{I} - \mathbf{D}_\eta^{-1}\,\mathbf{B}\,\mathbf{D}_\eta)]^{-1} = (\mathbf{I} - \mathbf{D}_\eta^{-1}\,\mathbf{B}\,\mathbf{D}_\eta)^{-1}\,\mathbf{D}_\eta^{-1}, \tag{66}$$

the standardized parameter arrays are defined as

$$\boldsymbol{\nu}^* = \boldsymbol{\nu}, \tag{67}$$
$$\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\,\mathbf{D}_\eta, \tag{68}$$
$$\mathbf{K}^* = \mathbf{K}, \tag{69}$$
$$\boldsymbol{\Theta}^* = \boldsymbol{\Theta}, \tag{70}$$
$$\boldsymbol{\alpha}^* = \mathbf{D}_\eta^{-1}\,\boldsymbol{\alpha}, \tag{71}$$
$$\mathbf{B}^* = \mathbf{D}_\eta^{-1}\,\mathbf{B}\,\mathbf{D}_\eta, \tag{72}$$
$$\mathbf{B}_{\eta d}^* = \mathbf{D}_\eta^{-1}\,\mathbf{B}_{\eta d}, \tag{73}$$
$$\mathbf{B}_d^* = \mathbf{B}_d, \tag{74}$$
$$\boldsymbol{\Gamma}^* = \mathbf{D}_\eta^{-1}\,\boldsymbol{\Gamma}, \tag{75}$$
$$\boldsymbol{\Psi}^* = \mathbf{D}_\eta^{-1}\,\boldsymbol{\Psi}\,\mathbf{D}_\eta^{-1}, \tag{76}$$
$$\boldsymbol{\Psi}_{d\eta}^* = \boldsymbol{\Psi}_{d\eta}\,\mathbf{D}_\eta^{-1}, \tag{77}$$
$$\boldsymbol{\Psi}_{x\eta}^* = \boldsymbol{\Psi}_{x\eta}\,\mathbf{D}_\eta^{-1}, \tag{78}$$
$$\boldsymbol{\Psi}_{xd}^* = \boldsymbol{\Psi}_{xd}. \tag{79}$$

APPENDIX 3

The parameter arrays may also be standardized with respect to the variances of $\mathbf{y}^*$ and $\mathbf{x}$. Let the diagonal matrices $\mathbf{D}_y$ and $\mathbf{D}_x$ represent $\boldsymbol{y^*}$ and $\mathbf{x}$ standard deviations, respectively, where for $\mathbf{y}^*$ corresponding to continuous y variables,

$$\mathbf{D}_y = diag[\boldsymbol{\Delta}[\boldsymbol{\Lambda}(\mathbf{I}{-}\mathbf{B})^{-1}\,\boldsymbol{\Gamma}\,\boldsymbol{\Sigma}_{xx}\,\boldsymbol{\Gamma}'\,(\mathbf{I}{-}\mathbf{B})'^{-1}\boldsymbol{\Lambda}'{+}\boldsymbol{\Lambda}(\mathbf{I}{-}\mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I}{-}\mathbf{B})'^{-1}\boldsymbol{\Lambda}'{+}\boldsymbol{\Theta}]\boldsymbol{\Delta}]^{1/2}, \quad (80)$$

while for categorical y variables the variance expression in (45) is used. $\mathbf{D}_x$ is formed from the square root of the diagonal elements of $\boldsymbol{\Sigma}_{xx}$. Standardizing also with respect to the variances of $\mathbf{y}$ and $\mathbf{x}$,

$$\boldsymbol{\nu}^{**} = \mathbf{D}_y^{-1}\,\boldsymbol{\nu}, \tag{81}$$
$$\boldsymbol{\Lambda}^{**} = \mathbf{D}_y^{-1}\,\boldsymbol{\Lambda}^*, \tag{82}$$
$$\mathbf{K}^{**} = \mathbf{D}_y^{-1}\,\mathbf{K}\,\mathbf{D}_x, \tag{83}$$
$$\boldsymbol{\Theta}^{**} = \mathbf{D}_y^{-1}\,\boldsymbol{\Theta}\,\mathbf{D}_y^{-1}, \tag{84}$$
$$\boldsymbol{\alpha}^{**} = \boldsymbol{\alpha}^*, \tag{85}$$
$$\mathbf{B}^{**} = \mathbf{B}^*, \tag{86}$$
$$\mathbf{B}_{\eta d}^{**} = \mathbf{D}_\eta^{-1}\,\mathbf{B}_{\eta d}\,\mathbf{D}_y, \tag{87}$$
$$\mathbf{B}_d^{**} = \mathbf{D}_y^{-1}\,\mathbf{B}_d^*\,\mathbf{D}_y, \tag{88}$$
$$\boldsymbol{\Gamma}^{**} = \boldsymbol{\Gamma}^*\,\mathbf{D}_x, \tag{89}$$
$$\boldsymbol{\Psi}^{**} = \boldsymbol{\Psi}^*, \tag{90}$$
$$\boldsymbol{\Psi}_{d\eta}^{**} = \mathbf{D}_y^{-1}\,\boldsymbol{\Psi}_{d\eta}\,\mathbf{D}_\eta^{-1}, \tag{91}$$
$$\boldsymbol{\Psi}_{x\eta}^{**} = \mathbf{D}_x^{-1}\,\boldsymbol{\Psi}_{x\eta}\,\mathbf{D}_\eta^{-1}, \tag{92}$$
$$\boldsymbol{\Psi}_{xd}^{**} = \mathbf{D}_x^{-1}\,\boldsymbol{\Psi}_{xd}\,\mathbf{D}_y^{-1}. \tag{93}$$

R^2 measures of the amount of variation explained in each outcome variable can be obtained from these standardized expressions. For $\boldsymbol{y^*}$ these are obtained as 1 minus the corresponding diagonal element of $\boldsymbol{\Theta}^{**}$, while for $\boldsymbol{\eta}$ they are obtained as 1 minus the corresponding diagonal element of $\boldsymbol{\Psi}^*$.

16

APPENDIX 4
ESTIMATORS

Estimation of model parameters is carried out using the maximum-likelihood method under normality assumptions. Weighted least squares methods are used with categorical outcomes. With continuous, non-normal data, limited-information methods are used which are based on the normal theory maximum-likelihood fitting function but do not assume normality in the calculation of standard errors or the χ^2 square test of model fit.

MAXIMUM-LIKELIHOOD (ML) ESTIMATION

A simultaneous, multiple-population analysis is performed under the customary assumption of i.i.d sampling from each of $G(g = 1, 2, \ldots, G)$ populations. With continuous y variables for which multivariate normality is assumed to hold for $\mathbf{y}_g$ conditional on $\mathbf{x}_g$, maximum-likelihood estimation is carried out by considering $\mathbf{v}_{gi} = (\mathbf{y}'_{gi}, \mathbf{x}'_{gi})'$ and

$$\boldsymbol{\mu}'_g = E(\mathbf{v}_{gi}), \tag{94}$$
$$\boldsymbol{\Sigma}'_g = V(\mathbf{v}_{gi}). \tag{95}$$

Here, the marginal means, variances, and covariances among the x variables need not be estimated given that no structure is imposed on the x variables. They are held fixed at the corresponding sample statistics which are the ML estimates.

ML estimation is obtained by minimizing the fitting function F_{ML} with respect to the vector of parameters $\boldsymbol{\pi}$,

$$F_{ML}(\boldsymbol{\pi}) = 1/2 \sum_{g=1}^{G} \{n_g[\ln|\boldsymbol{\Sigma}_g| + \; trace \; (\boldsymbol{\Sigma}_g^{-1}\mathbf{T}_g) - \ln|\mathbf{S}_g| - (p+q)]\}/n, \tag{96}$$

where n is the total sample size, n_g is the sample size in group g and

$$\mathbf{T}_g = \mathbf{S}_g + (\bar{\mathbf{v}}_g - \boldsymbol{\mu}_g)(\bar{\mathbf{v}}_g - \boldsymbol{\mu}_g)' \tag{97}$$

(see, e.g. Jöreskog & Sörbom, 1979; Sörbom, 1982; Browne & Arminger, 1995).

Under normality assumptions, standard errors of the parameter estimates $\hat{\boldsymbol{\pi}}$ are computed as in Jöreskog (1973, Appendix A2). Chi-square test of fit with ML is discussed in Appendix 5.

WEIGHTED LEAST-SQUARES ESTIMATION

Weighted least-squares estimation under the assumption of multivariate normality gives an estimator commonly referred to as GLS (generalized least squares). For GLS, the fitting

APPENDIX 4

function of (96) is modified as

$$F_{GLS}(\boldsymbol{\pi}) = \sum_{g=1}^{G}((n_g - 1) \; trace \; [(\boldsymbol{\Sigma} - \mathbf{S})\mathbf{S}^{-1}]^2 + n_g \; trace \; [\mathbf{S}^{-1}(\bar{\boldsymbol{v}}_g - \boldsymbol{\mu}_g)(\bar{\boldsymbol{v}}_g - \boldsymbol{\mu}_g)'])/n. \quad (98)$$

Weighted least squares is also used when the $\mathbf{y}$ vector is not multivariate normal conditional on $\mathbf{x}$. In this case the fitting function is

$$F_{WLS}(\boldsymbol{\pi}) = 1/2 \sum_{g=1}^{G}(\mathbf{s}_g - \boldsymbol{\sigma}_g)' \; \mathbf{W}_g^{-1} \; (\mathbf{s}_g - \boldsymbol{\sigma}_g). \quad (99)$$

When all y variables are continuous but not normal this gives ADF estimation (Browne, 1982, 1984). Muthén (1989a) presents $\mathbf{W}$ when the model includes a mean structure. When a correlation matrix is analyzed, computations of the ADF weight matrix are done according to Mooijaart (1985). When at least one y variable is categorical (99) gives the estimator of Muthén (1983, 1984). Asymptotic theory for the Muthén estimator is further discussed in Muthén and Satorra (1995a). For multiple-group analysis with categorical outcomes, see also Muthén and Christoffersson (1981).

There are three steps to the model estimation using weighted least squares by (99). First, the elements of $\boldsymbol{\tau}^*$, $E(\boldsymbol{y}^*_s|\mathbf{x})$ and $V(\boldsymbol{y}^*_s|\mathbf{x})$ are estimated as $\mathbf{s}$. When all variables in $\mathbf{y}$ are continuous, $\mathbf{s}$ simply contains the sample covariance matrix or the sample mean vector and the sample covariance matrix. When all variables in $\mathbf{y}$ are categorical, $\mathbf{s}$ is computed by a set of p probit regressions of each y on all x variables, followed by a set of $p(p-1)/2$ bivariate probit regressions of each pair of y variables on all x variables. These regressions use the probability expressions of (52) and (53). Here, $\mathbf{s}$ represent probit thresholds, slopes and residual correlations. Second, a weight matrix $\mathbf{W}$ is formed as an estimate of the asymptotic covariance matrix of $\mathbf{s}$. Third, the model parameters are estimated by minimizing the weighted least-squares fitting function.

Assuming for simplicity a single group and defining

$$\boldsymbol{\Delta} = \partial\boldsymbol{\sigma}(\boldsymbol{\pi})/\partial\boldsymbol{\pi}, \quad (100)$$

an estimate of the asymptotic variance matrix for $\hat{\boldsymbol{\pi}}$ is obtained when evaluating

$$aV(\hat{\boldsymbol{\pi}}) = n^{-1}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1} \quad (101)$$

at $\hat{\boldsymbol{\pi}}$. This variance estimator is sometimes referred to as the naive or model-based variance estimator.

A large-sample chi-square test of model fit is obtained as $2 \, n \times F_{WLS}(\hat{\boldsymbol{\pi}})$.

ROBUST ESTIMATION

Assume for simplicity a single group. The robust asymptotic covariance matrix for the estimated parameter vector $\hat{\boldsymbol{\pi}}$ is obtained using

$$\mathrm{aV}(\hat{\boldsymbol{\pi}}) = n^{-1}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Gamma}\mathbf{W}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1}, \tag{102}$$

where $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix for the vector of sample statistics $\mathbf{s}$. In the case where all y variables are continuous,

$$\mathbf{s} = vec\begin{pmatrix} \bar{\mathbf{v}} \\ \mathbf{S} \end{pmatrix}.$$

In this case, $\boldsymbol{\Gamma}$ is estimated by the ADF estimator for mean and covariance structures given in Muthén (1989a) using third- and fourth-order moments and $\mathbf{W}$ is estimated by the normal theory weight matrix for $\bar{v}$ and $\mathbf{S}$

$$\mathbf{W} = \begin{pmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & A \end{pmatrix}, \tag{103}$$

where

$$\mathbf{A} = 2\mathbf{C}'\left(\mathbf{S} \otimes \mathbf{S}\right)\mathbf{C}, \tag{104}$$

where $\mathbf{C}$ is the transition matrix for a symmetric matrix (see, e.g., Browne & Arminger, 1995).

