GENERALIZED MULTILEVEL STRUCTURAL EQUATION MODELING

SOPHIA RABE-HESKETH

UNIVERSITY OF CALIFORNIA, BERKELEY

ANDERS SKRONDAL

NORWEGIAN INSTITUTE OF PUBLIC HEALTH, OSLO

ANDREW PICKLES

THE UNIVERSITY OF MANCHESTER

A unifying framework for generalized multilevel structural equation modeling is introduced. The models in the framework, called generalized linear latent and mixed models (GLLAMM), combine features of generalized linear mixed models (GLMM) and structural equation models (SEM) and consist of a response model and a structural model for the latent variables. The response model generalizes GLMMs to incorporate factor structures in addition to random intercepts and coefficients. As in GLMMs, the data can have an arbitrary number of levels and can be highly unbalanced with different numbers of lower-level units in the higher-level units and missing data. A wide range of response processes can be modeled including ordered and unordered categorical responses, counts, and responses of mixed types. The structural model is similar to the structural part of a SEM except that it may include latent and observed variables varying at different levels. For example, unit-level latent variables (factors or random coefficients) can be regressed on cluster-level latent variables. Special cases of this framework are explored and data from the British Social Attitudes Survey are used for illustration. Maximum likelihood estimation and empirical Bayes latent score prediction within the GLLAMM framework can be performed using adaptive quadrature in gllamm, a freely available program running in Stata.

Key words: multilevel structural equation models, generalized linear mixed models, latent variables, random effects, hierarchical models, item response theory, factor models, adaptive quadrature, empirical Bayes, GLLAMM.

Introduction

Among the milestones in the development of statistical modeling are undoubtedly the advent of comprehensive methodologies for structural equation modeling (e.g., Jöreskog, 1973) and multilevel (regression) modeling (e.g., Goldstein, 1986) and the concomitant implementation in widely available software such as LISREL (Jöreskog & Sörbom, 1989) and MLwiN (Rasbash, Browne, Goldstein, Yang, Plewis, Healy, et al., 2000). Although developed separately and for different purposes, the modeling approaches have striking similarities. Both include latent variables in the models in order to induce, and therefore explain, correlations among the responses.

Multilevel regression models are used when the data structure is hierarchical with elementary units at level 1 nested in clusters at level 2, which in turn may be nested in (super)clusters at level 3, and so on. The latent variables, or *random effects*, can be interpreted as unobserved heterogeneity at the different levels inducing dependence among all lower-level units in the same higher-level unit. Whereas random intercepts represent heterogeneity between clusters in the

overall response, random coefficients represent heterogeneity in the relationship between the response and explanatory variables.

Structural equation models are used when the variables of interest cannot be measured perfectly. Instead, there are either sets of items reflecting a hypothetical construct (e.g., depression) or fallible measurements of a variable (e.g., calorie intake) using different instruments. The latent variables, or *factors*, can then be interpreted as the constructs, traits, or "true" variables, underlying the measured items and inducing dependence among them. The measurement model is sometimes of interest in its own right, but relations among the factors or between factors and observed variables (the structural part of the model) often define the substantive model of interest. Although structural equation models with latent variables were mainly developed in psychometrics, it is worth noting that the same ideas are often used in biostatistics, where latent variables are used in regression models to take account of measurement error in the covariates (e.g., Plummer & Clayton, 1993; Carroll, Ruppert, & Stefanski, 1995). Latent variables are also used in biometrical genetics to represent shared genetic and environmental influences (e.g., Neale & Cardon, 1992).

A synthesis of both methods, namely *multilevel structural equation modeling*, is required when the units of observation form a hierarchy of nested clusters and some variables of interest cannot be measured directly but are measured by a set of items or fallible instruments. Before contrasting different approaches it is useful to define three types of balance that may be required: (1) *complete multivariate responses* where there are no missing items for any unit (accidental or by design), (2) *balanced multilevel designs* where every unit at level l has the same number of units at level l-1, and (3) *balanced covariates*, which take on the same sets of values for each higher-level unit in a balanced multilevel design. An example of a balanced covariate is time in a longitudinal study if each unit contributes responses at the same sets of time points.

Multilevel structural equation models could be specified using either multilevel regression models or structural equation models as the starting point. An advantage of using multilevel regression models is that none of the three types of balance will be required. This is because in multilevel regression modeling, multivariate responses are typically handled by treating the different variables comprising the multivariate response as level-1 units and the original units as level-2 clusters (e.g., Chapter 4 of Goldstein, 1995; Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003). Incomplete multivariate responses then merely result in varying cluster sizes. Such unbalanced multilevel designs are typically handled by the estimation procedure, producing consistent estimators if responses are missing at random (MAR). Since level-1 covariates (e.g., time varying covariates in growth curve modeling) can be represented by a single covariate vector, unbalanced covariates pose no extra difficulties. Unfortunately, previous attempts to incorporate measurement models within the multilevel regression framework have rested on the unrealistic assumption that the measurement error variances are known (Goldstein, 1995) or that the factor loadings are known (Raudenbush & Sampson, 1999a,b; Raudenbush & Bryk, 2002).

When structural equation modeling is instead taken as starting point, we note that some limited multilevel structural equation modeling is possible using the traditional approaches where models are fitted to sample covariance matrices and sometimes means. This is achieved by treating the highest level of the multilevel model as "level 1" and the combinations of all lower-level units as a high-dimensional multivariate response (e.g., Muthén, 1997). For example, 4 variables observed at 3 time points on 2 twins from each of a large number of families could be modeled by letting the families be the "level 1" units and the responses for each family a 24-dimensional multivariate response. Obviously, this method can only be applied if there are not too many units at any of the levels apart from the highest, since the dimensionality of the response vector would otherwise become excessive. In addition, the method requires complete multivariate responses and balanced multilevel designs. Random coefficients can be included via factor models, but this is only possible if the corresponding covariates are balanced, thus requiring all three types of balance mentioned above.

Using the traditional approach to structural equation modeling, all three types of imbalance can be handled to some degree by multi-group analysis. However, a more flexible approach is to base parameter estimation on likelihood contributions from individual units (e.g., Arminger & Sobel, 1990) instead of sample covariance matrices and means from listwise samples. However, clusters with large numbers of lower-level units still pose problems due to excessive dimensionality. For example, if the families in the above example were also nested in counties with up to 100 families per county, the multivariate response would be of dimension 2400, with many missing values for those counties contributing fewer than 100 families.

Explicit two-level structural equation methodology has been developed to overcome the problem of large clusters. In the above example, county would be declared a "level 2" unit and family a "level 1" unit. For continuous responses or responses modeled by latent underlying variables, such multilevel structural equation models are typically defined by specifying separate models for the within-cluster and between-cluster covariance matrices (e.g., Longford & Muthén, 1992; Poon & Lee, 1992; Linda, Lee, & Poon, 1993; Muthén, 1994; Lee & Shi, 2001) although Goldstein and McDonald (Goldstein & McDonald, 1988; McDonald & Goldstein, 1989) propose a more general framework. There are six limitations to models specified via two separate conventional structural equation models. First, random coefficients cannot be included at "level 2." Second, random coefficients at the lower levels are not permitted if the corresponding covariates are continuous or highly unbalanced. Third, the models cannot include large clusters at more than one level ("level 2") since this would render the dimensionality excessive. Fourth, this framework does not allow direct inclusion of regressions of "level 1" latent variables on "level 2" latent variables. Such models must be specified in a roundabout way by using a large number of parameters and imposing a large number of nonlinear parameter constraints. We will return to this point later. Fifth, modeling the "level 1" covariance matrix presupposes that the responses are either continuous or can be viewed as generated by categorizing continuous underlying variables. This accommodates models for dichotomous and ordinal responses, but not, for instance, Poisson and gamma models often used for counts and durations in continuous time. Finally, the typical specification of multivariate normality for underlying variables corresponds to a probit link and therefore rules out several useful links such as the log, logistic, and complementary log-log.

In this paper we introduce a unifying framework for multilevel structural equation models. Our generalized linear latent and mixed modeling (GLLAMM) framework represents a generalization of multilevel regression models or generalized linear mixed models and does not require balanced data in any of the three senses described above (see also Rabe-Hesketh & Pickles, 1999; Rabe-Hesketh, Pickles, & Skrondal, 2001a; Skrondal & Rabe-Hesketh, 2004). Maximum likelihood estimation is implemented in the software gllamm (Rabe-Hesketh, Pickles, & Taylor, 2000; Rabe-Hesketh, Pickles, & Skrondal, 2001b; Rabe-Hesketh, Skrondal, & Pickles, 2002) running in the widely available statistical package Stata (StataCorp, 2003). Unlike previous methodologies, our framework allows

- an arbitrary number L of levels,
- automatic handling of data missing at random (MAR) and scope for modeling data not missing at random (NMAR),
- unbalanced multilevel designs,
- random coefficients of unbalanced covariates,
- general factor structures (with "free" factor loadings),
- regressions among latent variables (factors and/or random coefficients) varying at different levels,
- a wide range of response processes including ordered and unordered categorical responses, counts, and responses of mixed types.

Unlike previous contributions that have focused mostly on estimation issues, a major additional focus is to explore the different types of multilevel structural equation model that are likely to be useful in different types of application.

The outline of the paper is as follows. In the next section, we describe our modeling framework. We then explore multilevel structural equation models as special cases of the framework. Finally, we illustrate some of the ideas using an example.

A General Framework for Multilevel Structural Equation Modeling

We depart from the traditional multivariate framework for formulating factor and structural equation models and adopt an essentially univariate approach. All response variables for each unit are stacked in a single response vector with different variables distinguished one from another by a design matrix. In the next three subsections we outline the three parts of our framework: (1) the response model, (2) the structural model for the latent variables, and (3) the distribution of the latent variables. In the final subsection, we derive the marginal log-likelihood and describe our method of estimation and latent variable prediction.

Response Model

Conditional on the latent variables, the response model is a generalized linear model specified via a linear predictor, a link, and a distribution from the exponential family.

Linear Predictor

To simplify notation, we will not use subscripts for the units of observation at the various levels in this section. For a model with L levels and M_l latent variables at level $l > 1$, the linear predictor has the form

$$\nu = \boldsymbol{\beta}'\mathbf{x} + \sum_{l=2}^{L}\sum_{m=1}^{M_l}\eta_m^{(l)}\boldsymbol{\lambda}_m^{(l)'}\mathbf{z}_m^{(l)} \tag{1}$$

with the first element of $\boldsymbol{\lambda}_m^{(l)}$ typically set to 1, that is, $\lambda_{m1}^{(l)} = 1$. The elements of $\mathbf{x}$ are explanatory variables associated with the fixed effects or regression coefficients, $\boldsymbol{\beta}$. The mth latent variable at level l, $\eta_m^{(l)}$, is multiplied by a linear combination $\boldsymbol{\lambda}_m^{(l)'}\mathbf{z}_m^{(l)}$ of explanatory variables $\mathbf{z}_m^{(l)}$, where $\boldsymbol{\lambda}_m^{(l)}$ are parameters (usually factor loadings).

