Bayesian methods for latent trait modeling of longitudinal data

DAVID B. DUNSON

Biostatistics Branch, National Institute of Environmental Health Sciences

MD A3-03, P.O. Box 12233

Research Triangle Park, NC 27709, U.S.A.

dunson1@niehs.nih.gov

Summary

Latent trait models have long been used in the social science literature for studying variables that can only be measured indirectly through multiple items. However, such models are also very useful in accounting for correlation in multivariate and longitudinal data, particularly when outcomes have mixed measurement scales. Bayesian methods implemented with Markov chain Monte Carlo provide a flexible framework for routine fitting of a broad class of latent variable (LV) models, including very general structural equation models. However, in considering LV models, a number of challenging issues arise, including identifiability, confounding between the mean and variance, uncertainty in different aspects of the model, and difficulty in computation. Motivated by the problem of modeling multidimensional longitudinal data, this article reviews the recent literature, provides some recommendations, and highlights areas in need of additional research, focusing on methods for model uncertainty.

Keywords: Dynamic Dirichlet process; Factor analysis; Latent variables; Multidimensional longitudinal data; Random effects; Structural equation models

1 Introduction

In analyzing longitudinal data, it is clearly important to account for within-subject dependency in the repeated observations, and a wide variety of approaches have been considered (refer to Diggle et al., 2002 for a review). One of the most commonly-used is the random effects model, which allows the different subjects under study to have varying coefficients within a regression model, such as a linear or generalized linear model. Random effects are a type of latent variable (LV), and a broad class of LV models have been considered for longitudinal data analysis in the literature. This article provides a brief review of some of the recent literature.

Random effects models are most commonly considered in the setting in which a single outcome is measured repeatedly over time. For example, blood pressure measurements may be recorded for a patient at different visits to a clinic. One may then model these data using a linear mixed model (Laird and Ware, 1982) as follows:

$$y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{w}_{ij}'\mathbf{b}_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \tag{1}$$

where y_{ij} is blood pressure for patient i at visit j, $\mathbf{x}_{ij}$ is a $p \times 1$ vector of fixed effect predictors (e.g., age, body mass index, etc), $\boldsymbol{\beta}$ are fixed effect coefficients, $\mathbf{w}_{ij}$ is a $q \times 1$ vector of predictors having a varied effect (often chosen as a subset of $\mathbf{x}_{ij}$), and $\mathbf{b}_i \sim N_q(\mathbf{0}, \Omega)$ is a vector of random effects assumed to be normally distributed in the population with covariance Ω.

However, in many longitudinal studies, one does not simply collect a single outcome variable, such as blood pressure. Instead, many different patient characteristics are measured. For example, in a study of factors predictive of lung function, in addition to blood pressure, investigators may collect indicators of pulmonary symptoms experienced since the last clinic visit along with various tests of lung function. Such data are often referred to as multidimensional longitudinal data (Gray and Brookmeyer, 1998; 2000; Dunson, 2003). Although a common approach is to analyze the different outcomes separately, it is clearly appealing to consider joint models that account for dependency within and across time points. In addition to issues of efficiency and multiple testing, a clear substantive motivation for considering a multivariate approach is that investigators are seldom interested in one variable in isolation. For example, one seeks to study the impact of patient characteristics and clinical interventions on pulmonary function, which is a latent trait that is only imperfectly measured by any one variable or test.

One could argue compellingly that all biomedical studies involve latent traits, since one can almost never measure the variables of primary interest directly. Hence, it is appealing to consider statistical models that include latent traits, which then induce dependency in the different measured variables. Such latent trait models (LTMs) have long been a focus in the social science and psychometrics literature in which studies typically rely on questionnaire data or a battery of assessments to indirectly measure the variables of interest. If all the measured variables can be assumed to be normally distributed, then a linear structural relations (LISREL) model can potentially be used, and software is available for routine fitting (refer to Byrne, 1998 for a review). The LISREL framework can accommodate longitudinal data structures by allowing time-dependent latent traits.

Challenges arise in considering extensions of the LISREL model to allow nonlinear modeling structures, mixed scale measured variables, non-normal latent variables, model uncertainty and other complications. For example, in the lung function example, blood pressure, symptom indicators, and test data have different measurement scales, and it may be unclear whether it is appropriate to assume a single normally-distributed latent variable underlying these disparate outcomes.