For the case where at least one variable in $\mathbf{y}$ is categorical, $\boldsymbol{\Gamma}$ is estimated using the limited-information likelihood approach of Muthén (1984) and $\mathbf{W}$ is a diagonal matrix using the estimated variances of the $\mathbf{s}$ elements (see Muthén, Du Toit & Spisic, 1997).

Furthermore (cf. Satorra & Bentler, 1988, 1994; Satorra, 1992), a robust goodness-of-fit test is obtained as the mean-adjusted chi square defined as

$$G_M = 2\,n\,F(\hat{\boldsymbol{\pi}})/c, \tag{105}$$

where c is a scaling correction factor,

$$c = tr[\mathbf{U}\boldsymbol{\Gamma}]/d, \tag{106}$$

with

$$\mathbf{U} = (\mathbf{W}^{-1} - \mathbf{W}^{-1}\boldsymbol{\Delta}(\boldsymbol{\Delta}'\mathbf{W}^{-1}\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}'\mathbf{W}^{-1}) \tag{107}$$

and where d is the degrees of freedom of the model. A mean- and variance-adjusted goodness-of-fit statistic is defined as

$$G_{MV} = 2\,n\,F(\hat{\boldsymbol{\pi}})/c_2, \tag{108}$$

where

$$c_2 = [tr(\mathbf{U\Gamma})/d'], \tag{109}$$

where d' is computed as the integer closest to d^*,

$$d^* = (tr(\mathbf{U\Gamma}))^2/tr((\mathbf{U\Gamma})^2). \tag{110}$$

Chi-square difference testing using robust goodness-of-fit tests calls for special procedures and is discussed in Appendix 5.

NUMERICAL TECHNIQUES

To find model estimates, iterative techniques from numerical analysis are used to optimize the fitting function corresponding to a particular estimator. In most cases, a quasi-Newton technique is used. Following a few initial steps using a gradient method, the method uses an approximation to the second-order derivative matrix built up during iterations. The technical details of this optimization that can be modified by the user are the total number of iterations and the convergence criterion. The default number of iterations is 1000 and the default convergence criterion is that the absolute value of each first-order derivative has to be less than 0.00005.

There are two exceptions to the use of quasi-Newton optimization. These are for missing data and for mixture modeling, and are discussed below.

APPENDIX 5
MODEL TESTS OF FIT AND MODEL MODIFICATION

This Appendix describes model testing using χ^2, information criteria, root mean square error of approximation measures, and residual measures. Indices useful for modifying a model are also described.

With continuous y variables, the ML estimator assumes multivariate normality for $\mathbf{v}'_i = (\mathbf{y}'_i, \mathbf{x}'_i)$ and n i.i.d. observations in each of the several groups (group index omitted here),

$$logL = -c - n/2 \; log \; |\mathbf{\Sigma}| - 1/2 \; \mathbf{A}, \tag{111}$$

where $c = np/2 \; log \; (2\pi)$ and

$$\mathbf{A} = \sum_{i=1}^{n} (\mathbf{v}_i - \boldsymbol{\mu})' \; \mathbf{\Sigma}^{-1} \; (\mathbf{v}_i - \boldsymbol{\mu}) \tag{112}$$

$$= trace \; [\mathbf{\Sigma}^{-1} \sum_{i=1}^{n} (\mathbf{v}_i - \boldsymbol{\mu})(\mathbf{v}_i - \boldsymbol{\mu})'] \tag{113}$$

$$= n \; trace \; [\mathbf{\Sigma}^{-1} \; (\mathbf{S} + (\bar{\mathbf{v}} - \boldsymbol{\mu})(\bar{\mathbf{v}} - \boldsymbol{\mu})')]. \tag{114}$$

The standard H_1 model considers an unrestricted mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$. Under this model the maximum-likelihood value is

$$logL_{H_1} = -c - n/2 \; log|\mathbf{S}| - n/2 \; (p+q). \tag{115}$$

This leads to the ML fitting function given earlier for G groups (populations)

$$F_{ML}(\boldsymbol{\pi}) = 1/2 \sum_{g=1}^{G} \{n_g[\ln |\mathbf{\Sigma}_g| + \; trace \; (\mathbf{\Sigma}_g^{-1}\mathbf{T}_g) - \ln |\mathbf{S}_g| - (p+q)]\}/n, \tag{116}$$

where n is the total sample size, n_g is the sample size in group g and

$$\mathbf{T}_g = \mathbf{S}_g + (\bar{\boldsymbol{v}}_g - \boldsymbol{\mu}_g)(\bar{\boldsymbol{v}}_g - \boldsymbol{\mu}_g)'. \tag{117}$$

Note that

$$F_{ML}(\boldsymbol{\pi}) = -logL/n + logL_{H_1}/n. \tag{118}$$

Letting $\hat{\boldsymbol{\pi}}$ denote the ML estimate under H_0, the value of the likelihood-ratio χ^2-test of model fit for H_0 against H_1 is therefore obtained as $2 \; n \; F_{ML}(\hat{\boldsymbol{\pi}})$. In this expression, the factor n is often replaced by the factor $n - 1$, which gives slightly different χ^2 values for small samples. The $n - 1$ factor can be motivated by considering the Wishart distribution for $\mathbf{S}$ instead of the normal distribution for $\mathbf{v}$ (see, e.g., Browne & Arminger, 1995). The

APPENDIX 5

Wishart distribution holds when $\mathbf{S}$ has been created in the customary way by dividing by $n-1$. In (114), (115), and (116), however, the $\mathbf{S}$ matrix is obtained by division by n, not $n-1$. This definition of $\mathbf{S}$ gives the unrestricted ML estimate used in (115) and calls for using the factor n in $2\,n\,F_{ML}(\hat{\boldsymbol{\pi}})$. When a sample covariance matrix $\mathbf{S}$ is given as input to Mplus, it is assumed that it is given in the customary form dividing by $n-1$. In the ML estimation Mplus then transforms this matrix to being divided by n.

Robust goodness-of-fit tests discussed in Appendix 4 for continuous outcomes call for special chi-square difference testing procedures of nested models because the difference between two robust chi-square variates does not have a chi-square distribution (Satorra, 2000). The correct chi-square difference test procedure was outlined in Satorra (2000) and a simplified version for the mean-adjusted chi square was presented in Satorra and Bentler (1999). The simplified version is easily computed using the scaling correction factor c in (106), which is printed by Mplus. Denoting the regular chi-square test value for the more and less restrictive model T_0 and T_1, respectively, and the mean-adjusted counterparts as T_{m0} and T_{m1}, the mean-adjusted robust chi-square difference test T_{md} is

$$T_{md} = (T_0 - T_1)/c_d, \tag{119}$$
$$= (T_{m0}\,c_0 - T_{m1}\,c_1)/c_d, \tag{120}$$

where c_0 and c_1 are the scaling correction factors for the more and less restrictive model, respectively, and

$$c_d = (d_0\,c_0 - d_1\,c_1)/(d_0 - d_1), \tag{121}$$

where d_0 and d_1 are the degrees of freedom for the more and less restrictive model, respectively.

With maximum-likelihood estimation, Mplus computes information criteria which are useful for comparing non-nested models. The Akaike information criterion (AIC) is defined as

$$AIC = -2\,logL + 2\,r, \tag{122}$$

where r is the number of free model parameters (Akaike, 1987). The Bayesian information criterion (Schwartz, 1978) is defined as

$$BIC = -2\,logL + r\,ln\,n. \tag{123}$$

Sclove (1987) suggested a sample-size adjusted BIC, replacing n by

$$n^* = (n+2)/24. \tag{124}$$

Mplus also provides the root mean square error of approximation (RMSEA) model fit measure (see Browne & Cudeck, 1993; Steiger & Lind, 1980). With continuous outcomes, RMSEA is defined as

$$RMSEA = \sqrt{max[(2\,F_{ML}(\hat{\boldsymbol{\pi}})/d - 1/n), 0]}\,\sqrt{G}, \tag{125}$$

22

where d is the number of degrees of freedom of the model and G is the number of groups. This formula uses the Steiger (1998) modification for multiple groups. Browne and Cudeck (1993) recommend a value of less than 0.05 for reasonably well-fitting models. With continuous outcomes, Mplus prints an RMSEA confidence interval. For single group analysis, the probability that the value is less than or equal to 0.05 is also printed. With categorical outcomes, Mplus replaces d in (125) by $tr[\mathbf{U\Gamma}]$ in line with (106).

Other fit indices include the conventional indices TLI and CFI,

$$TLI = (\chi_B^2/d_B - \chi_{H_0}^2/d_{H_0})/(\chi_B^2/d_B - 1), \qquad (126)$$

$$CFI = 1 - max(\chi_{H_0}^2 - d_{H_0}, 0)/max(\chi_{H_0}^2 - d_{H_0}, \chi_B^2 - d_B, 0), \qquad (127)$$

where d_B and d_{H_0} denote the degrees of freedom of the baseline and H_0 models, respectively. TLI and CFI are available for both continuous and categorical outcomes. The baseline model has uncorrelated outcomes with unrestricted variances and unrestricted means and/or thresholds. With two-level models, the baseline model sets both the between and within covariances to zero. With categorical outcomes, the baseline model does not set to zero the covariances among the covariates of $\mathbf{x}$ because the x variables are not part of the model.

Mplus also computes the residual-based fit indices RMSR, SRMR and WRMR. The SRMR (standardized root mean square residual) fit index is defined as

$$SRMR = \sqrt{\sum_j \sum_k {}_j r_{jk}^2/e} \,. \qquad (128)$$

Here, $e = p(p+1)/2$, where p is the number of outcomes and r_{jk} is a residual in a correlation metric,

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} - \frac{\hat{\sigma}_{jk}}{\sqrt{\hat{\sigma}_{jj}}\sqrt{\hat{\sigma}_{kk}}}, \qquad (129)$$

where s_{jk} and $\hat{\sigma}_{jk}$ are the sample and model-estimated covariance between the continuous outcomes y_j and y_k, respectively. When $\hat{\sigma}_{jj} = s_{jj}$ and $\hat{\sigma}_{kk} = s_{kk}$ this coincides with the definition given in Hu and Bentler (1999). With categorical outcomes, SRMR is only available when all outcomes are categorical with no threshold structure or covariates, in which case the denominators of (129) are all unity. For RMSR r_{jk} is a residual correlation and the summation involves $k < j$.

The WRMR (weighted root mean square residual) fit index is defined as

$$WRMR = \sqrt{\sum_r^e \frac{(s_r - \hat{\sigma}_r)^2}{v_r}/e} \,, \qquad (130)$$

where s_r is an element of the sample statistics vector, $\hat{\sigma}_r$ is the estimated model counterpart, v_r is an estimate of the asymptotic variance of s_r, and e is the number of sample statistics.

APPENDIX 5

WRMR is suitable for models where sample statistics have widely varying variances, and when sample statistics are on different scales such as in models with mean and/or threshold structures. WRMR is also suitable with non-normal continuous outcomes. With continuous outcomes, WRMR is available with MLM and MLMV for which v_r is taken from the diagonal of $\mathbf{\Gamma}$ (see Appendix 4.) With categorical outcomes, WRMR is available using the diagonally-weighted least-squares estimators WLSM and WLSMV, so that WRMR is $\sqrt{2\,n\,F_{min}/e}$, where F_{min} is the minimum of the fitting function $F_{WLS}(\boldsymbol{\pi})$ in (99). Small values of WRMR indicate good fit. Generalizations to multiple groups use group-size weighted counterparts for SRMR and WRMR.

Hu and Bentler (1999) suggests the following fit index cut off value guide for good models with continuous outcomes: TLI > .95, CFI > .95, RMSEA < .06, SRMR < .08. Simulation studies in Yu and Muthén (2001) suggest that these cut off values are reasonable also with categorical outcomes, except that the SRMR cut off of .08 does not work well with small sample sizes (≤ 250). Yu and Muthén (2001) suggests WRMR < .90 for good models with continuous as well as with categorical outcomes.

When a model does not fit well, a modification can be guided by modification indices. For the case where all y variables are continuous and multivariate normal, Sörbom (1989) proposed an index called MI, which is based on first-order derivatives and expected second-order derivatives in line with Lagrangian multiplier tests. It is a meausure of how poorly a particular parameter constraint is chosen. For a parameter that is not freely estimated but either fixed or constrained to be equal to another parameter, MI gives the expected drop in the likelihood ratio chi-square statistic when this parameter is freed. An expected parameter change (EPC) statistic is also useful in evaluating possible model modifications (see Saris, Satorra & Sörbom, 1987). EPC is based on the ratio of the MI and the first-order derivative and represents the change in the parameter estimate when the parameter is freed. Parameters are clearly in need of being freed only when the MI values are large and the EPC values are large.