Links and Distributions

We will use $\boldsymbol{\eta}^{(l)} = (\eta_1^{(l)}, \ldots, \eta_{M_l}^{(l)})'$ to denote the vector of latent variables at level l for a given level-l unit with corresponding vector of explanatory variables $\mathbf{z}^{(l)} = (\mathbf{z}_1^{(l)'}, \ldots, \mathbf{z}_{M_l}^{(l)'})'$. Also define the vector of latent variables at all levels for a level-2 unit $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)'}, \ldots, \boldsymbol{\eta}^{(L)'})'$ with corresponding vector of explanatory variables $\mathbf{z} = (\mathbf{z}^{(2)'}, \ldots, \mathbf{z}^{(L)'})'$. The conditional expectation of the response y given $\mathbf{x}$, $\mathbf{z}$, and $\boldsymbol{\eta}$ is "linked" to the linear predictor ν via a link function $g(\cdot)$

$$g(\mathrm{E}[y|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}]) = \nu. \tag{2}$$

The specification is completed by choosing a distributional "family" for the conditional distribution of the response variable given the latent and explanatory variables. There are no latent variables at level 1, this level being reserved for the "error" of the chosen conditional distribution.

The following kinds of response variable are among those accommodated by our framework:

1. *Continuous responses*: An identity link and a normal distribution are usually assumed:

$$y = \nu + \epsilon,$$

with $f(\epsilon) = (2\pi\sigma^2)^{-1/2}\exp(-\epsilon^2/(2\sigma^2))$. The conditional density becomes

$$f(y|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \sigma^{-1}\phi(\nu\sigma^{-1}), \qquad (3)$$

where ϕ denotes the standard normal density. Heteroscedasticity can be modeled as

$$\log(\sigma) = \boldsymbol{\alpha}'\mathbf{z}^{(1)}, \qquad (4)$$

where σ is the standard deviation of ϵ, $\boldsymbol{\alpha}$ are parameters, and $\mathbf{z}^{(1)}$ covariates.

2. *Ordinal and dichotomous responses and discrete time durations*: To allow the generalized linear model framework to be applied to ordinal responses, Equation (2) is usually modified to a cumulative form

$$g(\Pr(y \leq y_s|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta})) = \kappa_s - \nu, \quad s = 1, \ldots, S - 1,$$

where $y_s, s = 1, \ldots, S$ are the response categories and $\kappa_s, s = 2, \ldots, S - 1$ with $0 = \kappa_1 < \cdots < \kappa_{S-1}$ are *thresholds* to be estimated. A multinomial distribution is invariably assumed together with a logit, probit, or complementary log-log link.

Alternatively, the models can be defined using the concept of an underlying continuous response y^*,

$$y^* = \nu + \epsilon^*,$$

with ordinal and dichotomous responses generated via a threshold model:

$$y = y_s \quad \text{if} \quad \kappa_{s-1} < y^* \leq \kappa_s, \quad \kappa_0 = -\infty, \kappa_1 = 0, \kappa_S = \infty.$$

The logit, probit, and complementary log-log links correspond to specifying $f(\epsilon) = \exp(-\epsilon)[1+\exp(-\epsilon)]^{-2}$, $f(\epsilon) = (2\pi)^{-1/2}\exp(-\frac{1}{2}\epsilon^2)$, and $f(\epsilon) = \exp(\epsilon - \exp(\epsilon))$, respectively. If the variance of ϵ is identifiable, scaled versions of these densities, corresponding to scaled links, can be used. The conditional distribution becomes

$$f(y = y_s|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = F(\sigma^{-1}[\kappa_s - \nu]) - F(\sigma^{-1}[\kappa_{s-1} - \nu]), \qquad (5)$$

where F is a cumulative distribution function and σ is a scale parameter that can be modeled as in (4). The parallel regression assumption can be relaxed by allowing the thresholds to depend on covariates.

Grouped or interval censored continuous responses can be modeled in the same way by constraining the threshold parameters to equal the limits of the censoring intervals. By allowing unit-specific right-censoring, this model can be used for discrete time durations. Another approach to discrete time durations is to model the probabilities of "surviving" each time interval using models for dichotomous responses. The logit link then corresponds to a continuation ratio model and the complementary log-log link to a proportional hazards model in continuous time; see Rabe-Hesketh, Yang, and Pickles (2001) for a review. Other models for ordinal responses, such as the adjacent category logit model, can be specified as special cases of the multinomial logit model for polytomous responses in (7). If identified, overdispersion can be modeled by including a random intercept at level 1.

3. *Counts and durations in continuous time*: The response model for counts (nonnegative integers) is typically specified via the log link and the Poisson distribution corresponding to

$$\ln E[y|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}] = \nu,$$

and

$$f(y = s|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \frac{[\exp(\nu)]^s}{s!}\exp(-\exp(\nu)). \tag{6}$$

The Poisson distribution can also be used for modeling durations in continuous time. If a piecewise exponential model is assumed, with constant hazards for intervals of time, each observed duration contributes a product of terms of the form of (6) to the likelihood, namely, one term for each interval it exceeds (e.g., Clayton, 1988). If the hazard changes at each event time, this is equivalent to Cox regression. Overdispersion can be handled by random intercepts at level 1. Unlike models for dichotomous and ordinal responses, models for counts cannot be written as latent response models.

4. *Polytomous responses, rankings and pairwise comparisons*: For such "comparative" responses we consider a latent response model of the form

$$y_s^* = \nu_s + \epsilon_s,$$

where $s = 1, 2, \ldots, S$ denotes unordered categories. For a polytomous response,

$$y = y_s \quad \text{if} \quad y_s^* > y_\ell^*, \quad \forall \ell, \; \ell \neq s.$$

If ϵ_s is specified as Gumbel (extreme value of Type I), $f(\epsilon_s) = \exp(-\epsilon_s - \exp(-\epsilon_s))$, the conditional probability becomes

$$f(y = y_s|\mathbf{x}, \mathbf{z}, \boldsymbol{\eta}) = \exp(\nu_s)\left[\sum_{\ell=1}^{S}\exp(\nu_\ell)\right]^{-1}, \tag{7}$$

a multinomial logit. This distribution can also serve as a building block for conditional distributions of rankings (Skrondal & Rabe-Hesketh, 2003a). In pairwise comparison data, the dichotomous preference indicators for pairs of alternatives can be modeled using probit or logit regression (e.g., Takane, 1987; Böckenholt, 2001).

5. *Mixed responses*: Different links and distributions can be specified for different responses. This allows modeling of left- or right-censored continuous responses by specifying an identity link and normal distribution for uncensored responses and a scaled probit link (with scale equal to the residual standard deviation of the uncensored responses) and binomial distribution otherwise. Mixed responses are common in selection models (e.g., Heckman, 1979) where selection is typically dichotomous but the response of interest is often not. An extension to multilevel selection models is treated in Skrondal, Rabe-Hesketh, and Pickles (2002). Other examples are covariate measurement error problems, for example, logistic regression with measurement errors in a continuous covariate (Rabe-Hesketh, Pickles, & Skrondal, 2003; Rabe-Hesketh, Skrondal, & Pickles, 2003; Skrondal & Rabe-Hesketh, 2003b). In structural equation models with several latent variables, the measurement models for different latent variables may require different links and/or families. Finally, composite links can be useful for specifying proportional odds models for right-censored responses, for handling missing categorical covariates and many other model types; see Skrondal and Rabe-Hesketh (2004b).

Skrondal and Rabe-Hesketh (2004a) discuss latent variable models with all these response types and present applications from various disciplines.

Structural Model for the Latent Variables

The structural model for the latent variables has the form

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}, \tag{8}$$

where $\mathbf{B}$ is an $M \times M$ parameter matrix, $M = \sum_l M_l$, $\mathbf{w}$ is a vector of Q covariates, $\boldsymbol{\Gamma}$ is an $M \times Q$ parameter matrix, and $\boldsymbol{\zeta}$ is a vector of M errors or disturbances. Note that (8) resembles single-level structural equation models (e.g., Muthén, 1984). The important difference is that the latent variables may vary at different levels in our framework. Each element of $\boldsymbol{\zeta}$ varies at the same level as the corresponding element of $\boldsymbol{\eta}$.

We do not allow latent variables to be regressed on latent or observed variables varying at a lower level, since such specifications do not appear to make sense. Furthermore, our current implementation is confined to recursive models, not permitting feedback effects among the latent variables. The two restrictions together imply that the matrix $\mathbf{B}$ is strictly upper diagonal if the elements of $\boldsymbol{\eta}^{(l)}$ are permuted appropriately, since the elements of $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)\prime}, \ldots, \boldsymbol{\eta}^{(L)\prime})^\prime$ are arranged in increasing order of l.

Distribution of the Latent Variables

The structure of the latent variables is specified by the number of levels L and the number of latent variables M_l at each level. A particular level may coincide with a level of clustering in the hierarchical dataset. However, there will often not be a direct correspondence between the levels of the model and the levels of the data hierarchy. For instance, in factor models items will be treated as units at level 1 and subjects as units at level 2. For the models, "unit at a level" and "latent variable at a level" are defined as follows:

- a unit at level 1 is an elementary unit of observation,
- a unit k at level $l > 1$ is a cluster of level-1 units,
- the level-1 units in cluster k at level $l > 1$ fall into $n_k^{(l-1)}$ subsets representing units at level $l - 1$,
- a latent variable $\eta^{(l)}$ at level l varies between the units at level l but not within the units,
- the units at level l are conditionally independent given the latent variables at levels $l + 1$ and above and any explanatory variables.

The distribution of the disturbances $\boldsymbol{\zeta}$ needs to be specified if a structural model is used; otherwise we must specify the distribution of $\boldsymbol{\eta}$. Here the basic assumption is that latent variables at the same level may be dependent, whereas latent variables at different levels are independent. In this article it is further assumed that the latent variables at level l have a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}_l$, although other multivariate distributions could be used. Alternatively, the distribution could be left unspecified by using nonparametric maximum likelihood estimation (e.g., Laird, 1978; Heckman & Singer, 1984; Rabe-Hesketh, Skrondal, & Pickles, 2003).