Section 2 provides a review of normal and underlying normal latent variable models, including the LISREL model and generalizations encompassing LISREL and random effects models. Section 3 considers recent extensions, which allow a richer class of parametric distributions for the latent and measured outcomes. Section 4 describes recent work, which allows the component distributions to be unknown by using finite mixture models or nonparametric Bayesian methods. Section 5 discusses the issue of uncertainty in the model structure, including the number of latent variables and their relationship, and Section 6 contains a discussion. The focus throughout is on longitudinal data.

2 Normal and Underlying Normal Models

2.1 Linear Structural Equation Models

To introduce the concepts, I focus initially on a normal structural equation for data collected at a single time point. In particular, suppose that $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})'$ is a multivariate vector of measured response variables and $\mathbf{x}_i = (x_{i1}, \ldots, x_{iq})'$ is a multivariate vector of measured predictors. Then, a structural equation model (SEM) (Bollen, 1989) can be specified in two components: (1) the measurement model, which relates the manifest or measured variables to latent variables; and (2) the latent variable or structural relations model which describes associations among the latent variables. For a recent review of SEMs, with a focus on applications in environmental epidemiology, refer to Sanchez et al. (2005). For an overview of Bayesian approaches to SEMs, refer to Palomo, Dunson and Bollen (2006).

For a normal linear SEM, the measurement model has the form:

$$\begin{aligned} \mathbf{y}_i &= \boldsymbol{\mu}_y + \boldsymbol{\Lambda}_y \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i^y, \quad \boldsymbol{\epsilon}_i^y \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_y), \\ \mathbf{x}_i &= \boldsymbol{\mu}_x + \boldsymbol{\Lambda}_x \boldsymbol{\xi}_i + \boldsymbol{\epsilon}_i^x, \quad \boldsymbol{\epsilon}_i^x \sim N_q(\mathbf{0}, \boldsymbol{\Sigma}_x), \end{aligned} \tag{2}$$

where $\boldsymbol{\mu}_y, \boldsymbol{\mu}_x$ are vectors of intercepts, $\boldsymbol{\Lambda}_y$ is a $p \times r$ factor loadings matrix, $\boldsymbol{\eta}_i$ is an $r \times 1$ vector of latent response variables, $\boldsymbol{\Lambda}_x$ is a $q \times s$ factor loadings matrix, $\boldsymbol{\xi}_i$ is an $s \times 1$ vector of latent predictors, and $\boldsymbol{\epsilon}_i^y, \boldsymbol{\epsilon}_i^x$ are normally distributed residual vectors, with $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_x$ covariance matrices (typically chosen to be diagonal).

Thus, one obtains multivariate data $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')'$ which contain indirect information about the latent variables $\boldsymbol{\phi}_i = (\boldsymbol{\eta}_i', \boldsymbol{\xi}_i')'$. Depending on the application, the latent traits may be introduced primarily for convenience in characterizing dependency in the measured variables, or the measurements may be designed to capture different aspects of the latent traits one is interested in studying. To provide an example of the first case, in the Agricultural Health Study (AHS), interest focuses on assessing the relationship between pesticide exposure and neurological health. Given the many different pesticides and the high within-subject correlation, it is not feasible to apply a typical regression model strategy. Instead one can reduce dimensionality by assessing the relationship between latent variables underlying the pesticide measurements and latent variables underlying neurologic symptom data. However, the pesticide measurements were not designed as indicators of latent traits, instead the latent traits are used as a dimensionality reduction technique and to simplify flexible modeling of the covariance. A useful strategy for very high dimensional data is to use a Bayesian latent variable model, with a prior that favors a sparse covariance structure (Carvalho et al., 2005).

To provide an example of the second case, which is historically more common in social science applications, suppose that researchers are interested in studying the relationship between stress induced by life events and severity of depression. Then, ξ_i may represent a latent continuous standardized stress score which ranks individual i relative to the other individuals in the population. One can estimate ξ_i based on $\mathbf{x}_i$, the elements of which consist of the individual's response to different questions and tests measuring stress. Supposing that $\mathbf{y}_i$ consist of responses to items related to depression, η_i then provides a standardized depression score for individual i.

In such an example, the final goal is to assess relationships among the latent traits, η_i and ξ_i, which relies on a linear structural relations model:

$$\boldsymbol{\eta}_i = \mathbf{B}\boldsymbol{\eta}_i + \boldsymbol{\Gamma}\boldsymbol{\xi}_i + \boldsymbol{\delta}_i, \quad \boldsymbol{\delta}_i \sim N_r(\mathbf{0}, \boldsymbol{\Omega}), \tag{3}$$

where $\boldsymbol{\xi}_i \sim N_s(\mathbf{0}, \boldsymbol{\Psi})$, $\mathbf{B}$ is an $r \times r$ matrix with 0's on the diagonal characterizing relationships among the elements of $\boldsymbol{\eta}_i$, $\boldsymbol{\Gamma}$ is an $r \times s$ matrix describing effects of the latent variables $\boldsymbol{\xi}_i$ on $\boldsymbol{\eta}_i$, and $\boldsymbol{\delta}_i$ is a normally-distributed residual. In the simple case considered in the example, this expression would simplify to $\eta_i = \gamma \xi_i + \delta_i$, $\delta_i \sim N(0, \omega)$, with γ characterizing the effect of stress on depression.