Modification indices and expected parameter change are not available in cases where $\mathbf{y}$ is not multivariate normal. In such situations, however, first-order derivatives of parameters are given and can be used as a guide to which parameters need to be freely estimated. First-order derivatives are, however, not directly related to the chi-square test of model fit and are dependent on the scale of the variables in the model.

APPENDIX 6
MISSING DATA

Missing data is allowed for in cases where all y variables are continuous and normally distributed (for mixture modeling, see Appendix 8). Maximum-likelihood estimation draws on theory in Little and Rubin (1987) assuming ignorable missingness with missing at random (MAR). MAR means that the probabilities of values being missing can be predicted by variables that are not missing, for instance x variables and variables observed at the first time point of a longitudinal study. Mplus performs maximum-likelihood estimation under MAR.

The vector of observations for individual i in group g is $\mathbf{v}_{gi} = (\mathbf{y}'_{gi}, \mathbf{x}'_{gi})'$ and

$$\boldsymbol{\mu}'_g = E(\mathbf{v}_{gi}), \tag{131}$$
$$\boldsymbol{\Sigma}'_g = V(\mathbf{v}_{gi}). \tag{132}$$

With missing data, it is convenient to first consider a model with no restrictions on $\boldsymbol{\mu}_g$ or $\boldsymbol{\Sigma}_g$. This model will be referred to as the unrestricted model, or the H_1, model. Maximum-likelihood estimates of the mean, variance, and covariance parameters of the H_1 model are obtained using the EM algorithm as discussed in Little and Rubin (1987).

The estimation of the latent variable model will be referred to as the H_0 model. In estimating the H_0 model, the marginal means, variances, and covariances among the x variables need not be estimated given that no structure is imposed on these variables. With missing data on x variables they are held fixed at the corresponding ML estimates obtained from estimating that part of the H_1 model. In the missing data case, a modification of the ML fitting function described in Appendix 4 is used to estimate the H_0 model,

$$F_{ML}(\boldsymbol{\pi}) = \sum_{g=1}^{G} \sum_{i=1}^{n_g} [\ln |\boldsymbol{\Sigma}_{gi}| + (\mathbf{v}_{gi} - \boldsymbol{\mu}_{gi})' \, \boldsymbol{\Sigma}_{gi}^{-1} \, (\mathbf{v}_{gi} - \boldsymbol{\mu}_{gi}) - c_g]. \tag{133}$$

Here, the population mean vectors of $\boldsymbol{\mu}_g$ and population covariance matrices of $\boldsymbol{\Sigma}_g$ have been given subscripts i to denote that they vary across the observations by the fact that the size of these arrays depend on how many variables are not missing for a given observation. For each group, observations are sorted into missing data patterns. Matrix inversion for different patterns is facilitated by the sweep operator (Little & Rubin, 1987). An offset term c_g is included in the fitting function, where c_g is the optimum value of the fitting function for each group under H_1. In this way, (133) evaluated at $\hat{\boldsymbol{\pi}}$ gives a chi-square measure of model fit for H_0 compared to H_1 as $2\, n\, F_{ML}(\hat{\boldsymbol{\pi}})$.

When the H_0 model is estimated, the estimation of the H_1 model is left as an option for the user. If the H_1 model is not estimated a χ^2 test of model fit of H_0 against the H_1 model is not obtained. Likelihood-ratio χ^2 testing of a series of nested models can, however, still be

obtained from the H_0 log likelihood values of these nested models. With large amounts of missing data, estimation of the H_1 model can be time-consuming and lead to computational difficulties and slow convergence. To help assess this difficulty, Mplus computes a covariance coverage matrix that describes the extent of missing data. In the covariance coverage matrix diagonal and off-diagonal elements give the proportion of available observations for each variable and pairs of variables, respectively. A default minimum coverage value of 0.10 is used to protect against computational difficulties. The estimation of the H_0 model is in general not as strongly influenced by low coverage as the H_1 model.

Missing by design gives zero coverage for certain elements in the covariance coverage matrix. Such zero coverage elements do not give computational difficulties because the corresponding elements of the covariance matrix are not used in the computations. The number of such zero coverage elements are also used to adjust the degrees of freedom of the model test of fit.

NUMERICAL TECHNIQUES

With missing data, an unrestricted model for the mean vector and covariance matrix is considered when estimating the H_1 model. This model is estimated using the EM algorithm described in Little and Rubin (1987). The default maximum number of iterations is 500. After ten iterations, convergence is checked based on the change from one iteration to the next. For each parameter value the change must be less than 0.0001 and when this is fullfilled a further requirement is that the $|2 \; n \; logL|$ change is less than 0.003. This $|2 \; n \; logL|$ value has been found sufficiently strict for the chi-square test of fit of H_0 against H_1 to be numerically precise. In problems where the $|2 \; n \; logL|$ criterion is not fulfilled, a stricter convergence criterion than 0.0001 for the parameters can be used so that the $|2 \; n \; logL|$ criterion becomes fulfilled.

APPENDIX 7
GROWTH MODELING

Random effect (random ccoefficient) growth modeling fits within the Mplus modeling framework described in Appendix 2. The central idea is that the random effects are treated as latent variables. To indicate the generality of the framework, model specifications for multiple-population analysis, multiple-indicator analyses, and categorical outcomes are shown. For an introduction to growth modeling in the latent variable framework, see Muthén and Khoo (1998). For a discussion of more advanced growth modeling in the latent variable framework, particularly with randomized interventions, see Muthén and Curran (1997). Growth modeling with clustered data, also known as three-level modeling, is discussed in Muthén (1997). Growth modeling with categorical outcomes is discussed in Muthén (1996). For recent developments in the area of growth mixture modeling with latent trajectory classes, see Muthén (2000a).

MULTIPLE-POPULATION GROWTH MODELING

As an example, consider a simple linear growth model with a single time-invariant covariate x. Let y_{git} denote the outcome for population (group) g, individual i, and timepoint t ($t = 1, 2, \ldots, T$),

$$Level - 1: y_{git} = \eta_{g0i} + \eta_{g1i} \ a_t + \epsilon_{git}, \tag{134}$$

$$Level - 2a: \eta_{g0i} = \alpha_{g0} + \gamma_{g0} \ x_{gi} + \zeta_{g0i}, \tag{135}$$

$$Level - 2b: \eta_{g1i} = \alpha_{g1} + \gamma_{g1} \ x_{gi} + \zeta_{g1i}, \tag{136}$$

where η_{g0i} is an intercept growth factor, η_{g1i} is a slope growth factor, a_t refers to a time score, level 1 refers to variation over time, and level 2 refers to variation over individuals. The time score a_t is typically transformed to $0, 1, \ldots, T - 1$ to define the intercept factor η_{g0i} as an initial status factor. In the growth model, level 1 corresponds to the measurement part of the general model in Appendix 2 while level 2 corresponds to the structural part.

Viewing the level 1 equation (134) as a measurement model shows the implicit assumption of time-invariant and population-invariant intercept 0 and slopes 1, t for the factors η_{g0i}, η_{g1i}. The level 2 parameters α_g, γ_g, $V(\zeta_g)$ capture structural differences across populations. As an equivalent, alternative parameterization, the model may be specified with a time- and population-invariant intercept ν in (134), instead fixing the intercept of the intercept factor equation to zero in the first group (g = 1), $\alpha_{10} = 0$.

APPENDIX 7

MULTIPLE INDICATOR GROWTH MODELING

Assume for simplicity a single population. Let y_{ijt} denote the outcome for individual i, indicator j, and timepoint t, and let η_{it} denote a latent variable construct,

$$Level - 1a\ (measurement\ part):$$
$$y_{ijt} = \nu_{jt} + \lambda_{jt}\ \eta_{it} + \epsilon_{ijt}, \tag{137}$$
$$Level - 1b: \eta_{it} = \eta_{0i} + \eta_{1i}\ a_t + \zeta_{it}, \tag{138}$$
$$Level - 2a: \eta_{0i} = \alpha_0 + \gamma_0\ x_i + \zeta_{0i}, \tag{139}$$
$$Level - 2b: \eta_{1i} = \alpha_1 + \gamma_1\ x_i + \zeta_{1i}. \tag{140}$$

Measurement invariance is specified by using time-invariant indicator intercepts and slopes:

$$\nu_{j1} = \nu_{j2} = \ldots = \nu_{jT} = \nu_j, \tag{141}$$
$$\lambda_{j1} = \lambda_{j2} = \ldots = \lambda_{jT} = \lambda_j, \tag{142}$$

where $\lambda_1 = 1$. The intercept of the level-2a equation is fixed at zero, $\alpha_0 = 0$. $V(\epsilon_{ijt})$ and $V(\zeta_{it})$ may vary over time. Structural differences are captured by letting $E(\eta_{it})$ and $V(\eta_{it})$ vary over time. With more than one population, across-population measurement invariance would be imposed and α_0 fixed to zero only in the first population.

GROWTH MODELING WITH ORDERED CATEGORICAL OUTCOME

Assume for simplicity a single population and a single outcome at each timepoint. With an ordered categorical outcome variable y_{it}, the level 1 equation in (134) is replaced by

$$Level - 1: y_{it}^* = \eta_{0i} + \eta_{1i}\ a_t + \epsilon_{it}. \tag{143}$$

In this model a key feature is the threshold $\tau_{t,c}$ for the c-category outcome variable y_{it} at time point t, $c = 0, 1, 2, \ldots, C-1$, where $\tau_0 = -\infty$, $\tau_C = \infty$. Across-time measurement invariance is imposed by the threshold specification

$$\tau_{1,c} = \tau_{2,c} = \ldots = \tau_{T,c}, \tag{144}$$

while the intercept of the level-2a equation is fixed at zero, $\alpha_0 = 0$. With more than one population, across-population invariance of thresholds would be imposed and α_0 fixed to zero only in the first population. Across-time differences in the variance of y_{it}^* need to be taken into account using the $\mathbf{\Delta}$ scaling matrix discussed in Appendix 2. These modeling issues are described in Muthén (1996).

APPENDIX 8
LATENT VARIABLE MIXTURE MODELING

Mixture modeling allows unobserved heterogeneity in the sample, where different individuals can belong to different subpopulations without the subpopulation membership being observed but instead inferred from the data. Mixture modeling captures this heterogeneity by a latent categorical variable. The modeling and estimation draws on Muthén and Shedden (1999) and Muthén, Shedden, and Spisic (1999) where technical details are given. Mixture modeling has a wide variety of applications. Overviews with latent class and growth mixture applications are given in Muthén (2000, 2001) and Muthén and Muthén (2000). Applications to randomized trials are given in Muthén, Brown, Masyn, Jo, Khoo, Yang, Wang, Kellam, Carlin, and Liao (2000), Jo (2000), and Jo and Muthén (2000, 2001). Applications to discrete-time survival analysis are given in Muthén and Masyn (2001).

THE GENERAL MODEL

Consider the observed variables $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{u}$, where $\mathbf{x}$ denotes a $q \times 1$ vector of covariates, $\mathbf{y}$ denotes a $p \times 1$ vector of continuous outcome variables, and $\mathbf{u}$ denotes an $r \times 1$ vector of binary and ordered polytomous categorical outcome variables. Consider latent variables $\boldsymbol{\eta}$ denoting an $m \times 1$ vector of continuous variables and $\mathbf{c}$ denoting a latent categorical variable with K classes, $\mathbf{c}_i = (c_{i1}, c_{i2}, \ldots, c_{iK})'$, where $c_{ik} = 1$ if individual i belongs to class k and zero otherwise. The model has three parts: $\mathbf{c}$ related to $\mathbf{x}$; $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$; and $\mathbf{y}$ related to $\mathbf{c}$ and $\mathbf{x}$.

The model relates $\mathbf{c}$ to $\mathbf{x}$ by multinomial logistic regression using the $K-1$-dimensional parameter vector of logit intercepts $\boldsymbol{\alpha}_c$ and the $(K-1) \times q$ parameter matrix of logit slopes $\boldsymbol{\Gamma}_c$, where for $k = 1, 2, \ldots, K$

$$P(c_{ik} = 1 | \mathbf{x}_i) = \frac{e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}{\sum_{k=1}^{K} e^{\alpha_{c_k} + \boldsymbol{\gamma}'_{c_k} \mathbf{x}_i}}, \tag{145}$$

where the last class is a reference class with coefficients standardized to zero, $\alpha_{c_K} = 0$, $\boldsymbol{\gamma}_{c_k} = \mathbf{0}$. The latent classes of $\mathbf{c}$ influence both $\mathbf{u}$ and $\mathbf{y}$. Consider first the $\mathbf{u}$ part of the model.