The Marginal Log-likelihood

The likelihood of the observed data is the likelihood marginal to all latent variables. Let $\boldsymbol{\theta}$ be the vector of all parameters including the regression coefficients $\boldsymbol{\beta}$, the factor loadings $\boldsymbol{\lambda}_m^{(l)}$, $m = 1, \ldots, M_l, l = 2, \ldots, L$, the nonduplicated elements of the covariance matrices $\boldsymbol{\Sigma}_l$, the threshold parameters κ_s, $s = 2, \ldots, S - 1$ for ordinal responses, and the parameters $\boldsymbol{\alpha}$ for modeling level-1 heteroscedasticity. The number of free parameters in $\boldsymbol{\theta}$ will be reduced if constraints

PSYCHOMETRIKA

are imposed. Further define $\mathbf{y}_{(l)}$ to be the response vector and $\mathbf{X}_{(l)}$ to be the matrix of explanatory variables [with rows $(\mathbf{x}', \mathbf{z}', \mathbf{z}^{(1)'}, \mathbf{w}')$] for all level-1 units belonging to a particular unit at level l. Let $\mathbf{y}$ and $\mathbf{X}$ be the response vector and matrix of explanatory variables for all units. Substituting the structural model into the response model, giving the reduced form, defines the conditional distribution of the responses given the latent and explanatory variables. The corresponding conditional density (or probability) of a response of a level-1 unit will be denoted as $f^{(1)}(\mathbf{y}_{(1)}|\mathbf{X}_{(1)}, \boldsymbol{\zeta}^{(2+)}; \boldsymbol{\theta})$, where $\boldsymbol{\zeta}^{(l+)} = (\boldsymbol{\zeta}^{(l)'}, \ldots, \boldsymbol{\zeta}^{(L)'})'$. The form of this density is implied by Equations (3), (5), (6), or (7), depending on the response process.

We will denote the multivariate normal density of the latent variables at level l as $h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta})$. The marginal likelihood is constructed recursively. The conditional density of the responses of a level-l unit, conditional on the latent variables at levels $l + 1$ and above, is:

$$f^{(l)}(\mathbf{y}_{(l)}|\mathbf{X}_{(l)}, \boldsymbol{\zeta}^{([l+1]+)}; \boldsymbol{\theta}) = \int h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)}|\mathbf{X}_{(l-1)}, \boldsymbol{\zeta}^{(l+)}; \boldsymbol{\theta})\mathrm{d}\boldsymbol{\zeta}^{(l)}, \quad (9)$$

where the product is over all level l-1 units within the level l unit. The total marginal likelihood is the product of the contributions from all highest-level units,

$$\ell(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X}) = \prod f^{(L)}(\mathbf{y}_{(L)}|\mathbf{X}_{(L)}; \boldsymbol{\theta}). \quad (10)$$

The gllamm program maximizes the numerically integrated marginal log-likelihood using a Newton–Raphson algorithm. For given parameter values, the multivariate integral over the latent variables $\boldsymbol{\zeta}^{(l)}$ is evaluated by integrating over M_l independent standard normally distributed latent variables $\mathbf{v}^{(l)}$ with $\boldsymbol{\zeta}^{(l)} = \mathbf{C}_l\mathbf{v}^{(l)}$, where $\mathbf{C}_l$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_l$. Letting $\mathbf{v}^{(l+)} = (\mathbf{v}^{(l)'}, \ldots, \mathbf{v}^{(L)'})'$, the integral can be approximated by Cartesian product quadrature as

$$\int h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)}|\mathbf{X}_{(l-1)}, \boldsymbol{\zeta}^{(l+)}; \boldsymbol{\theta})\mathrm{d}\boldsymbol{\zeta}^{(l)}$$

$$= \int \phi(v_{M_l}^{(l)}) \cdots \int \phi(v_1^{(l)}) \prod f^{(l-1)}(\mathbf{y}_{(l-1)}|\mathbf{X}_{(l-1)}, \mathbf{v}^{(l)}, \mathbf{v}^{([l+1]+)}; \boldsymbol{\theta})\mathrm{d}v_1^{(l)} \cdots \mathrm{d}v_{M_l}^{(l)}$$

$$\approx \sum_{r_{Ml}} \pi_{r_{Ml}} \cdots \sum_{r_1} \pi_{r_1} \prod f^{(l-1)}(\mathbf{y}_{(l-1)}|\mathbf{X}_{(l-1)}, \alpha_{r_1}, \ldots, \alpha_{r_{Ml}}, \mathbf{v}^{([l+1]+)}; \boldsymbol{\theta}),$$

where $\phi(\cdot)$ is the standard normal density and π_r and α_r are quadrature weights and locations, respectively. Standard Gauss–Hermite quadrature rules can be used, but a superior approach is to use adaptive quadrature, which essentially places the locations where the integrand is concentrated. The method exploits the fact that the integrand is proportional to the posterior density of the latent variables, which is approximately multivariate normal for large cluster sizes (see Rabe-Hesketh, Skrondal, & Pickles, 2004). Apart from improving parameter estimation, adaptive quadrature also provides better empirical Bayes predictions of the latent variables, which are produced by numerical integration for all latent variables along with posterior standard deviations as one step of the algorithm. Multiple integrals can be evaluated more efficiently using spherical instead of Cartesian product quadrature. For details of our implementation of adaptive quadrature, see Rabe-Hesketh, Skrondal, and Pickles (2002, 2004).

An advantage of estimation using quadrature is that accuracy can be assessed by comparing solutions with different numbers of quadrature points. Since adaptive quadrature is derived by assuming approximate normality of the posterior density of the latent variables, it tends to work well for continuous responses, large counts, and large clusters, precisely where standard quadrature often works poorly (Lesaffre & Spiessens, 2001). However, the normality approximation may be crude for small clusters of dichotomous responses, a problem compounded when the predicted probabilities are close to 0 or 1, which is common for high intraclass correlations. We have

carried out an extensive simulation study for dichotomous responses with a large range of cluster sizes and intraclass correlations. The performance of adaptive quadrature was found to be good in all cases, with larger numbers of quadrature points required for more difficult situations (Rabe-Hesketh, Skrondal, & Pickles, 2004). Comparing gllamm with other software (using, e.g., IGLS, PQL, MCMC, and quadrature) for continuous, dichotomous, ordinal, and polytomous responses, counts and rankings, good agreement was found between parameter estimates, standard errors, and log-likelihood values (Rabe-Hesketh, Pickles, & Skrondal, 2001b; Rabe-Hesketh, Skrondal, & Pickles, 2002). The exceptions were cases where simulation or parametric bootstrapping demonstrated that PQL produced biased estimates in contrast to quadrature (Rabe-Hesketh, Pickles, & Skrondal, 2001b; Dohoo, Tillard, Stryhn, & Faye, 2001; Rabe-Hesketh, Toulopoulou, & Murray, 2001).

Special Cases of Multilevel Structural Equation Models

In the first subsection we develop multilevel factor models. Conventional factor and Multiple Indicator Multiple Cause (MIMIC) models are expressed as models within the framework in order to familiarize the reader with the notion of using dummy variables for specifying multivariate models for a univariate response vector. This provides the starting point for our multilevel extensions. In addition to standard multilevel factor models, we suggest more structured "variance components factor models," which represent useful building blocks for multilevel structural equation models discussed in the second subsection. Here, a structural model is specified for latent variables varying at different levels and measured by either same level or lower level items. In the final subsection we first show how conventional multilevel random coefficient models (for unbalanced covariates) can be specified using the response model in our framework. The model is then extended by including both latent responses and latent covariates. All models can be combined with any of the links and distributions described earlier. It should be noted that the special cases do not exhaust the model framework but are chosen to explore types of model structure that may prove useful in multilevel modeling.

Multilevel Factor Models

Factor models can be specified by using dummy variables to associate particular factor loadings with particular responses. To introduce notation, first consider a conventional single-level factor model with up to I items i observed on units j. Define dummy variables for the items

$$d_{ih} = \begin{cases} 1 & \text{if } i = h \\ 0 & \text{otherwise} \end{cases} , \quad i = 1, \ldots, I \tag{11}$$

placed in the dummy vector $\mathbf{d}_i = (d_{i1}, \ldots, d_{iI})'$ with the ith element equal to 1 and all other elements equal to 0. A multidimensional factor model can be defined as

$$\nu_{ij} = \boldsymbol{\beta}' \mathbf{d}_i + \sum_{m=1}^{M_2} \eta_{mj}^{(2)} \boldsymbol{\lambda}_m^{(2)'} \boldsymbol{\delta}_{mi}^{(2)}, \tag{12}$$

where $\boldsymbol{\lambda}_m^{(2)}$ is a vector of the nonzero factor loadings for the mth common factor $\eta_{mj}^{(2)}$ and $\boldsymbol{\delta}_{mi}^{(2)}$ is the subvector of $\mathbf{d}_i$ corresponding to the items loading on the factor. For example, if items 2 and 5 load on factor 1, $\boldsymbol{\delta}_{1i}^{(2)}$ contains the second and fifth elements of $\mathbf{d}_i$, $\boldsymbol{\delta}_{1i}^{(2)} = (d_{i2}, d_{i5})'$, and $\boldsymbol{\lambda}_1^{(2)} = (\lambda_{21}, \lambda_{51})$. The traditional way of writing the model in (12) for continuous, ordinal, or dichotomous data is (e.g., Muthén, 1984)

$$\mathbf{y}_j^* = \boldsymbol{\beta} + \boldsymbol{\Lambda} \boldsymbol{\eta}_j + \boldsymbol{\epsilon}_j,$$

where $\mathbf{y}_j^*$ is an I-dimensional vector of underlying continuous responses for unit j, $\boldsymbol{\Lambda}$ is the $I \times M_2$ matrix of factor loadings, $\boldsymbol{\eta}_j$ are common factors, and $\boldsymbol{\epsilon}_j$ unique factors.

In our framework there is no explicit term for the unique factors because the level-1 variability is implicit in the distribution family of the chosen generalized linear response model. If identified, the unique factor variances can be allowed to differ between items with $\text{var}(\epsilon_{ij}) = \sigma_i^2$ by introducing heteroscedasticity at level 1 using (4) with $\mathbf{z}^{(1)} = \mathbf{d}_i$. Conditional dependence between pairs of items can be induced by additional latent variables at the cost of increasing the dimensionality of integration. For categorical responses, multivariate conditional response probabilities can be specified directly using interaction parameters (e.g., Harper, 1972; Hagenaars, 1988).

Before considering how *unobserved* heterogeneity at the cluster level could be incorporated in factor models, we briefly consider three ways in which *observed* heterogeneity is traditionally incorporated: MIMIC models, MIMIC models with direct effects, and multi-group factor models.

A MIMIC model (e.g., Jöreskog & Goldberger, 1975) has the response model in (12) with structural model

$$\boldsymbol{\eta}_j^{(2)} = \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j^{(2)}.$$

In the MIMIC approach, item bias results if explanatory variables have direct effects on the items, in addition to their indirect effects via the common factors (Muthén, 1985). To accommodate the direct effect of an explanatory variable w_q on item h, the fixed term $\beta_{I+1}(d_{ih}w_q)$ can be included in the response model for v_{ij}.

Multiple group factor models (e.g., Jöreskog, 1971) can be specified using

$$v_{ijg} = \sum_{h=1}^{G} \boldsymbol{\beta}_h'(\mathbf{d}_i d_{gh}) + \sum_{h=1}^{G} \sum_{m=1}^{M_{h2}} \eta_{mhj}^{(2)} \boldsymbol{\lambda}_{mh}^{(2)\prime} (\boldsymbol{\delta}_{mhi}^{(2)} d_{gh}),$$

where g indexes group, $g = 1, \ldots, G$. Here, in contrast to MIMIC models, the number of common factors M_{g2} and their covariance matrices can also differ between groups.

In the multilevel setting, with units j nested in clusters k, one possible way of modeling unobserved heterogeneity at the cluster level is analogous to a MIMIC model. Instead of observed variables $\mathbf{w}$, unobserved cluster-level variables $\boldsymbol{\eta}_k^{(3)}$ affect the items via the factors $\boldsymbol{\eta}_{jk}^{(2)}$

$$v_{ijk} = \boldsymbol{\beta}'\mathbf{d}_i + \sum_{m=1}^{M_2} \eta_{mjk}^{(2)} \boldsymbol{\lambda}_m^{(2)\prime} \boldsymbol{\delta}_{mi}^{(2)}$$

$$\boldsymbol{\eta}_{jk}^{(2)} = \mathbf{B}^* \boldsymbol{\eta}_k^{(3)} + \boldsymbol{\zeta}_{jk}^{(2)}, \tag{13}$$

where $\mathbf{B}^*$ is an identity matrix, representing the part of the $\mathbf{B}$ matrix of Equation (8) that relates the unit-level common factors to the cluster-level common factors, so that

$$\eta_{mjk}^{(2)} = \eta_{mk}^{(3)} + \zeta_{mjk}^{(2)}. \tag{14}$$

We will call this type of multilevel factor model, where a common factor defined through factor loadings at the unit level varies at higher levels, a *variance components factor model*. In the unidimensional case, this model represents an obvious generalization of item response theory (IRT) models useful if, for example, children's mean latent abilities vary randomly between schools (see, e.g., Fox & Glas, 2001). Such a model is illustrated in path diagram form in Figure 1(a).

Following the conventions of path diagrams, circles represent latent variables and rectangles observed variables. The nested frames represent the nested levels; variables located within the
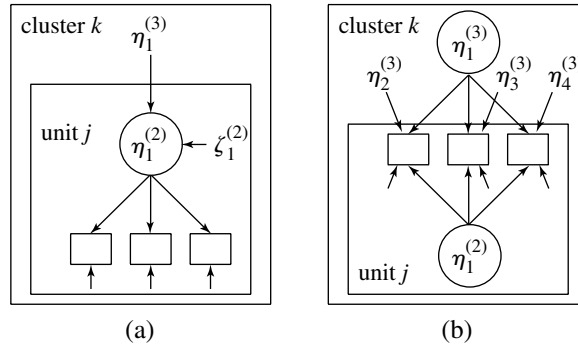
FIGURE 1.
(a) A variance components factor model and (b) a general two-level factor model.

outer frame labeled "cluster k" vary between clusters and have a k subscript and variables also inside the inner frame labeled "unit j" vary between units and have both the j and k subscripts. We have enclosed in circles only those latent variables that are interpretable as latent covariates or latent response variables. Latent variables serving as residuals, disturbances, or random effects (not enclosed in circles) are referred to as ζ only if they correspond to disturbances in the structural model. Arrows connecting latent and/or observed variables represent linear relations except for arrows to observed responses, which represent possibly nonlinear relations depending on the link functions. Level-1 variability is represented by a short unlabeled arrow pointing at the observed response. Unlike conventional measurement models for continuous responses where the arrow represents an additive error term, the arrow could represent, for instance, binomial or Poisson variability depending on the response distribution.

The analogue to item bias, where unobserved cluster-level variables have a direct effect on particular items h, can be modeled by including a cluster-level unique factor, $\eta^{(3)}_{M_3+1,k}d_{ih}$ for that item in the model for v_{ijk}. The unobserved heterogeneity analogue to multiple group analysis is latent class analysis (e.g., Clogg, 1995), which will not be discussed further in this article.

A general two-level factor model with unique factors at both unit and cluster levels can be defined as

$$v_{ijk} = \boldsymbol{\beta}'\mathbf{d}_i + \sum_{m=1}^{M_2} \eta^{(2)}_{mjk}\boldsymbol{\lambda}^{(2)\prime}_m\boldsymbol{\delta}^{(2)}_{mi} + \sum_{m=1}^{M_3} \eta^{(3)}_{mk}\boldsymbol{\lambda}^{(3)\prime}_m\boldsymbol{\delta}^{(3)}_{mi} + \sum_{m=1}^{I} \eta^{(3)}_{M_3+m,k}d_{im}. \tag{15}$$

The latent variables in the last sum are mutually independent unique factors at the cluster level and are independent of the cluster-level common factors in the second sum. An example of this type of model with a single factor at both levels, $M_2 = M_3 = 1$, is given in path diagram form in Figure 1(b). For continuous responses or responses that can be modeled using an underlying normal response y^*_{ijk}, the general model in (15) corresponds to allowing completely different factor models for the within and between cluster covariance matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, where the I-dimensional vector of responses is modeled as

$$\mathbf{y}^*_{jk} \sim N(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1)$$
$$\boldsymbol{\mu}_k \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_2),$$

and $\boldsymbol{\mu}$ and $\boldsymbol{\mu}_k$ are intercepts (e.g., Longford & Muthén, 1992; Poon & Lee, 1992; Longford, 1993; Linda, Lee, & Poon 1993; Muthén, 1994; Lee & Shi, 2001).

The variance components factor model in (13) and (14) results if $M_2 = M_3$, $\boldsymbol{\lambda}_m^{(2)} = \boldsymbol{\lambda}_m^{(3)}$ and the unique factors are omitted. It is surprising that this much simpler model is hardly ever considered in the multilevel structural equation modeling literature, since it is a special case with an intuitive interpretation analogous to MIMIC models.

As mentioned in the introduction, specification of multilevel structural equation models via separate structural equation models for the within and between cluster covariance matrices is often cumbersome. For example, incorporating a direct effect of a cluster-level latent variable on a unit-level latent variable as in Figure 1(a) requires the set-up in Figure 1(b), where the factor loadings in the within and between model are constrained equal and the unique factor variances in the between model are constrained to 0. Estimating the corresponding regression coefficient b (in a model where it is identified), requires nonlinear constraints of the form $\boldsymbol{\lambda}_1^{(3)} = b\boldsymbol{\lambda}_1^{(2)}$, which can be imposed via "phantom" variables (Rindskopf, 1984).

Multilevel Structural Equation Models

In addition to being of interest in their own right, multilevel factor models can represent measurement models within a structural equation model. In multilevel structural equation models, there are several ways of specifying regressions of factors on latent or observed explanatory variables. Consider for simplicity the case of a single cluster-level observed covariate w_k. If a unidimensional version of the multilevel factor model in (15) is used for the response, the regression part of the reduced form becomes (a) $b_1 w_k \boldsymbol{\lambda}_1^{(2)} \boldsymbol{\delta}_{1i}^{(2)}$ if the unit-level common factor $\eta_{1jk}^{(2)}$ is regressed on w_k, (b) $b_2 w_k \boldsymbol{\lambda}_1^{(3)} \boldsymbol{\delta}_{1i}^{(2)}$ if the cluster-level common factor $\eta_{1jk}^{(3)}$ is regressed on w_k, (c) $b_1 w_k \boldsymbol{\lambda}_1^{(2)} \boldsymbol{\delta}_{1i}^{(2)} + b_2 w_k \boldsymbol{\lambda}_1^{(3)} \boldsymbol{\delta}_{1i}^{(3)}$ if both the unit and cluster-level common factors are regressed on w_k, where b_1 and b_2 are elements of $\mathbf{B}$. These models are not equivalent and case (c) can be estimated to distinguish between effects on the cluster and unit-level factors. However, if a variance components factor model is used for the latent response, the reduced form terms due to regressions of the unit or cluster-level latent variables are identical, since $\boldsymbol{\lambda}_1^{(2)} = \boldsymbol{\lambda}_1^{(3)}$, greatly simplifying model specification. The same issues apply to latent covariates, which may themselves be regressed on observed covariates. In order to limit conceptual and notational complexity, we will in the sequel confine the explicit development to unidimensional variance components factor models.

Latent response and explanatory variables can vary at different levels. For example, predictors of pupils' performance may include a class-level latent variable teachers' experience and a school-level latent covariate such as headmaster's attitude. The response model will therefore typically include responses varying at different levels. In this section we illustrate different multilevel structures for latent covariates assuming that the latent response variables vary at the unit level. Similar structures could also be specified for latent response variables. Two kinds of multilevel structural equation models are discussed, models where the latent covariates are either measured by same-level or lower-level items.

Latent Covariates Measured by Same-Level Items

Consider first a model where the latent covariate varies at the cluster level. All responses are stacked into a single response vector $\mathbf{y}$ and a single response model specifies the measurement models for all latent covariates and response variables. We therefore find it convenient to use subscripts R and C for terms corresponding to responses and covariates, respectively. Where these subscripts are sufficient to uniquely label different dummy vectors and factor loading vectors, we will omit the (l) superscript. Let $\boldsymbol{\delta}_{Ri}$ be the dummy vector for the items measuring the unit level

latent response and $\boldsymbol{\delta}_{Ci}$ the dummy vector for the items measuring a cluster-level covariate. A typical model might then be

$$v_{ijk} = \boldsymbol{\beta}' \mathbf{x}_{ijk} + \eta_{Rjk}^{(2)} \boldsymbol{\lambda}_R' \boldsymbol{\delta}_{Ri} + \eta_{Ck}^{(3)} \boldsymbol{\lambda}_C' \boldsymbol{\delta}_{Ci}$$

$$\begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Ck}^{(3)} \end{bmatrix} = \begin{bmatrix} 0 & b_{12} & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Ck}^{(3)} \\ \eta_{Rk}^{(3)} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} w_{Rjk} \\ w_{Ck} \end{bmatrix} + \begin{bmatrix} \zeta_{Rjk}^{(2)} \\ \zeta_{Ck}^{(3)} \end{bmatrix}. \quad (16)$$