For $\mathbf{u}$, conditional independence is assumed given $\mathbf{c}_i$ and $\mathbf{x}_i$,

$$P(u_{i1}, u_{i2}, \ldots, u_{ir} | \mathbf{c}_i, \mathbf{x}_i) = P(u_{i1} | \mathbf{c}_i, \mathbf{x}_i) \, P(u_{i2} | \mathbf{c}_i, \mathbf{x}_i) \ldots P(u_{ir} | \mathbf{c}_i, \mathbf{x}_i). \tag{146}$$

The categorical variable $u_{ij}(j = 1, 2, \ldots, r)$ with S_j ordered categories follows an ordered polytomous logistic regression (proportional odds model), where for categories $s =$

APPENDIX 8

$0, 1, 2, \ldots, S_j - 1$ and $\tau_{j,k,0} = -\infty$, $\tau_{j,k,S_j} = \infty$,

$$u_{ij} = s, \ if \quad \tau_{j,k,s} < u_{ij}^* \le \tau_{j,k,s+1}, \tag{147}$$

$$P(u_{ij} = s | \mathbf{c}_i, \mathbf{x}_i) = F_{s+1}(u_{ij}^*) - F_s(u_{ij}^*), \tag{148}$$

$$F_s(u^*) = \frac{1}{1 + e^{-(\tau_s - u^*)}}, \tag{149}$$

where for $\mathbf{u}_i^* = (u_{i1}^*, u_{i2}^*, \ldots, u_{ir}^*)'$, $\boldsymbol{\eta}_{ui} = (\eta_{u_1 i}, \eta_{u_2 i}, \ldots, \eta_{u_f i})'$, and conditional on class k,

$$\mathbf{u}_i^* = \boldsymbol{\Lambda}_{u_k} \boldsymbol{\eta}_{ui} + \mathbf{K}_{u_k} \mathbf{x}_i, \tag{150}$$

$$\boldsymbol{\eta}_{ui} = \boldsymbol{\alpha}_{u_k} + \boldsymbol{\Gamma}_{u_k} \mathbf{x}_i, \tag{151}$$

where $\boldsymbol{\Lambda}_{u_k}$ is an $r \times f$ logit parameter matrix varying across the K classes, $\mathbf{K}_{u_k}$ is an $r \times q$ logit parameter matrix varying across the K classes, $\boldsymbol{\alpha}_{u_k}$ is an $f \times 1$ vector logit parameter vector varying across the K classes, and $\boldsymbol{\Gamma}_{u_k}$ is an $f \times q$ logit parameter matrix varying across the K classes. The thresholds may be stacked in the $\sum_{j=1}^r (S_j - 1) \times 1$ vectors $\boldsymbol{\tau}_k$ varying across the K classes.

It should be noted that (150) does not include intercept terms given the presence of τ parameters. Furthermore, τ parameters have opposite signs than u^* in (150) because of their interpretation as thresholds or cutpoints that a latent continuous response variable u^* exceeds or falls below (see also Agresti, 1990, pp. 322-324). For example, with a binary u_j scored 0/1 (148) leads to

$$P(u_{ij} = 1 | \mathbf{c}_i, \mathbf{x}_i) = 1 - \frac{1}{1 + e^{-(\tau - u^*)}}, \tag{152}$$

$$= \frac{1}{1 + e^{- \ logit}}, \tag{153}$$

where $logit = -\tau + u^*$. For example, the higher the τ the higher u^* needs to be to exceed it, and the lower the probability of $u = 1$.

The model structure in (150) and (151) is useful when the $\mathbf{u}$ vector represents repeated measures and the latent classes correspond to different trajectory classes. In this case, the elements of $\boldsymbol{\eta}_u$ correspond to growth factors in random effects growth modeling, except that $\boldsymbol{\eta}_u$ has zero variance conditional on $\mathbf{x}$. The parameterization of this type of growth model is shown in the section Latent Class Growth Analysis below.

Consider next the $\mathbf{y}$ part of the model. Multivariate normality is assumed for $\mathbf{y}$ conditional on $\mathbf{x}$ and class k,

$$\mathbf{y}_i = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \boldsymbol{\eta}_i + \mathbf{K}_k \mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{154}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha}_k + \mathbf{B}_k \boldsymbol{\eta}_i + \boldsymbol{\Gamma}_k \mathbf{x}_i + \boldsymbol{\zeta}_i, \tag{155}$$

where the residual vector $\boldsymbol{\epsilon}_i$ is $N(\mathbf{0}, \boldsymbol{\Theta}_k)$ and the residual vector $\boldsymbol{\zeta}_i$ is $N(\mathbf{0}, \boldsymbol{\Psi}_k)$, both assumed to be uncorrelated with other variables. This part of the mixture model builds on the general model of Appendix 2 generalized to the K classes of the mixture, except that multiple-group analysis is not included.

ESTIMATION

The Mplus mixture model is estimated by maximum-likelihood. The observed-data log likelihood is

$$logL = \sum_{i=1}^{n} log[\mathbf{y}_i, \mathbf{u}_i|\mathbf{x}_i], \tag{156}$$

where $[\mathbf{y}_i, \mathbf{u}_i|\mathbf{x}_i]$ is a mixture distribution defined as

$$\sum_{k=1}^{K} P(c_{ik} = 1|\mathbf{x}_i) \, [\mathbf{u}_i|c_{ik} = 1, \mathbf{x}_i] \, [\mathbf{y}_i|c_{ik} = 1, \mathbf{x}_i], \tag{157}$$

where $[\mathbf{y}_i|c_{ik} = 1, \mathbf{x}_i]$ is $N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$,

$$\boldsymbol{\mu}_i = \boldsymbol{\nu}_k + \boldsymbol{\Lambda}_k \, (\mathbf{I} - \mathbf{B}_k)^{-1} \, (\boldsymbol{\alpha}_k + \boldsymbol{\Gamma}_k \, \mathbf{x}_i) + \mathbf{K}_k \, \mathbf{x}_i, \tag{158}$$

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Lambda}_k \, (\mathbf{I} - \mathbf{B}_k)^{-1} \, \boldsymbol{\Psi}_k \, (\mathbf{I} - \mathbf{B}_k)^{-1'} \, \boldsymbol{\Lambda}_k' + \boldsymbol{\Theta}_k. \tag{159}$$

The maximum-likelihood estimation uses an EM algorithm (see Muthén, Shedden, & Spisic, 1999), where data on $\mathbf{c}$ are considered missing. The complete-data log likelihood is

$$\sum_{i=1}^{n} (log[\mathbf{c}_i|\mathbf{x}_i] + log[\mathbf{u}_i|\mathbf{c}_i, \mathbf{x}_i] + log[\mathbf{y}_i|\mathbf{c}_i, \mathbf{x}_i]). \tag{160}$$

In the E step, the conditional expectation of the complete-data log likelihood given the data involves the conditional probability of individual i belonging to class k, given the observed data,

$$p_{ik} = P(c_{ik} = 1|\mathbf{y}_i, \mathbf{u}_i, \mathbf{x}_i) = P(c_{ik} = 1|\mathbf{x}_i) \, [\mathbf{u}_i|c_{ik} = 1, \mathbf{x}_i] \, [\mathbf{y}_i|\mathbf{c}_i, \mathbf{x}_i]/[\mathbf{y}_i, \mathbf{u}_i|\mathbf{x}_i]. \tag{161}$$

From a Bayesian point of view this is the posterior probability of group membership.

Maximizing the expected complete-data log likelihood leads to a separate M step for each of the three model parts: $\mathbf{c}$ related to $\mathbf{x}$; $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$; and $\mathbf{y}$ related to $\mathbf{c}$ and $\mathbf{x}$.

The maximization for $\mathbf{c}$ related to $\mathbf{x}$ leads to a multinomial regression optimization,

$$\sum_{n=1}^{n} \sum_{k=1}^{K} p_{ik} \, log \, P(c_{ik} = 1|\mathbf{x}_i). \tag{162}$$

The maximization for $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$ leads to a logistic regression optimization,

$$\sum_{i=1}^{n} \sum_{j=1}^{r} \sum_{k} p_{ik} \, log \, P(u_{ij} = 1|\mathbf{c}_i, \mathbf{x}_i). \tag{163}$$

APPENDIX 8

For $\mathbf{y}$ related to $\mathbf{c}$ and $\mathbf{x}$,

$$[\mathbf{y}_i|\mathbf{c}_i, \mathbf{x}_i] = [\mathbf{y}_i|\mathbf{x}_i]_1^{c_{i1}} \ [\mathbf{y}_i|\mathbf{x}_i]_2^{c_{i2}} \ \ldots [\mathbf{y}_i|\mathbf{x}_i]_k^{c_{ik}} \ldots [\mathbf{y}_i|\mathbf{x}_i]_K^{c_{iK}}, \tag{164}$$

where $[\mathbf{y}_i|\mathbf{x}_i]_k$ is multivariate normal with mean vector and covariance matrix given in (158) and (159). It follows that in (160),

$$\sum_{i=1}^{n} log[\mathbf{y}_i|\mathbf{c}_i, \mathbf{x}_i] = \sum_{i=1}^{n}\sum_{k=1}^{K} c_{ik} \ log[\mathbf{y}_i|\mathbf{x}_i]_k. \tag{165}$$

so that the maximization considers

$$E(\sum_{i=1}^{n} log[\mathbf{y}_i|\mathbf{c}_i, \mathbf{x}_i] \ |\mathbf{u}_i, \mathbf{y}_i, \mathbf{x}_i) = \sum_{i=1}^{n}\sum_{k=1}^{K} p_{ik} \ log[\mathbf{y}_i|\mathbf{x}_i]_k, \tag{166}$$

which corresponds to multiple-group analysis of K groups with posterior-probability weighted sample mean vectors and covariance matrices. Missing data is allowed for on $\mathbf{y}$ and $\mathbf{u}$, assuming missing at random (MAR; Little & Rubin, 1987).

Standard errors of parameter estimates are computed using three alternative methods (see, e.g., Amemiya, 1985, chapter 4, and also McLachlan & Peel, 2000, chapter 2). Writing the observed-data log likelihood as

$$logL = \sum_{i=1}^{n} logL_i, \tag{167}$$

the MLF approach approximates the Fisher information matrix using

$$I_{MLF} = \sum_{i=1}^{n} \frac{\partial logL_i}{\partial \boldsymbol{\pi}} \times \frac{\partial logL_i}{\partial \boldsymbol{\pi}'}, \tag{168}$$

the ML approach approximates the Fisher information matrix using

$$I_{ML} = -\sum_{i=1}^{n} \frac{\partial^2 logL_i}{\partial \boldsymbol{\pi} \ \partial \boldsymbol{\pi}'}, \tag{169}$$

and the MLR approach approximates the Fisher information matrix using

$$I_{MLR} = I_{ML}^{-1} \ I_{MLF} \ I_{ML}^{-1}. \tag{170}$$

The I_{MLR} alternative is designed to be robust against misspecification of the likelihood. Limited Monte Carlo simulations under correct specification of the likelihood indicate that for smaller sample sizes the MLR standard errors perform slightly better than those of ML and the standard errors of ML perform better than those of MLF.

It is often difficult to verify that a mixture model is identified. The invertability of the estimated information matrix used to produce the standard errors provides an empirical guide to assessing if a model is identified. Observing a change in the log likelihood when freeing a parameter in an identified model is another empirical guide.

TRAINING DATA

The numerical and statistical performance of mixture modeling benefits from confirmatory analysis. Mplus mixture modeling offers the same kind of confirmatory analysis as in regular non-mixture modeling, using a priori restrictions on the parameters. With mixture modeling, however, there is also a second type of confirmatory analysis. With a categorical latent variable c, a researcher may want to incorporate the hypothesis that certain individuals are known to represent certain classes. Individuals with known class membership are referred to as training data (see also McLachlan & Basford, 1988; Hosmer, 1973). As is the case with using parameter restrictions, the numerical and statistical performance of the mixture analysis benefits from incorporating training data. Multiple-group modeling corresponds to the case of all sample units contributing training data so that c is in effect an observed categorical variable. An example of mixture analysis with training data is the Complier Average Causal Effect modeling of Little and Yau (1998), see also Jo and Muthén (2000, 2001), drawing on Rubin's causal model.

The training data typically consists of 0 and 1 class membership values for all individuals, where 1 denotes which classes an individual may belong to. Known class membership for an individual corresponds to having training data value of 1 for the known class and 0 for all other classes. Unknown class membership for an individual is specified by the value 1 for all classes. With class membership training data, the class probabilities in (161) are renormed for each individual to add to one over the admissible set of classes. Fractional training data is also allowed, corresponding to class probabilities adding to unity for each individual. With fractional training data, the class probabilities are taken to be fixed quantities, which reduces the sampling variability accounted for in the standard error calculations. Fractional training data where each individual has a probability of 1 for one class and 0's for the other classes is equivalent to training data with class membership value 1 for only one class for each individual. Using training data with a value of 1 for one class and 0's for the other classes makes it possible to perform multinomial logistic regression with an unordered, polytomous observed dependent variable using the Mplus model part where **c** is related to **x**. More general analyses are also possible using training data in this way, such as path analysis with an unordered, polytomous observed mediating variable, extending the model to include the **u** and **y** parts.

MODEL TESTS OF FIT

Tests of model fit can be obtained from the log likelihood value for a given model. Based on this, a likelihood ratio chi-square value can be computed for nested models. It is not appropriate to use such values for comparing models with different number of classes, however, given that this involves inadmissible parameter values of zero class probabilities. For such comparisons, AIC and BIC information criteria can be used instead. The sample-size

adjusted BIC given in Appendix 5 was found to give superior performance in a simulation study for latent class analysis models in Yang (1998).

The degree to which the latent classes are clearly distinguishable by the data and the model can be assessed by using the estimated posterior probabilities in (161) for each individual. By classifying each individual into his/her most likely class, a table can be constructed with rows corresponding to individuals classified into a given class, and where for individuals in each row, the column entries give the average conditional probabilities (Nagin, 1999). High diagonal and low off-diagonal values indicate good classification quality. A summary measure of the classification is given by the entropy measure (see, e.g., Ramaswamy, DeSarbo, Reibstein, Robinson, 1993),

$$E_K = 1 - \frac{\sum_i \sum_k (-\hat{p}_{ik} \, ln\hat{p}_{ik})}{n \, lnK}, \tag{171}$$

where $\hat{p}_{ik}$ denotes the estimated conditional probability for individual i in class k. Entropy values range from zero to one, where entropy values close to one indicate clear classifications in that the entropy decreases for probability values that are not close to zero or one.

The fit of the model to the data can be studied by comparing for each class estimated moments with moments created by weighting the individual data by the estimated posterior probabilities (Roeder, Lynch & Nagin, 1999). The Mplus TECH7 output gives the means, variances, and covariances computed in this way. Plots are useful to study how well the model fits individual observations. For example, with growth mixture modeling one may check how closely the estimated average trajectory within each class matches the data by randomly assigning individuals to classes based on individual estimated posterior class probabilities and for each class plotting the observed individual trajectories together with the model-estimated average trajectory (Bandeen-Roche et al., 1997; Muthén et al., 2000).

When the model contains only **u**, Pearson and likelihood ratio chi-square tests against the unrestricted multinomial alternative are provided,

$$\chi_P^2 = \sum_{cells} \frac{(o_i - e_i)^2}{e_i}, \tag{172}$$

$$\chi_L^2 = 2 \sum_{cells} o_i \, log \, o_i/e_i, \tag{173}$$

where o_i is the observed frequency in cell i of the multivariate frequency table for **u** and e_i is the corresponding frequency estimated under the model. Low cell frequency often arise when there are many u variables and this makes the chi-square approximation deteriorate (Agresti, 1990). Cells with zero observed frequency do not cause numerical problems, but cells with non-zero observed frequencies and very low expected frequencies are detrimental. A cell that has non-zero observed frequency and expected frequency less than .01 is not included in the χ^2 computation as the default. With missing data on **u**, the EM algorithm

described in Little and Rubin (1987; chapter 9.3, pp. 181-185) is used to compute the estimated frequencies in the unrestricted multinomial model. In this case, a test of MCAR for the unrestricted model is also provided (Little & Rubin, 1987, pp. 192-193).

NUMERICAL ISSUES AND TECHNIQUES

Mixture modeling can involve numerical and statistical problems. Mixture modeling is known to sometimes generate a likelihood function with several local maxima. The occurrence of this depends on the model and the data. It is therefore recommended that for a given dataset and a given model different optimizations are carried out using different sets of starting values. Some models are also less stable or less well defined, as evidenced by very slow convergence, decreases in the log likelihood due to failed M step iterations, a non-positive definite Fisher information matrix, or other computational problems. Models that are in theory identified can in certain samples and with certain starting values lead to a non-positive definite Fisher information matrix, showing singularity or a saddle point. Models with across-class variation in covariance matrices are known to be generally less stable than models with no such across-class variation. Certain models are also prone to classes collapsing, leading to zero class counts. Information on these problems is provided in the Mplus output. General references for mixture modeling include Everitt and Hand (1981), McLachlan and Basford (1988), McLachlan and Peel (2000), and Titterington, Smith, and Makov (1985).

The M steps for $\mathbf{c}$ related to $\mathbf{x}$ and for $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$ are carried out using Newton-Raphson with a switch to quasi-Newton if the Hessian is not negative definite. The M step for $\mathbf{y}$ related to $\mathbf{c}$ and $\mathbf{x}$ is carried out using quasi-Newton. The M steps need not be carried to full convergence (McLachlan & Krishnan, 1997). The default settings are 2 iterations for Newton-Raphson and full convergence for quasi-Newton. Convergence is monitored using the observed-data log likelihood. The following four criteria must all be fulfilled. First, the absolute and relative log likelihood change must both be smaller than 0.0000001. Second, the change in any class count must be smaller than 0.001. Third, the observed-data log likelihood derivative criterion must be fulfilled, where the default derivative criterion is that the observed-data log likelihood derivatives divided by n for all free parameters in the full model are less than 0.000001. Fourth, it must be possible to compute the standard errors (information matrix positive definite).

In the M step for $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$, the conditional u probabilities go toward zero or one in some applications, corresponding to infinitely small or large logit parameter values, where for a binary u,

$$Probability = \frac{1}{1 + e^{-Logit}}. \tag{174}$$

As a default for standard models ($\mathbf{K}_u = \mathbf{0}$ and no parameter constraints), these logit parameter values are not allowed to go outside the range of -15 to $+15$ (probability range of

0.0000003 to 0.9999997) in order to avoid numerical difficulties. When such extreme parameter values have been encountered, the model can be re-estimated holding these parameters fixed.

COMPUTATIONAL IMPLEMENTATION

In the current computational implementation, $\mathbf{\Lambda}_{u_k}$ is an array of constants. The parameter array $\mathbf{K}_{u_k}$ is only used with binary u's, whereas with polytomous u's these direct effects are handled by $\mathbf{\Gamma}_{u_k}$, defining an η_u variable behind the u variable so that $u_i^* = \eta_{ui}$. With binary u's and $\boldsymbol{\eta}_u$ not included in the model, a set of intercepts referred to as $\mathbf{\Lambda}_u$ replace the first term on the right-hand side of (150). With polytomous u's, or when $\boldsymbol{\eta}_u$ is included in the model, the subscript u is replaced by the subscript f (growth factor) in (150) and (151). The computational implementation for the $\mathbf{y}$ part of the model is that of Appendix 2 for the case where all y variables are continuous.

GROWTH MIXTURE MODELING

Growth mixture modeling was introduced in Muthén and Shedden (1999), Muthén (2000), and Muthén et al. (2000). As an example, consider a simple linear growth model with a single time-invariant covariate x. Let y_{kit} denote the outcome for latent class k, individual i, and timepoint t ($t = 1, 2, \ldots, T$),

$$y_{kit} = \eta_{k0i} + \eta_{k1i} \ a_t + \epsilon_{kit}, \tag{175}$$

$$\eta_{k0i} = \alpha_{k0} + \gamma_{k0} \ x_{ki} + \zeta_{k0i}, \tag{176}$$

$$\eta_{k1i} = \alpha_{k1} + \gamma_{k1} \ x_{ki} + \zeta_{k1i}, \tag{177}$$

where $V(\epsilon_{kit}) = \mathbf{\Theta}_k$, $V(\zeta_{ki}) = \mathbf{\Psi}_k$. Because there is no intercept term ν in (175), both α_{k0} and α_{k1} are free to be estimated in all classes.

LATENT CLASS GROWTH ANALYSIS

Latent class growth analysis refers to growth mixture modeling with $\mathbf{\Psi}_k = \mathbf{0}$. This type of modeling was discussed in Nagin (1999). Consider the case where the outcomes are categorical so that the elements of the categorical variable vector $\mathbf{u}$ correspond to repeated measures. The random effects are captured by $\boldsymbol{\eta}_u$ in (151), where it should be noted that $\boldsymbol{\eta}_u$ has a zero covariance matrix given $\mathbf{x}$. Assume for simplicity a single outcome at each timepoint, $\mathbf{u}_i^* = (u_{i1}, u_{i2}, \ldots, u_{it}, \ldots, u_{iT})'$, and the simple growth model corresponding to (150),

$$Level - 1 : u_{it}^* = \eta_{0i} + \eta_{1i} \ a_t, \tag{178}$$

where a_t are fixed time scores represented in $\mathbf{\Lambda}_u$. With an ordered categorical outcome variable u_{it}, let $\tau_{t,k,s}$ be the sth threshold in class k at timepoint t, $s = 0, 1, 2, \ldots, S_t - 1$, where $\tau_{t,k,0} = -\infty$, $\tau_{t,k,S_t} = \infty$. Across-time and across-class measurement invariance is imposed by the threshold specification

$$\tau_{1,1,s} = \tau_{2,1,s} = \ldots = \tau_{T,1,s} = \ldots = \tau_{1,K,s} = \ldots = \tau_{T,K,s}, \tag{179}$$

for each s value. In the level-2 equation corresponding to (151), the α mean of the intercept growth factor η_{0i} is fixed at zero in the first class for identification purposes. The mean of the intercept growth factor is free to be estimated in the remaining classes. An example of latent class growth analysis with repeated measures on multiple binary outcomes is given in Muthén (2001).

DISCRETE-TIME SURVIVAL ANALYSIS

Discrete-time survival analysis (see, e.g. Singer & Willet, 1993) uses the categorical variable vector $\mathbf{u}$ to represent events modeled by a logistic hazard function (cf Muthén & Masyn, 2001). Consider a set of binary 0/1 variables $u_j, j = 1, 2, \ldots, r$, where $u_{ij} = 1$ if individual i experiences the non-repeatable event in time period j and define j_i as the last time period in which data were collected for individual i. The hazard is the probability of experiencing the event in time period j given that it was not experienced prior to j. The likelihood is

$$\prod_{i=1}^{n} \prod_{j=1}^{j_i} h_{ij}^{u_{ij}} (1 - h_{ij})^{1-u_{ij}}, \tag{180}$$

where with q covariates $\mathbf{x}$ the hazard is written as

$$h_{ij} = \frac{1}{1 + e^{-(-\tau_j + \kappa'_j \mathbf{x}_i)}}, \tag{181}$$

where a proportional-odds assumption is obtained by dropping the j subscript for κ'_j. The survival function is

$$S_{ij} = \prod_{k=1}^{j} (1 - h_{ik}). \tag{182}$$

Discrete-time survival analysis is fitted into the general mixture model above by noting that the likelihood in (180) is the same as for $\mathbf{u}$ related to $\mathbf{c}$ and $\mathbf{x}$ in a single-class model, i.e. with u's independent conditional on $\mathbf{x}$. κ'_j is the jth row of $\mathbf{K}_u$ in (150). The fact that individual i does not have observations on u after time period j_i is handled as missing data. For example, with five time periods ($r = 5$), an individual who experiences the event in period four has the data vector $\mathbf{u}'_i$

$$\begin{pmatrix} 0 & 0 & 0 & 1 & 999 \end{pmatrix},$$

with 999 representing missing data. An individual who is censored in period five has the data vector $\mathbf{u}'_i$

$$\left(\begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \end{array} \right),$$

while an individual who is censored in period four has the data vector $\mathbf{u}'_i$

$$\left(\begin{array}{ccccc} 0 & 0 & 0 & 0 & 999 \end{array} \right).$$

Muthén and Masyn (2001) also proposes general mixture discrete-time survival analysis models, where different latent classes have different hazard and survival functions. For example, a growth mixture model for $\mathbf{y}$ can be combined with a survival model for $\mathbf{u}$, where the threshold parameters are held invariant across the latent classes,

$$\tau_{j1} = \tau_{j2} = \ldots = \tau_{jK}; j = 1, 2, \ldots, r, \tag{183}$$

but where the latent class membership for the growth model influences the hazard function through logit shift parameters α as in latent class growth analysis with an intercept growth factor,

$$u^*_{it} = \eta_{0i}, \tag{184}$$

where the α mean of the intercept growth factor η_{0i} is fixed at zero in the first class for identification purposes and free to be estimated in the remaining classes.