The first term of the response model accommodates the part of the mean structure not specified through the structural model for the latent variables. The second term represents the measurement model for the latent response variable and the third term represents the measurement model for a cluster-level covariate. In the structural model, the unit-level latent response variable $\eta_{Rjk}^{(2)}$ is regressed on the latent cluster-level covariate $\eta_{Ck}^{(3)}$ and on another cluster-level latent variable $\eta_{Rk}^{(3)}$, representing a random intercept that is uncorrelated with the latent covariate. In addition, the unit-level latent response variable is regressed on a unit-level covariate w_{Rjk} and the cluster-level latent covariate is regressed on a cluster-level covariate w_{Ck}. We have omitted latent variables on the left-hand side of the structural model that are not regressed on other latent or observed variables ($\eta_{Rk}^{(3)}$ in this case) and will continue to do so in the sequel. The model is shown as a path diagram in the upper left panel of Figure 2, where we have labeled the paths in the structural
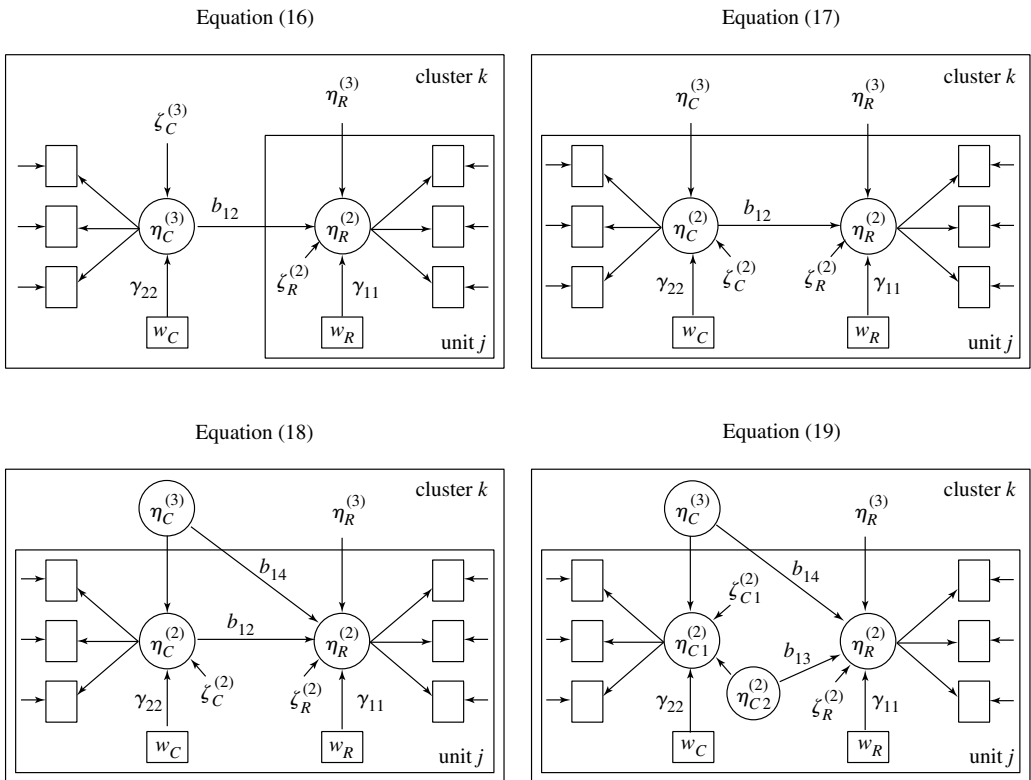

FIGURE 2.
Path diagrams for structural equation models with latent covariates at different levels.

model except if the coefficient is set to 1. Further latent covariates could be added where latent covariates at the same level could be mutually correlated.

Now consider a latent covariate that varies at the unit level instead of the cluster level and has a cluster-level component of variation. Letting $\boldsymbol{\delta}_{Ri}$ and $\boldsymbol{\delta}_{Ci}$ represent the dummy vectors for the items measuring the latent response variable $\eta_{Rjk}^{(2)}$ and the unit-level latent covariate $\eta_{Cjk}^{(2)}$, respectively, the model is constructed as

$$\nu_{ijk} = \boldsymbol{\beta}'\mathbf{x}_{ijk} + \eta_{Rjk}^{(2)}\boldsymbol{\lambda}_R'\boldsymbol{\delta}_{Ri} + \eta_{Cjk}^{(2)}\boldsymbol{\lambda}_C'\boldsymbol{\delta}_{Ci}$$

$$\begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Cjk}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & b_{12} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Cjk}^{(2)} \\ \eta_{Rk}^{(3)} \\ \eta_{Ck}^{(3)} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} w_{Rjk} \\ w_{Cjk} \end{bmatrix} + \begin{bmatrix} \zeta_{Rjk}^{(2)} \\ \zeta_{Cjk}^{(2)} \end{bmatrix}. \quad (17)$$

The latent response variable $\eta_{Rjk}^{(2)}$ is regressed on the latent covariate $\eta_{Cjk}^{(2)}$ and has a cluster-level random intercept $\eta_{Rk}^{(3)}$. The latent covariate $\eta_{Cjk}^{(2)}$ has a component of variation $\eta_{Ck}^{(3)}$ at the cluster level. The measurement model for the latent covariate is therefore a variance components factor model as in (13) and (14). A path diagram for this model is shown in the upper-right panel of Figure 2. Fox and Glas (2003) use both cluster and unit-level latent covariates in a model with an observed response variable. However, their model does not include a cluster-level component of variation for the unit-level latent covariate.

Latent Covariates (Contexts) Measured by Lower-Level Items

Effects of ecological contexts are often of scientific interest in multilevel settings. An example is school climate, which is reflected in individual pupils' attitudes. In a variance components factor model for items measuring these attitudes, the school-level latent variable can be interpreted as school climate and could be included as a predictor of, for example, individual-level ability. The response model coincides with that of Equation (17) and the structural model is

$$\begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Cjk}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & b_{12} & 1 & b_{14} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{Cjk}^{(2)} \\ \eta_{Rk}^{(3)} \\ \eta_{Ck}^{(3)} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} w_{Rjk} \\ w_{Cjk} \end{bmatrix} + \begin{bmatrix} \zeta_{Rjk}^{(2)} \\ \zeta_{Cjk}^{(2)} \end{bmatrix}, \quad (18)$$

where b_{14} represents the effect of school climate, $\eta_{Ck}^{(3)}$. Here an effect b_{12} of the pupil's own attitude $\eta_{Cjk}^{(2)}$ has also been included. A path diagram for this model is shown in the lower-left panel of Figure 2.

Another possibility is to consider the effect of deviations of pupils' latent attitudes from the school means. The notion of such centered effects is common in sociological research, for instance, when studying relative deprivation. In the model above $\zeta_{Cjk}^{(2)}$ represents such a centered latent variable since it is the error term of the regression of pupils' attitude on their schools' mean attitudes. However, $\zeta_{Cjk}^{(2)}$ cannot be included as a regressor in our framework. We can get around this problem by using the following specification:

$$
\begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{C1jk}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 & 0 & b_{13} & 1 & b_{14} \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_{Rjk}^{(2)} \\ \eta_{C1jk}^{(2)} \\ \eta_{C2jk}^{(2)} \\ \eta_{Rk}^{(3)} \\ \eta_{Ck}^{(3)} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & 0 \\ 0 & \gamma_{22} \end{bmatrix} \begin{bmatrix} w_{Rjk} \\ w_{Cjk} \end{bmatrix} + \begin{bmatrix} \zeta_{Rjk}^{(2)} \\ \zeta_{C1jk}^{(2)} \end{bmatrix},
$$

$$(19)$$

where $\mathrm{var}(\zeta_{C1jk}^{(2)}) = 0$. Here the latent variable $\eta_{C2jk}^{(2)}$ acts as the disturbance of the regression for $\eta_{C1jk}^{(2)}$ since the variance of $\zeta_{C1jk}^{(2)}$, which is required by the framework, is set to zero. b_{13} therefore represents the effect of the centered latent covariate. $\eta_{C1jk}^{(2)}$ takes the place of $\eta_{Cjk}^{(2)}$ in the response and structural model (18). A path diagram for this model is shown in the lower-right panel of Figure 2.

Raudenbush and Sampson (1999b) use latent variables to represent the neighborhood-level contexts or "ecological constructs" physical and social disorder. Items reflecting physical and social disorder (level 1) are measured at the housing block level (level 2) within neighborhoods (level 3). A three-level model is developed where correlated random intercepts at level three represent the ecological constructs of interest. However, their models do not include factor loadings (one parameter item-response models are used at level 1) or regressions among latent variables.

The models described in this subsection have assumed that a reflective (not a formative) measurement model (e.g., Bollen, 1989) applies for the context. If this assumption is appropriate, it would generally be problematic to substitute sumscores or factor scores for the latent variables (Skrondal & Laake, 2001), an approach often adopted in multilevel modeling.

Multilevel Structural Equation Models with Random Coefficients

Observed Response Variable

A conventional multilevel regression model can be specified as

$$
v_i = \boldsymbol{\beta}' \mathbf{x}_i + \sum_{l=2}^{L} \sum_{m=1}^{M_l} \eta_{bm}^{(l)} z_{bm}^{(l)},
$$

where we have suppressed all observation indices except i. Here there is one observed variable $z_{bm}^{(l)}$ for each latent variable $\eta_{bm}^{(l)}$ and the corresponding factor loading $\lambda_{bm}^{(l)}$ is set to 1 so that

$$
\sum_{m=1}^{M_l} \eta_{bm}^{(l)} z_{bm}^{(l)} = \boldsymbol{\eta}_b^{(l)'} \mathbf{z}_b^{(l)}.
$$

We now add a unit-level latent covariate. Stacking the response and the items measuring the covariate in a single response vector and using dummy vectors $\boldsymbol{\delta}_{Ri} = d_{ir}$ for the response (assumed to be in position $i = r$) and $\mathbf{d}_i$ for all items, the model can be written as

$$
v_i = \boldsymbol{\beta}'(\mathbf{x}_i d_{ir}) + \eta_C^{(2)} \boldsymbol{\lambda}_C' \mathbf{d}_i + \sum_{l=3}^{L} \boldsymbol{\eta}_b^{(l)'} (\mathbf{z}_b^{(l)} d_{ir}). \tag{20}
$$

The first and third terms represent the multilevel regression model for the response variable with the lowest-level random effects now at level 3 since the units form level 2 and the items level 1. The second term represents both the measurement model for the latent covariate $\eta_C^{(2)}$ and the regression of the response variable on the latent covariate with regression coefficient λ_{Cr}. The

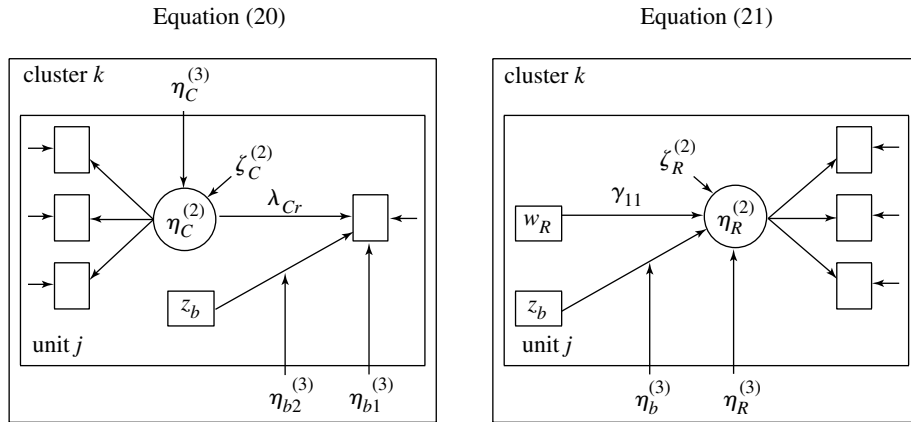Equation (20)   Equation (21)

FIGURE 3.
Random coefficient models with latent covariate (left panel) and latent response variable (right panel).

scale of the latent covariate is that of the first item (since $\lambda_{C1} = 1$). The latent covariate may have a component of variation at level 3 (or higher) which can be specified in the structural model as

$$\eta_C^{(2)} = b_{12}\eta_C^{(3)} + \zeta_C^{(2)}, \quad b_{12} = 1.$$

A path diagram of this model for $L = 3$ with a single random coefficient is shown in the left panel of Figure 3. The random coefficient is represented by a latent variable having an arrow pointing from it to the path representing the corresponding fixed coefficient, a convention previously used in Pickles, Pickering, Simonoff, Meyer, Silberg, & Maes (1998).