APPENDIX 9
AGGREGATED ANALYSIS UNDER COMPLEX SAMPLING

Mplus can analyze two-level data obtained by cluster sampling and unequal probability weights. This analysis assumes continuous y variables. These variables do not, however, need to be normally distributed. Missing data is not allowed, but a listwise present sample is assumed.

Aggregated modeling is described in Muthén and Satorra (1995b). It is defined as analysis of a conventional sample mean vector and covariance matrix. If weights are present, a weighted sample mean vector and a weighted sample covariance matrix is used. The F_{ML} fitting function of (96) is computed. Standard errors of estimates and chi-square model testing are provided using methods that are robust against the non-independence of observations due to the clustering as well as robust against deviations from normality. These are the methods described in Appendix 4 with a complex-sample specific modification of how the estimate of the $\mathbf{\Gamma}$ matrix is computed.

The $\mathbf{\Gamma}$ matrix used for robust standard error and chi-square model test computations is obtained as follows. Considering a certain group (population) g, but dropping the group index for simplicity, let $\mathbf{v}'_{ci} = (\mathbf{y}'_{ci}, \mathbf{x}'_{ci})$ denote the vector of observed variables for individual i in cluster c ($c = 1, 2, \ldots, C$) and define the data vector $\mathbf{d}_c$ for cluster c

$$
\mathbf{d}_c = \sum_{i=1}^{n_c} w_{ci} \begin{pmatrix} -v_{ci1} \\ \vdots \\ -v_{cip} \\ (v_{ci1} - \bar{v}_1) \quad (v_{ci1} - \bar{v}_1) \\ (v_{ci2} - \bar{v}_2) \quad (v_{ci1} - \bar{v}_1) \\ \vdots \end{pmatrix}
$$

where w_{ci} denotes the sampling weights and

$$
\begin{pmatrix} \bar{\mathbf{v}} \\ \mathbf{s} \end{pmatrix} = \mathrm{vec} \begin{pmatrix} \bar{\mathbf{v}} \\ \mathbf{S} \end{pmatrix} = n^{-1} \sum_{c=1}^{C} \mathbf{d}_c
$$

where $n = \sum_{c=1}^{C} n_c$. The asymptotic covariance matrix can be expressed as

$$
\mathbf{\Gamma} = V \begin{pmatrix} \bar{\mathbf{v}} \\ \mathbf{S} \end{pmatrix} = n^{-2} \sum_{c=1}^{C} V(\mathbf{d}_c)
$$

where $V(\mathbf{d}_c)$ is estimated by the sampling variance across clusters (Muthén & Satorra, 1995b).

APPENDIX 9

APPENDIX 10
TWO-LEVEL (DISAGGREGATED) ANALYSIS UNDER COMPLEX SAMPLING

This appendix gives a brief technical summary of some key aspects of multilevel latent variable modeling in Mplus. For a fuller description, see Muthén (1989b, 1990, 1994) and Muthén and Satorra (1995b). An application of multilevel path analysis is given in Muthén (1989b), an application of multilevel factor analysis is given in Muthén (1991), three-level modeling of growth in two parallel processes is described in Muthén (1997), and multilevel multiple-group factor analysis modeling is described in Muthén, Khoo, and Gustafsson (1998).

The multilevel analysis in Mplus is referred to as two-level modeling or disaggregated analysis. This modeling is suitable for hierarchical data observed on two levels such as students within schools. Two-level modeling is also suitable for three-level data when one of the three levels corresponds to repeated measures over time because the repeated measures are not treated as a third level but as a multivariate observation vector. In multilevel regression terms, the Mplus two-level modeling handles random intercepts but not random slopes. An exception is growth models for repeated measures where random slopes can be handled.

The two levels of the model are referred to as between and within. The following general mean and covariance structure model is used for the between and within levels,

$$\boldsymbol{\mu} = \boldsymbol{\nu}_B + \boldsymbol{\Lambda}_B \left(\mathbf{I} - \mathbf{B}_B\right)^{-1} \boldsymbol{\alpha}_B, \tag{185}$$

$$\boldsymbol{\Sigma}_B = \boldsymbol{\Lambda}_B \left(\mathbf{I} - \mathbf{B}_B\right)^{-1} \boldsymbol{\Psi}_B \left(\mathbf{I} - \mathbf{B}_B\right)'^{-1} \boldsymbol{\Lambda}'_B + \boldsymbol{\Theta}_B, \tag{186}$$

$$\boldsymbol{\Sigma}_W = \boldsymbol{\Lambda}_W \left(\mathbf{I} - \mathbf{B}_W\right)^{-1} \boldsymbol{\Psi}_W \left(\mathbf{I} - \mathbf{B}_W\right)'^{-1} \boldsymbol{\Lambda}'_W + \boldsymbol{\Theta}_W. \tag{187}$$

Various special cases of modeling that lead to this mean and covariance structure model are described in the papers listed above. From a general point of view, two-level modeling considers the vector of observed variables containing cluster-specific (level 2) variables $\mathbf{z}_c$ ($c = 1, 2, \ldots, C$) and individual-specific (level 1) variables ($\mathbf{y}_{ci}$ and $\mathbf{x}'_{ci}$) for individual i in cluster c, where

$$\mathbf{v}_{ci} = \begin{pmatrix} \mathbf{z}_c \\ \mathbf{y}_{ci} \\ \mathbf{x}_{ci} \end{pmatrix} = \mathbf{v}_c^* + \mathbf{v}_{ci}^* = \begin{pmatrix} \mathbf{v}_{zc}^* \\ \mathbf{v}_{yc}^* \\ \mathbf{v}_{xc}^* \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{v}_{yci}^* \\ \mathbf{v}_{xci}^* \end{pmatrix},$$

$$E(\mathbf{v}_{ci}) = E(\mathbf{v}_c^*) + \mathbf{0} = \boldsymbol{\mu}, \tag{188}$$

$$V(\mathbf{v}_{ci}) = V(\mathbf{v}_c^*) + V(\mathbf{v}_{ci}^*) = \boldsymbol{\Sigma}_B + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_W \end{pmatrix}, \tag{189}$$

APPENDIX 10

where

$$\mathbf{\Sigma}_B = \left(\begin{array}{cc} \mathbf{\Sigma}_{Bzz} & \mathbf{\Sigma}'_{Byx,z} \\ \mathbf{\Sigma}_{Byx,z} & \mathbf{\Sigma}_{Byx,yx} \end{array} \right).$$

Using the between-cluster (level 2) model

$$\mathbf{v}_c^* = \boldsymbol{\nu}_B + \mathbf{\Lambda}_B \, \boldsymbol{\eta}_{Bc} + \boldsymbol{\epsilon}_{Bc}, \tag{190}$$

$$\boldsymbol{\eta}_{Bc} = \boldsymbol{\alpha}_B + \mathbf{B}_B \, \boldsymbol{\eta}_{Bc} + \boldsymbol{\zeta}_{Bc} \tag{191}$$

and the within-cluster (level 1) model

$$\left(\begin{array}{c} \mathbf{v}_{yci}^* \\ \mathbf{v}_{xci}^* \end{array} \right) = \mathbf{\Lambda}_W \, \boldsymbol{\eta}_{Wci} + \boldsymbol{\epsilon}_{Wci}, \tag{192}$$

$$\boldsymbol{\eta}_{Wci} = \mathbf{B}_W \, \boldsymbol{\eta}_{Wci} + \boldsymbol{\zeta}_{Wci}, \tag{193}$$

results in the general mean and covariance structure model given in (185), (186), and (187). Assuming normality for $\mathbf{v}_{ci}$, the likelihood can be expressed in terms of a sum of D distinct level 2 terms involving both between- and within-cluster parameters and one level 1 term involving within-cluster parameters (cf. Muthén, 1990). The level 2 terms have the mean vector and covariance matrix structure

$$\sqrt{s} \, \boldsymbol{\mu}, \tag{194}$$

$$\mathbf{\Sigma}_B^* = \left(\begin{array}{cc} s \, \mathbf{\Sigma}_{Bzz} & s \, \mathbf{\Sigma}'_{Byx,z} \\ s \, \mathbf{\Sigma}_{Byx,z} & \mathbf{\Sigma}_W + s \, \mathbf{\Sigma}_{Byx,yx} \end{array} \right), \tag{195}$$

where s is the cluster size. The level 1 term has mean vector zero and covariance matrix $\mathbf{\Sigma}_W$. When data have equal cluster sizes (balanced data), the maximum-likelihood estimator considers the fitting function

$$C \left\{ ln \mid \mathbf{\Sigma}_B^* \mid + tr[\mathbf{\Sigma}_B^{*-1} (\mathbf{S}_B + s \, (\bar{\mathbf{v}} - \boldsymbol{\mu})(\bar{\mathbf{v}} - \boldsymbol{\mu})')] \right\} + (n - C) \left\{ ln \mid \mathbf{\Sigma}_W \mid + tr[\mathbf{\Sigma}_W^{-1} \mathbf{S}_W] \right\}, \tag{196}$$

where $\mathbf{\Sigma}_B^*$ is given in (195) and n is the total sample size. In structural equation modeling terms, this can be represented as a two-group model for between and within. This fitting function has been found to give estimates close to maximum likelihood also when cluster sizes are not equal (unbalanced data) and represents an estimator in its own right. This is called the MUML estimator and is used in Mplus. In the unbalanced case, the standard errors and chi-square tests of fit obtained when applying conventional two-group structural equation modeling are not correct. Mplus uses the estimator (196) with robust standard errors and chi-square tests of model fit specifically derived for unbalanced data.

It is instructive to consider the between and within sample covariance matrices for the within-level observation vector $\mathbf{y}$,

$$\mathbf{S}_B = (C - 1)^{-1} \sum_{c=1}^{C} n_c \, (\bar{\mathbf{y}}_c - \bar{\mathbf{y}})(\bar{\mathbf{y}}_c - \bar{\mathbf{y}})', \tag{197}$$

$$\mathbf{S}_W = (n - C)^{-1} \sum_{c=1}^{C} \sum_{i=1}^{n_c} (\mathbf{y}_{ci} - \bar{\mathbf{y}}_c)(\mathbf{y}_{ci} - \bar{\mathbf{y}}_c)', \tag{198}$$

where n_c is the size of cluster c. When dividing by C instead of $C - 1$ in (197), (196) gives maximum-likelihood estimation in the balanced case. Using $C - 1$,

$$E(\mathbf{S}_B) = \mathbf{\Sigma}_W + s\,\mathbf{\Sigma}_B, \tag{199}$$

$$E(\mathbf{S}_W) = \mathbf{\Sigma}_W, \tag{200}$$

where s is defined as (see, e.g., Searle, Casella, McCullogh, 1992, p. 71)

$$s = [n^2 - \sum_{c=1}^{C} n_c^2]\,[n(C - 1)]^{-1}. \tag{201}$$

This definition of s is used in the fitting function (196). It simplifies to the common cluster size in the balanced case. It follows that while $\hat{\mathbf{\Sigma}}_W = \mathbf{S}_W$, $\mathbf{\Sigma}_B$ is not estimated by $\mathbf{S}_B$ but instead as

$$\hat{\mathbf{\Sigma}}_B = (\mathbf{S}_B - \mathbf{S}_W)/s. \tag{202}$$

The intraclass correlation for y_j, ρ_j, is estimated as the ratio of between variance and total variance,

$$\rho_j = [\hat{\mathbf{\Sigma}}_B]_{jj}/([\hat{\mathbf{\Sigma}}_B]_{jj} + [\hat{\mathbf{\Sigma}}_W]_{jj}). \tag{203}$$

COMPUTATIONAL IMPLEMENTATION

The computational implementation considers an observed-variable vector with both level 1 and level 2 variables, $\mathbf{v}' = (\mathbf{z}_y', \mathbf{z}_x', \mathbf{y}', \mathbf{x}')$, where $\mathbf{z}$ is divided into outcome variables $\mathbf{z}_y$ and covariates $\mathbf{z}_x$. Artificial latent variables are introduced to capture the level-2 variation in the observed variables and to reproduce the structures in (194) and (195). For two-level modeling the latent variable vector $\boldsymbol{\eta}' = (\boldsymbol{\eta}_B', \boldsymbol{\eta}_W', \boldsymbol{\eta}_{Bzy}', \boldsymbol{\eta}_{Bzx}', \boldsymbol{\eta}_{By}', \boldsymbol{\eta}_{Wyd}', \boldsymbol{\eta}_{Bx}', \boldsymbol{\eta}_{Wx}')$, where $\boldsymbol{\eta}_B$ contains between-level (level 2) latent variables, $\boldsymbol{\eta}_W$ contains within-level (level 1) latent variables, $\boldsymbol{\eta}_{Bzy}$, $\boldsymbol{\eta}_{Bzx}$ and $\boldsymbol{\eta}_{By}$ contain artificial variable-specific between-level latent variables behind observed z_y and z_x variables and observed y variables, $\boldsymbol{\eta}_{Wyd}$ contains artificial latent variables behind observed y variables that are regressed on other variables for which they are not indicators, $\boldsymbol{\eta}_{Bx}$ contains artificial variable-specific between-level latent variables behind x variables, and $\boldsymbol{\eta}_{Wx}$ contains artificial variable-specific within-level latent variables behind x variables.

For level 2 terms, the parameter matrices are arranged as follows. In line with Chapter 2, y variables that are regressed on other variables for which they are not indicators, are transferred into latent variables in $\boldsymbol{\eta}_{Wyd}$ by letting $\mathbf{A}$ be a matrix of 0's and 1's with the number of rows equal to the number of y variables and the number of columns equal to

the length of $\boldsymbol{\eta}_{Wyd}$, the number of y variables that are involved, and letting $\mathbf{A}$ have unit elements for y variables that are involved. Let other parameter arrays affected by this transfer be denoted in line with Chapter 2.

$$\boldsymbol{\nu}'_2 = (\mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}'), \quad (204)$$

$$\boldsymbol{\Lambda}_2 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \sqrt{s}\,\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{s}\,\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{Wynd} & \mathbf{0} & \mathbf{0} & \sqrt{s}\,\mathbf{I} & \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sqrt{s}\,\mathbf{I} & \mathbf{I} \end{pmatrix} (205)$$

$$\boldsymbol{\Theta}_2 = \begin{pmatrix} \mathbf{0} & & & symm. \\ \mathbf{0} & \mathbf{0} & & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_{Wynd} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} (206)$$

$$\boldsymbol{\alpha}_2 = \begin{pmatrix} \boldsymbol{\alpha}_B \\ \mathbf{0} \\ \boldsymbol{\nu}_{Bzy} \\ \boldsymbol{\nu}_{Bzx} \\ \boldsymbol{\nu}_{By} \\ \mathbf{0} \\ \boldsymbol{\mu}_{Bx} \\ \mathbf{0} \end{pmatrix}, \quad (207)$$

$$\mathbf{B}_2 = \begin{pmatrix} \mathbf{B}_{BB} & \mathbf{0} & \mathbf{B}_{BBzy} & \boldsymbol{\Gamma}_{Bzx} & \mathbf{B}_{BBy} & \mathbf{0} & \boldsymbol{\Gamma}_{Bx} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{WW} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{WWyd} & \mathbf{0} & \boldsymbol{\Gamma}_{Wx} \\ \boldsymbol{\Lambda}_{Bzy} & \mathbf{0} & \mathbf{B}_{BzyBzy} & \mathbf{K}_{Bzx} & \mathbf{B}_{BzyBy} & \mathbf{0} & \mathbf{B}_{BzyBzx} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{BzxBy} & \mathbf{0} & \mathbf{B}_{BzxBx} & \mathbf{0} \\ \boldsymbol{\Lambda}_{By} & \mathbf{0} & \mathbf{B}_{ByBzy} & \mathbf{0} & \mathbf{B}_{ByBy} & \mathbf{0} & \mathbf{K}_{Bx} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{Wyd} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{WydWyd} & \mathbf{0} & \mathbf{K}_{Wx} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad (208)$$

$$\boldsymbol{\Psi}_2 = \begin{pmatrix} \boldsymbol{\Psi}_{BB} & & & & & & & symm. \\ \mathbf{0} & \boldsymbol{\Psi}_{WW} & & & & & & \\ \boldsymbol{\Psi}_{BzyB} & \mathbf{0} & \boldsymbol{\Theta}_{Bzy} & & & & & \\ \boldsymbol{\Psi}_{BzxB} & \mathbf{0} & \boldsymbol{\Psi}_{BzxBzy} & \boldsymbol{\Sigma}_{Bzx} & & & & \\ \boldsymbol{\Psi}_{ByB} & \mathbf{0} & \boldsymbol{\Psi}_{ByBzy} & \boldsymbol{\Psi}_{ByBzx} & \boldsymbol{\Theta}_{By} & & & \\ \mathbf{0} & \boldsymbol{\Psi}_{WydW} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_{Wyd} & & \\ \boldsymbol{\Psi}_{BxB} & \mathbf{0} & \boldsymbol{\Psi}_{BxBzy} & \boldsymbol{\Sigma}_{BxBzx} & \boldsymbol{\Psi}_{BxBy} & \mathbf{0} & \boldsymbol{\Sigma}_{Bx} & \\ \mathbf{0} & \boldsymbol{\Psi}_{WWx} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{WxWyd} & \mathbf{0} & \boldsymbol{\Sigma}_{Wx} \end{pmatrix}. (209)$$

$$(210)$$

For level 1 terms, the parameter matrices are arranged as follows.

$$\boldsymbol{\nu}'_1 = \left(\ \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}'\ \right), \tag{211}$$

$$\boldsymbol{\Lambda}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{Wynd} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{212}$$

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} \mathbf{I} & & & symm. \\ \mathbf{0} & \mathbf{I} & & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_{Wynd} & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{213}$$

$$\boldsymbol{\alpha}'_1 = (\mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}', \mathbf{0}'), \tag{214}$$

$$\mathbf{B}_1 = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_{WW} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{WWyd} & \mathbf{0} & \boldsymbol{\Gamma}_{Wx} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{Wyd} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{B}_{WydWyd} & \mathbf{0} & \mathbf{K}_W \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{215}$$

$$\boldsymbol{\Psi}_1 = \begin{pmatrix} \mathbf{0} & & & & & & & symm. \\ \mathbf{0} & \boldsymbol{\Psi}_{WW} & & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & & & \\ \mathbf{0} & \boldsymbol{\Psi}_{WydW} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Theta}_{Wyd} & & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \\ \mathbf{0} & \boldsymbol{\Psi}_{WWx} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{\Psi}_{WxWyd} & \mathbf{0} & \boldsymbol{\Sigma}_{Wx} \end{pmatrix}. \tag{216}$$

$$\tag{217}$$

For two-level analysis the default quasi-Newton convergence criterion is slightly stricter than for single-level analysis, 0.000001.

APPENDIX 10

APPENDIX 11
ESTIMATION OF FACTOR SCORES

This appendix describes factor score estimation with continuous y variables and with binary y variables.

Consider again the modeling framework of Appendix 2,

$$\boldsymbol{y^*}_i = \boldsymbol{\nu} + \boldsymbol{\Lambda}\,\boldsymbol{\eta}_i + \mathbf{K}\,\mathbf{x}_i + \boldsymbol{\epsilon}_i, \tag{218}$$

$$\boldsymbol{\eta}_i = \boldsymbol{\alpha} + \mathbf{B}\,\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\,\mathbf{x}_i + \boldsymbol{\zeta}_i. \tag{219}$$

Let $V(\boldsymbol{\epsilon}) = \boldsymbol{\Theta}$, $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$ as before.

For each individual i,

$$E(\boldsymbol{\eta}_i|\mathbf{x}_i) = \boldsymbol{\mu}_i = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha} + (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\,\mathbf{x}_i, \tag{220}$$

$$V(\boldsymbol{\eta}_i|\mathbf{x}_i) = \boldsymbol{\Sigma} = (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}. \tag{221}$$

Consider Bayesian estimation of $\boldsymbol{\eta}_i$ given $\mathbf{y}_i$ and $\mathbf{x}_i$. The posterior distribution of $\boldsymbol{\eta}_i$ is

$$g(\boldsymbol{\eta}_i|\mathbf{y}_i, \mathbf{x}_i) \propto \phi(\boldsymbol{\eta}_i|\mathbf{x}_i) f(\mathbf{y}_i|\boldsymbol{\eta}_i, \mathbf{x}_i), \tag{222}$$

where the prior $\phi(\boldsymbol{\eta}_i|\mathbf{x}_i)$ is multivariate normal with mean vector $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}$. Mplus considers three cases: all variables in $\mathbf{y}$ continuous; at least one variable in $\mathbf{y}$ binary or ordered categorical; and mixture models.

In the case of all y variables being continuous, $f(\mathbf{y}_i|\boldsymbol{\eta}_i, \mathbf{x}_i)$ is obtained from (218) and (219) using the computational implementation of (59), (60), where

$$E(\boldsymbol{\eta}_v) = \boldsymbol{\mu}_v = (\mathbf{I} - \mathbf{B}_v)^{-1}\boldsymbol{\alpha}_v, \tag{223}$$

$$V(\boldsymbol{\eta}_v) = \boldsymbol{\Sigma}_v = (\mathbf{I} - \mathbf{B}_v)^{-1}\boldsymbol{\Psi}_v(\mathbf{I} - \mathbf{B}_v)'^{-1}. \tag{224}$$

In this case, maximizing the log of the posterior with respect to $\boldsymbol{\eta}_i$ gives the usual factor score estimates from the regression method with correlated factors (see, e.g. Lawley & Maxwell, 1971)

$$\hat{\boldsymbol{\eta}}_i = \boldsymbol{\mu}_v + \mathbf{C}\,(\mathbf{v}_i - \boldsymbol{\nu}_v - \boldsymbol{\Lambda}_v\,\boldsymbol{\mu}_v), \tag{225}$$

where the factor score coefficient matrix $\mathbf{C}$ is expressed as

$$\mathbf{C} = \boldsymbol{\Sigma}_v\,\boldsymbol{\Lambda}'_v\,(\boldsymbol{\Lambda}_v\,\boldsymbol{\Sigma}_v\,\boldsymbol{\Lambda}'_v + \boldsymbol{\Theta}_v)^{-1}. \tag{226}$$

The quality of the factor score estimates can be expressed in terms of factor determinacy defined as the correlation between $\hat{\eta}_j$ and η_j.

APPENDIX 11

In the case where at least one of the ys is binary or ordered categorical, conditional independence is assumed

$$f(\mathbf{y}_i|\boldsymbol{\eta}_i, \mathbf{x}_i) = \prod_{j=1}^{p} f_j(y_{ij}|\boldsymbol{\eta}_i, \mathbf{x}_i), \qquad (227)$$

so that $\boldsymbol{\Theta}$ is assumed to be diagonal. For categorical outcomes, $\boldsymbol{\Theta}$ is further restricted as

$$\boldsymbol{\Theta} = \boldsymbol{\Delta}^{-2} - diag\,[\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B})'^{-1}\,\boldsymbol{\Lambda}']. \qquad (228)$$

The restriction on the residual variances of $\boldsymbol{\Theta}$ is a natural consequence of the latent response variables of $\boldsymbol{y}^*$ being measured by categorical y variables (see Appendix 1), while the assumption of zero residual covariances is made to simplify the factor score estimation. Consider a categorical y_j variable with categories $s = 0, 1, 2, \ldots, S_j - 1$ and $\tau_{j,k,0} = -\infty$, $\tau_{j,k,S_j} = \infty$ and define the probability of y_j being observed in category s,

$$f_j(y_{ij}|\boldsymbol{\eta}_i, \mathbf{x}_i) = \Phi[(\tau_{j,s+1} - \boldsymbol{\lambda}'_j\,\boldsymbol{\eta}_i - \ '_j\,\mathbf{x}_i)\,\theta_{jj}^{-1/2}] - \Phi[(\tau_{j,s} - \boldsymbol{\lambda}'_j\,\boldsymbol{\eta}_i - \ '_j\,\mathbf{x}_i)\,\theta_{jj}^{-1/2}], \quad (229)$$

while for a continuous y_j variable

$$f_j(y_{ij}|\boldsymbol{\eta}_i, \mathbf{x}_i) = e^{-1/2\,(y_{ij}-\nu_j-\boldsymbol{\lambda}'_j\,\boldsymbol{\eta}_i - \ '_j\,\mathbf{x}_i)^2\,\theta_{jj}^{-1}}, \qquad (230)$$

where $\boldsymbol{\lambda}'_j$ is the jth row of $\boldsymbol{\Lambda}$, $\ '_j$ is the jth row of $\mathbf{K}$, and θ_{jj} is the jth diagonal element of $\boldsymbol{\Theta}$. The factor score estimate $\hat{\boldsymbol{\eta}}_i$ is obtained from the mode of the posterior (222) of $\boldsymbol{\eta}_i$ by minimizing the function F with respect to $\boldsymbol{\eta}_i$,

$$F = 1/2\,(\boldsymbol{\eta}_i - \boldsymbol{\mu}_i)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\eta}_i - \boldsymbol{\mu}_i) - \sum_{j=1}^{p} ln f_j(y_{ij}|\boldsymbol{\eta}_i, \mathbf{x}_i). \qquad (231)$$

The minimization needs to be carried out by iterative techniques. Mplus uses quasi-Newton techniques where only first-order derivatives of F are needed.