Further variables measured with error, including higher-level covariates or "contexts," can be added to the model using possibly correlated latent variables at levels 2 and higher. However, in the present framework, a latent covariate cannot have a random coefficient since $\boldsymbol{\Gamma}$ in the structural model is a matrix of constant parameters. This feature has been included in a model by Fox (2001).

If the variable $z_{bm}^{(l)}$ associated with the mth random coefficient at level l varies at a level lower than l (as in the path diagram), $\eta_{bm}^{(l)}$ can be interpreted as a conventional random coefficient. However, if $z_{bm}^{(l)}$ varies at the same or higher level than l, $\eta_{bm}^{(l)}z_{bm}^{(l)}$ represents a heteroscedastic random intercept at level l. A random intercept whose variance depends on several covariates $\mathbf{z}_{bm}^{(l)}$, can be modeled using $\eta_{bm}^{(l)}\boldsymbol{\lambda}_{bm}^{(l)\prime}\mathbf{z}_{bm}^{(l)}$. Similarly, a heteroscedastic random coefficient for a covariate $z_{bm}^{(l)}$ can be modeled using $\eta_{bm}^{(l)}\boldsymbol{\lambda}_{bm}^{(l)\prime}(z_{bm}^{(l)}\mathbf{v}_m^{(l)})$, where $\mathbf{v}_m^{(l)}$ are the variables inducing heteroscedasticity.

Latent Response Variable

If the response variable is latent instead of observed and measured by multiple items, a multilevel random intercept model with a single random coefficient at level 3 can be specified as

$$\nu_i = \boldsymbol{\beta}'\mathbf{x}_i + \eta_R^{(2)}\boldsymbol{\lambda}_R'\boldsymbol{\delta}_{Ri} + \eta_b^{(3)}\boldsymbol{\lambda}_b'(\boldsymbol{\delta}_{Ri}z_b), \quad \boldsymbol{\lambda}_b = \boldsymbol{\lambda}_R$$

$$\eta_R^{(2)} = \boldsymbol{\gamma}'\mathbf{w}_R + \mathbf{b}_1'\boldsymbol{\eta}_R + \zeta_R^{(2)}, \quad \mathbf{b}_1 = \mathbf{1}. \tag{21}$$

The second term in the response model represents the measurement model for the latent response variable $\eta_R^{(2)}$. The structural model defines a random intercept model for the latent response

variable where $\boldsymbol{\eta}_R$ is a vector of random intercepts at levels $l = 3, \ldots, L$. Since the random coefficient term $\eta_b^{(3)} z_b$ cannot be included in the structural model, it is included in the response model as $\eta_b^{(3)} \boldsymbol{\lambda}_b'(\boldsymbol{\delta}_{Ri} z_b) = \eta_b^{(3)} z_b \boldsymbol{\lambda}_R' \boldsymbol{\delta}_{Ri}$ by imposing the constraint $\boldsymbol{\lambda}_b = \boldsymbol{\lambda}_R$. The model is shown in the right panel of Figure 3 for the three-level case with one observed covariate w_R in the structural model. Further random coefficients or covariates measured with error can be included in a straightforward manner.

Finally, models can include regressions between factors and random coefficients. For example, headmasters' attitude, measured by a number of items and represented by a factor at the school level, may affect the random slope of pupils' progress over time. Let items i be nested in occasions t nested in pupils j nested in schools k and all responses stacked into a single response vector where d_{ir} is the dummy variable for the pupil-level outcome and and $\boldsymbol{\delta}_{Ci}$ is the dummy vector for the headmaster-level items. The model can then be written as

$$v_{itjk} = \boldsymbol{\beta}' \mathbf{x}_{itjk} + \eta_{Rjk}^{(2)} d_{ir} + \eta_{bjk}^{(2)}(z_{btjk} d_{ir}) + \eta_{Ck}^{(3)} \boldsymbol{\lambda}_C' \boldsymbol{\delta}_{Ci}$$

$$\eta_{Rjk}^{(2)} = b_{13} \eta_{Ck}^{(3)} + b_{14} \eta_{Rk}^{(3)} + \zeta_{Rjk}^{(2)}, \quad b_{14} = 1$$

$$\eta_{bjk}^{(2)} = b_{23} \eta_{Ck}^{(3)} + b_{25} \eta_{bk}^{(3)} + \zeta_{bjk}^{(2)}, \quad b_{25} = 1.$$

The second and third terms of the response model specify a growth curve model for pupils' outcome with a random intercept $\eta_{Rjk}^{(2)}$ and slope $\eta_{bjk}^{(2)}$ of time z_{btjk} at the pupil level. The last term specifies the measurement model for headmasters' attitude $\eta_{Ck}^{(3)}$. In the structural model, the pupil-level random intercepts and slopes are regressed on headmasters' latent attitude. In addition, there are school-level random effects $\eta_{Rk}^{(3)}$ and $\eta_{bk}^{(3)}$ for the intercept and slope, which may be correlated. Note that these random effects could alternatively be included directly in the response model as $\eta_{Rk}^{(3)} d_{ir}$ and $\eta_{bk}^{(3)}(z_{btjk} d_{ir})$. The disturbances $\zeta_{Rjk}^{(2)}$ and $\zeta_{bjk}^{(2)}$ may also be correlated. The model is shown as a path diagram in Figure 4. The first diagram uses an arrow pointing at a path to represent the random coefficient. This diagram is appropriate for a balanced or unbalanced covariate z_{btjk} whereas the second diagram assumes that the covariate takes on the same (three) values z_{bt} for each unit. The second diagram uses the conventional path diagram representation of random intercepts and coefficients. Here the factor loadings for the random intercept would be set to 1 and those for the random coefficient to the times z_{b1}, z_{b2} and z_{b3}. Alternatively, one of the factor loadings could be estimated to allow for nonlinear growth (e.g., McArdle, 1986; Meredith & Tisak, 1990).

In the above example a random coefficient and intercept were regressed on a factor, but factors could also be regressed on random coefficients. More complex structural models could be specified in which random coefficients and/or factors at different levels can be predictors, intervening variables or responses.

Application

Respondents in the British Social Attitudes Survey Panel 1983–1986 (Social and Community Planning Research, 1987)[1] were asked whether or not abortion should be allowed by law under the following circumstances:

- the woman decides on her own she does not wish to have the child [woman]
- the couple agree that they do not wish to have the child [couple]
- the woman is not married and does not wish to marry the man [marriage]

---

[1]Data were supplied by the UK Data Archive. Neither the original data collectors nor the archive are responsible for the present analyses.
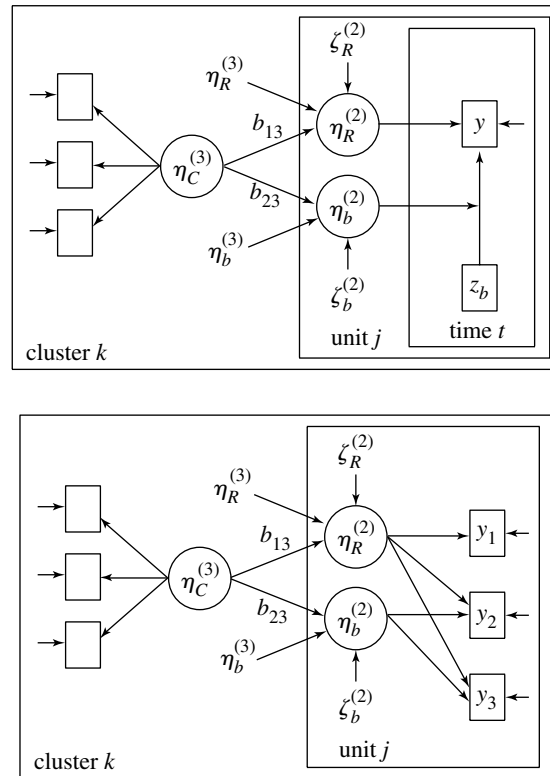
FIGURE 4.
Level 2 random intercept and slope regressed on level 3 latent covariate. (a) general representation (b) representation for balanced covariate. Correlations between the the random effects $\eta_R^{(3)}$ and $\eta_b^{(3)}$ or between $\zeta_R^{(2)}$ and $\zeta_b^{(2)}$ would be indicated by curved double-headed arrows.

- the couple cannot afford any more children [financial]
- there is a strong chance of a defect in the baby [defect]
- the woman's health is seriously endangered by the pregnancy [risk]
- the woman became pregnant as a result of rape [rape]

The data have a multilevel structure with panel waves nested in individuals nested in polling districts. There were 14143 responses to the seven items over the four panel waves from 734 individuals in 57 polling districts. The multilevel design was highly unbalanced with 49% of subjects responding to at least one item in all four panel waves, 12% in three waves, 13% in two waves and 25% in one wave. Unit nonresponse was therefore common, but if an interview took place, item nonresponse occurred in only 7% of cases. We will not explicitly model unit or item nonresponse and therefore assume that responses are missing at random (MAR).

Previous multilevel analyses of these data have used raw sumscores or scores constructed from item response models as response variable (Knott, Albanese, & Galbraith, 1990; Wiggins, Ashworth, O'Muircheartaigh, & Galbraith, 1990). However, using such constructed scores as proxies for latent variables has been demonstrated to be highly problematic, leading to biased standard errors and often to inconsistent parameter estimates (Skrondal & Laake, 2001). Hence, we use multilevel factor models with a logit link for the dichotomous items. We used the change in deviance to choose between competing models. Each model is fitted a number of times using

adaptive quadrature with different numbers of quadrature points per dimension to ensure reliable results.