Factor scores for continuous latent variables in mixture models are computed by noting that

$$E(\boldsymbol{\eta}_i|\mathbf{x}_i, \mathbf{y}_i) = E_{c|x,y}\,[E(\boldsymbol{\eta}_i|\mathbf{x}_i, \mathbf{y}_i, \mathbf{c}_i)], \qquad (232)$$

where $E(\boldsymbol{\eta}_i|\mathbf{x}_i, \mathbf{y}_i, \mathbf{c}_i)$ is estimated as discussed above for continuous outcomes using class-specific parameter values. Here, $E_{c|x,y}$ denotes expectation over the conditional distribution of $\mathbf{c}$ given $\mathbf{x}$ and $\mathbf{y}$, which reduces to a summation over classes using the posterior probabilities. Factor scores are printed for both the factor scores weighted across classes as given in (232) and the class-specific factor scores $E(\boldsymbol{\eta}_i|\mathbf{x}_i, \mathbf{y}_i, \mathbf{c}_i)$ for the class with the highest posterior class probability.

48

APPENDIX 12
MONTE CARLO SIMULATIONS

Monte Carlo simulations are useful to study properties of estimators. In these simulations, data are generated by random number draws from a prespecified model. The random draws and the analysis are repeated over several replications and the distribution of the estimates for these replications is summarized. Except for mixture models, the starting point for the generation of data is the mean vector

$$E \begin{pmatrix} \mathbf{y}^* \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_y^* \\ \mathbf{0} \end{pmatrix}$$

and covariance matrix

$$V \begin{pmatrix} \boldsymbol{\Sigma}_{y^*y^*} & symm. \\ \boldsymbol{\Sigma}_{xy^*} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}.$$

Random draws of the vector $(\mathbf{y}^{*\prime}, \mathbf{x}')'$ are made from a multivariate normal distribution with this mean vector and covariance matrix. The variables of $\mathbf{y}^*$ may then be transformed into categorical y variables. The multivariate normal data are created by generating uniformly distributed random numbers by the $ran2$ procedure in Press, Teukolsky, Vetterling and Flannery (1992, p. 272), normal random numbers as in Box and Mueller (1958), and multivariate normal variables by Cholesky decomposition as in Kennedy and Gentle (1980, pp. 294-301). Summaries across the replications include the parameter coverage defined as the proportion of the replications that a population parameter value is covered by the 95% interval based on the parameter estimate and its estimated standard error.

With mixture models, the distribution for the x variables is generated in a first step using the above multivariate normal approach, and allowing categorization of x's. Based on these values, the categorical outcomes for $\mathbf{c}$ and $\mathbf{u}$ are generated using uniformly distributed random numbers by a procedure equivalent to $ran2$, and the continuous outcomes $\mathbf{y}$ are generated using the multivariate normal procedure above with a mean vector that varies across $\mathbf{x}_i$ values. Missing data can be generated for $\mathbf{u}$ and $\mathbf{y}$, using MCAR or missingness predicted by $\mathbf{x}$ and $\mathbf{c}$.

Summaries across R replications include the mean and standard deviation (using $R-1$) of the parameter estimates over the replications, the mean across replications of the estimated standard errors, the mean square error,

$$\sum_{r=1}^{R} (\hat{\pi} - \bar{\hat{\pi}})^2/R + (\bar{\hat{\pi}} - \pi)^2, \tag{233}$$

the 95% coverage, i.e. proportion of replications for which the 95% confidence interval includes the true parameter value, and the 95% reject proportion, i.e. the proportion of

APPENDIX 12

replications for which the absolute value of the estimate to standard error ratio exceeds 1.96.

REFERENCES

Agresti, A. (1990). Categorical data analysis. New York: John Wiley & Sons.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317-332.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature, 19,* 1483-1536.

Amemiya, T. (1985). *Advanced Econometrics.* Cambridge, Mass.: Harvard University Press.

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

Bollen, K.A. 1989. *Structural Equations with Latent Variables.* New York: John Wiley.

Browne, M.W. (1982). Covariance structures. In D.M. Hawkins (ed.), *Topics in Applied Multivariate Analysis.* Cambridge University Press.

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 62-83

Browne, M.W. & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C.C. Clogg & M.E. Sobel (Eds.), *Handbook of Statistical Modeling for the Social and Behavioral Sciences* (pp. 311-359). New York: Plenum Press.

Browne, M.W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J.S. Long (eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park: Sage Publications.

Box, G.E.P & Mueller, M.E. (1958). A note on generation of normal deviates. *Annals of Mathematical Statistics, 28,* 610-611.

Everitt, B.S. & Hand, D.J. (1981). *Finite Mixture Distributions.* London: Chapman & Hall.

Hosmer, D. W. (1973). A comparison of iterative maximum likelihood estimates of the parameters of a mixture of two normal distributions under three different types of sample. *Biometrics, 29*, 761-770.

Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression.* New York: John Wiley & Sons.

Hu, L.T. & Bentler, P.M. (1999). Cutoff criteria for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1 - 55.

Jo, B. (2000). Estimation of intervention effects with noncompliance: alternative model

REFERENCES

specifications. Under review.

Jo, B. & Muthén, B. (2000). Longitudinal studies with intervention and noncompliance: Estimation of causal effects in growth mixture modeling. To appear in N. Duan & S. Reise (Eds.), *Multilevel Modeling: Methodological Advances, Issues, and Applications*, Multivariate Applications Book Series, Lawrence Erlbaum Associates.

Jo, B. & Muthén, B. (2001). Modeling of intervention effects with noncompliance: A latent variable approach for randomized trials. In G. A. Marcoulides & R.E. Schumacker (eds.), *Advanced Structural Equation Modeling: New Developments and Techniques*, Lawrence Erlbaum Associates, pp. 57-87.

Jöreskog, K.G. (1973). A general method for estimating as linear structural equation system. In *Structural Equation Models in the Social Sciences*, A.S. Goldberger and O.D. Duncan (Eds.). New York: Seminar Press, pp. 85-112.

Jöreskog, K.G., & Sörbom, D. (1979). *Advances in Factor Analysis and Structural Equation Models.* Cambridge, MA: Abt Books.

Kennedy, W.J. & Gentle, J.E. (1980). *Statistical Computing.* New York: Marcel Dekker, Inc.

Lawley, D.N. & Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method.* London: Butterworths.

Little, R.J.A., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley & Sons.

Little, R.J. & Yau, L.H.Y. (1998). Statistical techniques for analyzing data from prevention trials: Treatment of no-shows using Rubin's causal model. *Psychological Methods, 3,* 147-159.

McLachlan, G.J. & Basford, K.E. (1988). *Mixture Models. Inference and Applications to Clustering.* New York: Marcel Dekker.

McLachlan, G.J., & Krishnan, T. (1997). *The EM Algorithm and Extensions.* New York: John Wiley & Sons.

McLachlan, G.J. & Peel, D. (2000). *Finite Mixture Models.* New York: Wiley & Sons.

Mooijaart. A. (1985). A note on computational efficiency in asymptotically distribution-free correlational models. *British Journal of Mathematical and Statistical Psychology, 38,* 112-115.

Muthén, B. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association, 74,* 807-811.

Muthén, B. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics, 22,* 48-65.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, B. (1989a). Multiple-group structural modeling with non-normal continuous variables. *British Journal of Mathematical and Statistical Psychology, 42,* 55-62.

Muthén, B. (1989b). Latent variable modeling in heterogeneous populations. Presidential address to the Psychometric Society, July, 1989. *Psychometrika, 54,* 557-585.

Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Paper presented at the Psychometric Society meeting in Princeton, NJ, June 1990. *UCLA Statistics Series 62.*

Muthén, B. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28*, Winter 1991, 338-354.

Muthén, B. (1994). Multilevel covariance structure analysis. In J. Hox, & I. Kreft (Eds.), Multilevel Modeling, a special issue of *Sociological Methods & Research, 22* 376-398.

Muthén, B. (1996). Growth modeling with binary responses. In A. V. Eye, & C. Clogg (Eds.), *Categorical Variables in Developmental Research: Methods of Analysis* (pp. 37-54). San Diego, CA: Academic Press.

Muthén, B. (1997). Latent variable modeling with longitudinal and multilevel data. In A. Raftery (Ed.), *Sociological Methodology 1997* (pp. 453-480). Boston: Blackwell Publishers.

Muthén, B. (2000). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. To appear in Collins, L.M. & Sayer, A. (Eds.), *New Methods for the Analysis of Change.* Washington, D.C.: APA.

Muthén, B. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R.E. Schumacker (eds.), *Advanced Structural Equation Modeling: New Developments and Techniques*, Lawrence Erlbaum Associates, pp. 1-33.

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S., Carlin, J., & Liao, J. (2000). General growth mixture modeling for randomized preventive interventions. Under review.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407-419.

Muthén, B. & Curran, P. (1997). General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371-402.

Muthén, B., du Toit, S.H.C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. Accepted for publication in *Psychometrika.*

Muthén, B. & Khoo, S.T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences*, Special issue: latent growth

REFERENCES

curve analysis, 10, 73-101.

Muthén, B., Khoo, S.T. & Gustafsson, J.E. (1998). Multilevel latent variable modeling in multiple populations. Under review.

Muthén, B. & Masyn, K. (2001). Mixture discrete-time survival analysis.

Muthén, B. & Muthén, L. (2000). Integrating person-centered and variable-centered analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism: Clinical and Experimental Research, 24*, 882-891.

Muthén, B., & Satorra, A. (1995a). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika, 60,* 489-503.

Muthén, B., & Satorra, A. (1995b). Complex sample data in structural equation modeling. In P.V. Marsden (Ed.), *Sociological Methodology* (pp. 267-316). Washington D.C.: The American Sociological Association.

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

Muthén, B., Shedden, K. & Spisic, D. (1999). General latent variable mixture modeling. Technical report.

Nagin, D.S. (1999). Analyzing developmental trajectories: a semi-parametric, group-based approach. *Psychological Methods, 4*, 139-157.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., & Flannery, B.P. (1992). *Numerical Recipes in FORTRAN. The Art of Scientific Computing.* Second edition. Cambridge, England: Cambridge University Press.

Ramaswamy, V., DeSarbo, W., Reibstein, D., & Robinson, W. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science, 12*, 103-124.

Roeder, K., Lynch, K.G., & Nagin, D.S. (1999). Modeling uncertainty in latent class membership: A case study in criminology. *Journal of the American Statistical Association, 94*, 766-776.

Saris, W.E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C.C. Clogg (Ed.), *Sociological Methodology 1987* (pp. 105-129). Washington D.C.: American Sociological Association.

Satorra, A. (1992). Asymptotic robust inferences in the analysis of mean and covariance structures. In P.V. Marsden (Ed.), *Sociological Methodology 1992* (pp. 249-78). Oxford, England: Blackwell Publishers.

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), *Innovations in Multivariate Statistical Analysis. A Festschrift for Heinz Neudecker* (pp.233-247).

London: Kluwer Academic Publishers.

Satorra, A. & Bentler, P.M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *1988 Proceedings of the Business and Economic Statistics Section of the American Statistical Association,* 308-313.

Satorra, A. & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye and C.C. Clogg (eds.), *Latent Variable Analysis: Applications to Developmental Research* (pp. 399-419). Newbury Park: SAGE Publications.

Satorra, A. & Bentler, P.M. (1999). A scaled difference chi-square test statistic for moment structure analysis. Technical report, University of California, Los Angeles.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464.

Sclove L.S. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333-343.

Searle, S., Casella, G., & McCullogh, C.E. (1992). *Variance Components.* New York: Wiley & Sons.

Singer, J.D. & Willett, J.B. (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18,* 155-195.

Steiger, J.H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling, 5(4),* 411-419.

Steiger, J.H. & Lind, J.M. (1980). Statistically based tests for the number of common factors. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Sörbom, D. (1982). Structural equation models with structured means. In K.G. Jöreskog & H. Wold (Eds.), *Systems under Indirect Observation: Causality, Structure, Prediction.* Amsterdam: North Holland.

Sörbom, D. (1989). Model modification. *Psychometrika, 54,* 371-384.

Titterington, D.M., Smith, A.F.M., & Makov, U.E. (1985). *Statistical Analysis of Finite Mixture Distributions.* Chichester, U.K.: John Wiley & Sons.

Yang, C.C. (1998). Finite mixture model selection with psychometric applications. Unpublished doctoral dissertation. University of California, Los Angeles.

Yu, C.-Y. & Muthén, B. (2001). Evaluation of model fit indices for latent variable models with categorical and continuous outcomes. Technical report in preparation.

REFERENCES

INDEX

INDEX