Initially, we focus on between subject heterogeneity and subsequently also include heterogeneity between polling districts. It is plausible that in addition to a "general attitude" factor $\eta_{Gjk}^{(3)}$ measured by all items, there may be an independent "extreme circumstance" factor $\eta_{Ejk}^{(3)}$ representing people's additional inclination to be in favor of abortion when there is a strong chance of a defect in the baby, a high risk to the woman, or where the pregnancy was a result of rape. Using indices i for item or situation (level 1), t for time (level 2), j for subject (level 3) and k for polling district (level 4), the two-factor model can be written as

$$v_{itjk} = \boldsymbol{\beta}'\mathbf{d}_i + \eta_{Gjk}^{(3)}\boldsymbol{\lambda}_G'\mathbf{d}_i + \eta_{Ejk}^{(3)}\boldsymbol{\lambda}_E'\boldsymbol{\delta}_{Ei}, \tag{22}$$

where $\boldsymbol{\delta}_{Ei}$ is a vector of dummy variables for the three extreme situations. A unidimensional factor model appears to be inadequate since removing the extreme circumstance factor increases the deviance by 207.7 with 3 degrees of freedom.

Since there are repeated responses for each subject and item, item specific unique factors can be included at the subject level:

$$v_{itjk} = \boldsymbol{\beta}'\mathbf{d}_i + \eta_{Gjk}^{(3)}\boldsymbol{\lambda}_G'\mathbf{d}_i + \eta_{Ejk}^{(3)}\boldsymbol{\lambda}_E'\boldsymbol{\delta}_{Ei} + \sum_{m=1}^{7} \eta_{Umjk}^{(3)}d_{im}, \tag{23}$$

where the latent variables are mutually independent. The unique factors $\eta_{Umjk}^{(3)}$ in the last term can be interpreted as heterogeneity between subjects in their attitudes to specific items, inducing additional dependence between responses over time not accounted for by the common factor. Evaluation of the log-likelihood for this model requires integration over nine dimensions at level 3. To reduce the dimensionality of integration, the items i can be treated as level 2 units so that time becomes level 1 and the model is reparameterized as

$$v_{tijk} = \boldsymbol{\beta}'\mathbf{d}_i + \eta_{Uijk}^{(2)}\boldsymbol{\lambda}_U'\mathbf{d}_i + \eta_{Gjk}^{(3)}\boldsymbol{\lambda}_G'\mathbf{d}_i + \eta_{Ejk}^{(3)}\boldsymbol{\lambda}_E'\boldsymbol{\delta}_{Ei}. \tag{24}$$

Here the last term in (23) that evaluates to $\eta_{Uijk}^{(3)}$ for item i has been replaced by $\eta_{Uijk}^{(2)}\lambda_{Ui}$. Whereas the $\eta_{Uijk}^{(3)}$ are treated as separate latent variables for the items, $i = 1, \ldots, 7$, $\eta_{Uijk}^{(2)}$ is a single latent variable with different realizations for different items i. The purpose of $\boldsymbol{\lambda}_U$ is to allow the unique factor variances to differ between the items. The models are equivalent since both $\eta_{Uijk}^{(3)}$ and $\eta_{Uijk}^{(2)}\lambda_{Ui}$ vary between items, are uncorrelated across items, and have item-specific variances. The advantage of (24) is that a nine-dimensional integral at level 3 has been replaced by a one-dimensional integral at level 2 and a two-dimensional integral at level 3. It is often possible to reduce the dimensionality of integration by reparameterization (e.g., Rabe-Hesketh & Skrondal, 2001). Adding unique factors at the subject level to the two-factor model decreases the deviance by 12.6, a small change for seven additional parameters.

Introducing district-level latent variables in addition to subject-level latent variables, the common factors can be allowed to vary between polling districts, giving two-dimensional variance components factor models. Allowing the general attitude factor to vary between districts decreases the deviance by 8.2 with one extra parameter whereas the deviance decreases by only 3.2 for the extreme circumstance factor. The retained model is therefore the response model in (24) plus the structural model

$$\eta_{Gjk}^{(3)} = \eta_{Gk}^{(4)} + \zeta_{Gij}^{(3)}.$$

TABLE 1.
Estimates for the Multilevel Variance Components Logit Factor Model

	General	Extreme
Fixed Part		
Intercepts:		
woman	−0.83 (0.14)	
couple	−0.17 (0.15)	
marriage	−0.28 (0.16)	
financial	−0.01 (0.14)	
defect	3.79 (0.27)	
risk	5.90 (0.56)	
rape	4.82 (0.39)	
Random Part: Subject level		
Factor loadings:	General	Extreme
woman	1	0
couple	1.13 (0.08)	0
marriage	1.21 (0.09)	0
financial	1.01 (0.08)	0
defect	0.78 (0.09)	1
risk	0.73 (0.13)	1.53 (0.26)
rape	0.72 (0.11)	1.23 (0.21)
Common factor variance:	5.22 (0.67)	3.30 (0.80)
Random Part: District level		
Common factor variance:	0.36 (0.17)	0
Log-likelihood	−5160.9	

Including unique factors at the district level increases the dimension of integration at level 4 from 1 to 8. The dimensionality cannot be reduced by reparameterization in this case. We therefore included each unique factor separately, but the changes in deviance were small.

Estimates for the retained multilevel variance components logit factor model are given in Table 1. These were obtained using adaptive quadrature with 10 points per dimension which gave very similar results to 8 and 5 points per dimension. As expected, the intercepts for the extreme circumstance items were much larger than for the others due the larger prevalence of endorsing these items. Both a general attitude and extreme circumstance factor were required at the subject level. Only the general attitude factor appeared to vary at the polling district level, but with a relatively small variance.

Concluding Remarks

A unifying framework for multilevel structural equation modeling has been developed. The framework handles various response processes including continuous, censored, grouped, ordinal, dichotomous, and unordered polytomous responses, as well as counts and durations in continuous or discrete time. All kinds of imbalance (missing data, unbalanced multilevel designs, and unbalanced covariates) are accommodated. Furthermore, explicit selection models can be incorporated if the assumption of data missing at random (MAR) is untenable. The framework allows an arbitrary number of hierarchical levels to be specified and structural relations among factors and/or random coefficients at different levels to be included. Important special cases of the general framework were outlined and potential applications described. However, these models in no way exhaust the possibilities of the framework.

As in conventional structural equation modeling, it is necessary to impose identification restrictions on the parameters of models within the framework. Unfortunately, the generality of

the model framework exacerbates the complexity of identification analysis and prohibits derivation of general identification conditions. Identification must thus proceed on a case by case basis. Fragile identification, where the model is technically identified but identification hinges on scarce information, should be avoided (e.g., Rabe-Hesketh & Skrondal, 2001).

We have developed a maximum likelihood approach for estimating all models in the framework. This is implemented in a freely available program gllamm, which runs in the general purpose statistical package Stata (StataCorp, 2003). gllamm also produces empirical Bayes predictions of all latent variables with corresponding standard errors and can be used to simulate responses from a model. If there are many latent variables gllamm can be slow, but the number of latent variables can often be reduced by reparameterization as shown in our application.

Other currently available approaches tend to be limited to continuous responses, certain model and data structures, and rely on ad hoc estimation procedures. Perhaps the simplest approach is to first estimate separate structural equation models for the clusters and then obtain the between-cluster covariance matrix of the parameter estimates (e.g., Chou, Bentler, & Pentz, 2000). However, all parameters, including factor loadings, are in this case treated as random with unconstrained covariance matrix, severely limiting the types of multilevel structural equation model that can be formulated. Furthermore, parameter estimation is fragile for small clusters; the method does not provide proper standard errors and the properties of the estimators are unknown. A better approach for continuous responses is implemented in Mplus (Muthén & Muthén, 1998; Muthén, 2002), which fits models to estimated within and between cluster covariance matrices. This yields a maximum likelihood estimator in the case of balanced multilevel designs (Muthén, 1989) and a consistent so-called MUML estimator in unbalanced multilevel designs (Muthén, 1989, 1994). However, missing data are not permitted except if they are handled by multi-group analysis. Goldstein (1995) suggests an alternative two-stage method. First, the within and between (residual) covariance matrices are estimated using conventional multilevel regression modeling, which handles unbalanced data. Second, structural equation modeling is performed based on these covariance matrices. This approach has some intuitive appeal for models where the covariance matrices represent sufficient statistics. Although the estimators are consistent, the statistical properties of the two-stage method are generally unknown. Hox (2002) gives a useful overview of most of these ad-hoc approaches.

Methods based on separate models for the within and between covariance matrices suffer from the six limitations discussed in the introduction, although the first two of these are addressed in ongoing work (Asparouhov & Muthén, 2004) which is implemented in Mplus. Some multilevel models with structural relations among latent variables can be estimated in HLM 5 (Raudenbush, Bryk, & Congdon, 2000) but only for the case of known factor loadings. The current version of MLwiN includes Markov Chain Monte Carlo (MCMC) methods for multilevel factor models with continuous responses (Goldstein & Browne, 2002).

Other estimation methods for more general models have been suggested but are not generally available in software. For continuous data, maximum likelihood estimation has been proposed for unbalanced multilevel designs with missing items (Longford & Muthén, 1992), for example using an EM algorithm (Raudenbush, 1995; Lee & Tsang, 1999) or a generalization of the iterated generalized least squares algorithm (Yang, Pickles, & Taylor, 1999; Yang & Pickles, 2004). An EM algorithm for unbalanced continuous two-level data has recently been implemented in EQS (e.g., Liang & Bentler, 2003). For binary data, MCMC methods have been proposed by Ansari and Jedidi (2000) and Fox and Glas (2001). BUGS can be used to estimate general models by MCMC (Spiegelhalter, Thomas, Best, & Gilks, 1996), although model specification and monitoring of stationarity require some expertise.

The framework discussed in this paper does not include models with discrete, nonnormal or "nonparametric" latent variables. Such models are discussed in Skrondal and Rabe-Hesketh (2004) and can be estimated in gllamm. Currently the structural model cannot include simultane-

ous effects or interactions between latent and observed variables and the linear predictor cannot include products of latent variables. Such extensions would be useful for example for specifying models with random coefficients for latent covariates and/or random intercepts with randomly varying variances. Our framework encompasses a very large range of multilevel structural equation models and we are only beginning to understand the scope of multilevel structural equation modeling.

References

Ansari, A., & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika*, *65*, 475–496.

Arminger, G., & Sobel, M. (1990). Pseudo maximum likelihood estimation of mean- and covariance structures with missing data. *Journal of the American Statistical Association*, *85*, 195–203.

Asparouhov, T., & Muthén, B.O. (2004). Full-information maximum-likelihood estimation of general two-level latent variable models. Draft.

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

Böckenholt, U. (2001). Hierarchical modeling of paired comparison data. *Psychological Methods*, *6*, 49–66.

Carroll, R.J., Ruppert, D., & Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*. London: Chapman & Hall.

Chou, C.-P., Bentler, P.M., & Pentz, M.A. (2000). A two-stage approach to multilevel structural equation models: application to longitudinal data. In T.D. Little, K.U. Schnabel, & J. Baumert (Eds.), *Modeling Longitudinal and Multilevel Data* (pp. 33–49). Mahwah, NJ: Erlbaum.

Clayton, D. (1988). The analysis of event history data: a review of progress and outstanding problems. *Statistics in Medicine*, *7*, 819–841.

Clogg, C. C. (1995). Latent class 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.

Dohoo, I.R., Tillard, E., Stryhn, H., & Faye, B. (2001). The use of multilevel models to evaluate sources of variation in reproductive performance in dairy cattle. *Preventive Veterinary Medicine*, *50*, 127–144.

Fox, J.P. (2001). *Multilevel IRT: A Bayesian Perspective on Estimating Parameters and Testing Statistical Hypotheses*. Ph.D. thesis, University of Twente, Enschede.

Fox, J.P., & Glas, C. A.W. (2001). Bayesian estimation of a multilevel IRT model using Gibbs sampling. *Psychometrika*, *66*, 271–288.

Fox, J.P., & Glas, C. A.W. (2003). Bayesian modeling of measurement error in predictor variables using item response theory. *Psychometrika*, *68*, 169–191.

Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalised least squares. *Biometrika*, *73*, 43–56.

Goldstein, H. (1995). *Multilevel Statistical Models*. London: Arnold.

Goldstein, H., & Browne, W. (2002). Multilevel factor analysis modeling using Markov Chain Monte Carlo (MCMC) estimation. In G.A. Marcoulides & I. Moustaki (Eds.), *Latent Variable and Latent Structure Models*. Mahwah, NJ: Erlbaum.

Goldstein, H., & McDonald, R.P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, *53*, 455–467.

Hagenaars, J.A. (1988). Latent structure models with direct effects between indicators: local dependence models. *Sociological Methods & Research*, *16*, 379–405.

Harper, D. (1972). Local dependence latent structure models. *Psychometrika*, *37*, 53–59.

Heckman, J.J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*, 153–161.

Heckman, J.J., & Singer, B. (1984). A method of minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, *52*, 271–320.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum.

Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.

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

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

Jöreskog, K.G., & Sörbom, D. (1989). *LISREL 7: A Guide to the Program and Applications*. Chicago, IL: SPSS Publications.

Knott, M., Albanese, M.T., & Galbraith, J.I. (1990). Scoring attitudes to abortion. *The Statistician*, *40*, 217–223.

Laird, N.M. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, *73*, 805–811.

Lee, S.-Y., & Shi, J.-Q. (2001). Maximum likelihood estimation of two-level latent variable models with mixed continuous and polytomous data. *Biometrics*, *57*, 787–794.

Lee, S.-Y., & Tsang, S.-Y. (1999). Constrained maximum likelihood estimation of two-level covariance structure models via EM type algorithms. *Psychometrika*, *64*, 435–450.

Lesaffre, E., & Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example. *Applied Statistics*, *50*, 325–335.

Liang, J., & Bentler, P. M. (2003). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, in press.

Linda, N.Y., Lee, S.-Y., & Poon, W.-Y. (1993). Covariance structure analysis with three level data. *Computational Statistics & Data Analysis, 15*, 159–178.

Longford, N.T. (1993). *Random Coefficient Models*. Oxford: Oxford University Press.

Longford, N.T., & Muthén, B.O. (1992). Factor analysis for clustered observations. *Psychometrika*, *57*, 581–597.

McArdle, J.J. (1986). Latent variable growth within behavior genetic models. *Behavior Genetics*, *16*, 163–200.

McDonald, R.P., & Goldstein, H. (1989). Balanced and unbalanced designs for linear structural relations in two-level data. *British Journal of Mathematical and Statistical Psychology*, *42*, 215–232.

Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*, 107–122.

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

Muthén, B.O. (1985). A method for studying the homogeneity of test items with respect to other relevant variables. *Journal of Educational Statistics*, *10*, 121–132.

Muthén, B.O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, *54*, 557–585.

Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376–398.

Muthén, B.O. (1997). Latent variable modeling of longitudinal and multilevel data. In A.E. Raftery (Ed.), *Sociological Methodology 1997* (pp. 453–480). Cambridge, MA: Blackwell.

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

Muthén, L.K., & Muthén, B.O. (1998). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.

Neale, M.C., & Cardon, L.R. (1992). *Methodology for Genetic Studies of Twins and Families*. London: Kluwer.

Pickles, A., Pickering, K., Simonoff, E., Meyer, J., Silberg, J., & Maes, H. (1998). Genetic clocks and soft events: A twin model for pubertal development and other recalled sequences of developmental milestones. *Behavior Genetics*, *28*, 243–253.

Plummer, M., & Clayton, D. (1993). Measurement error in dietary assessment: an investigation using covariance structure models. Part II. *Statistics in Medicine*, *12*, 937–948.

Poon, W.-Y., & Lee, S.-Y. (1992). Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statistics and Probability Letters*, *14*, 25–30.

Rabe-Hesketh, S., & Pickles, A. (1999). Generalised linear latent and mixed models. In H. Friedl, A. Berghold & G. Kauermann (Eds.), *Proceedings of the 14th International Workshop on Statistical Modeling* (pp. 332–339). Graz, Austria.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001a). GLLAMM: A general class of multilevel models and a Stata program. *Multilevel Modelling Newsletter*, *13*, 17–23.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2001b). *GLLAMM Manual*. Tech. rept. 2001/01. Department of Biostatistics and Computing, Institute of Psychiatry, King's College, University of London. Downloadable from http://www.gllamm.org.

Rabe-Hesketh, S., Pickles, A., & Skrondal, A. (2003). Correcting for covariate measurement error in logistic regression using nonparametric maximum likelihood estimation. *Statistical Modelling, 3*, 215-232.

Rabe-Hesketh, S., Pickles, A., & Taylor, C. (2000). sg129: Generalized linear latent and mixed models. *Stata Technical Bulletin*, *53*, 47–57.

Rabe-Hesketh, S., & Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics*, *57*, 1256–1264.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2002). Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, *2*, 1–21.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2003). Maximum likelihood estimation of generalized linear models with covariate measurement error. *The Stata Journal*, *3*, 386–411.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, in press.

Rabe-Hesketh, S., Toulopoulou, T., & Murray, R. (2001). Multilevel modeling of cognitive function in schizophrenic patients and their first degree relatives. *Multivariate Behavioral Research*, *36*, 279–298.

Rabe-Hesketh, S., Yang, S., & Pickles, A. (2001). Multilevel models for censored and latent responses. *Statistical Methods in Medical Research*, *10*, 409–427.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., & Lewis, T. (2000). *A User's Guide to MLwiN (second edition)*. London: Institute of Education, University of London.

Raudenbush, S.W. (1995). Maximum likelihood estimation for unbalanced multilevel covariance structure models via the EM algorithm. *British Journal of Mathematical and Statistical Psychology*, *48*, 359–370.

Raudenbush, S.W., Bryk, A.S., Cheong, Y. F., & Congdon, R. (2000). *HLM 5: Hierarchical Linear and Nonlinear Modeling*. Lincolnwood, IL: Scientific Software International.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Raudenbush, S.W., & Sampson, R. (1999a). Assessing direct and indirect effects in multilevel designs with latent variables. *Sociological Methods & Research*, *28*, 123–153.

Raudenbush, S.W., & Sampson, R. (1999b). Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. In P.V. Marsden (Ed.), *Sociological Methodology 1999*, vol. 29 (pp. 1–41). Oxford: Blackwell.

Rijmen, F., Tuerlinckx, F., De Boeck, P., and Kuppens, P. (2003). A nonlinear mixed model framework for IRT models. *Psychological Methods*, *8*, 185–205.

Rindskopf, D. (1984). Using phantom and imaginary latent variables to parameterize constraints in linear structural models. *Psychometrika*, *49*, 37–47.

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*, 563–576.

Skrondal, A., & Rabe-Hesketh, S. (2003a). Multilevel logistic regression for polytomous data and rankings. *Psychometrika*, *68*, 267–287.

Skrondal, A., & Rabe-Hesketh, S. (2003b). Some applications of generalized linear latent and mixed models in epidemiology: Repeated measures, measurement error and multilevel modeling. *Norwegian Journal of Epidemiology*, *13*, 265-278.

Skrondal, A., & Rabe-Hesketh, S. (2004a). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC.

Skrondal, A., & Rabe-Hesketh, S. (2004b). Generalized linear latent and mixed models with composite links and exploded likelihoods. In A. Biggeri, E. Dreassi, C. Lagazio & M. Marchi (Eds.), *Proceedings of the 19th International Workshop on Statistical Modeling* (pp. 27–39). Florence, Italy: Firenze University Press.

Skrondal, A., Rabe-Hesketh, S., & Pickles, A. (2002). Informative dropout and measurement error in cluster randomised trials. Paper presented at the International Biometric Society Conference 2002, Freiburg.

Social and Community Planning Research (1987). *British Social Attitudes Panel Survey, 1983-1986 [Computer file] SN: 2197*. Colchester, Essex: The Data Archive [Distributor].

Spiegelhalter, D., Thomas, A., Best, N., & Gilks, W. (1996). *BUGS 0.5 Bayesian Analysis using Gibbs sampling. Manual (version ii)*. Cambridge: MRC-Biostatistics Unit. Downloadable from `http://www.mrc-bsu.cam.ac.uk/bugs/documentation/contents.shtml`.

StataCorp. (2003). *Stata Statistical Software: Release 8.0*. College Station, TX: Stata Corporation.

Takane, Y. (1987). Analysis of covariance structures and probabilistic binary data. *Communication & Cognition*, *20*, 45–62.

Wiggins, R.D., Ashworth, K., O'Muircheartaigh, C.A., & Galbraith, J.I. (1990). Multilevel analysis of attitudes to abortion. *The Statistician*, *40*, 225–234.

Yang, S., & Pickles, A. (2004). Multilevel latent models for multivariate responses subject to measurement error. *Submitted for publication*.

Yang, S., Pickles, A., & Taylor, C. (1999). Multilevel latent variable model for analysing two-phase survey data. In H. Friedl, A. Berghold, & G. Kauermann (Eds.) *Proceedings of the 14th International Workshop on Statistical Modeling* (pp. 402–408). Graz, Austria.