In studies of latent traits, it is clear that expressions (2) and (3) provide a convenient and general framework for inferences on relationships among the latent traits. A nice feature is that the approach avoids arbitrary collapsing of the measured data $\mathbf{y}_i$ and $\mathbf{x}_i$ into pre-specified scores; say by averaging different tests or focusing on a single representative test. However, one must be careful to remember that $\boldsymbol{\eta}_i$ and $\boldsymbol{\xi}_i$ are not observed directly for any of the subjects, so that the interpretation of these latent traits relies entirely on the quality of the measured variables and the particular identifiability constraints made. Regarding identifiability, it is clear that one cannot estimate all the parameters in models (2) and (3) on the basis of the data alone. It is necessary to fix some of the elements of the factor loadings matrices $\boldsymbol{\Lambda}_y, \boldsymbol{\Lambda}_x$, variance components $\boldsymbol{\Sigma}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Omega}$ and structural parameters $\mathbf{B}, \boldsymbol{\Gamma}$ to ensure identifiability.

Detailed consideration of the issue of identifiability has been provided elsewhere for normal linear SEMs (refer to Bollen, 1989). I make only brief comments here. A typical constraint is to assume $\boldsymbol{\Sigma}_y, \boldsymbol{\Sigma}_x, \boldsymbol{\Omega}$ are diagonal. Then, one either restricts the diagonal elements of the factor loadings matrices to be equal to one, or fixes the variances of the latent variables to equal one. This sets the scale of the latent variables, which is intrinsically arbitrary. Finally, off-diagonal elements of the loadings matrices $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ need to be fixed, because when one marginalizes over the latent variable distributions only $\boldsymbol{\Lambda}_y \boldsymbol{\Lambda}_y'$ and $\boldsymbol{\Lambda}_x \boldsymbol{\Lambda}_x'$ appear in the likelihood. Without constraints, one can obtain an equivalent likelihood incorporating a rotation matrix - say by replacing $\boldsymbol{\Lambda}_y$ with $\boldsymbol{\Lambda}_y \mathbf{P}$, where $\mathbf{PP'} = \mathbf{I}_r$ is the identity matrix. Often, appropriate restrictions are suggested naturally by the application.

For example, if the first 5 elements of $\mathbf{y}_i$ measure η_{i1} but not η_{i2}, then one should set the first 5 elements of the 2nd column of $\mathbf{\Lambda}_y$ equal to 0. In cases in which the latent variables are introduced for convenience in characterizing dependency, one can instead set the upper triangular elements of $\mathbf{\Lambda}_y$ and $\mathbf{\Lambda}_x$ equal to zero as a default to ensure identifiability (Lopes and West, 2004).

2.2 Extensions

The previous subsection describes the usual specification of an SEM used in the social sciences. There are clearly many modifications and extensions that have been considered. Firstly, the predictors $\mathbf{x}_i$ are modeled as random and normally distributed. One can envision fixed predictors $\mathbf{w}_i$ by instead letting $\xi_{il} = w_{il}$ for a subset of the latent variables, with the residual variance $\omega_l \to 0$. For example, w_{il} may be dose of an environmental exposure measured in the blood, which one wants to relate to neurological functioning, which is a latent response. Alternatively, if one wants to adjust for fixed predictors that may affect the measurement of the different items but not the structural relationship among the latent variables, one can include a regression component in place of the intercept in expression (2).

In considering the relationship between expressions (2)-(3) and the linear mixed model (1), one may note that the SEM approach is more flexible in how it accounts for baseline dependency among the multiple outcomes. In particular, the SEM model includes unknown factor loadings multiplying the latent variables. This allows for a non-exchangeable correlation structure and differences in scale between the measured variables. This flexibility is clearly needed when the measured outcomes are not simply repeated measurements of the same variable, but instead have different interpretations and measurement scales. However, the linear mixed model structure has the advantage of allowing heterogeneity in regression coefficients among subjects.

To make this concept clear, it is useful to consider an example. In particular, suppose that interest focuses on assessing the effect of a pesticide on neurological functioning. Newborn mice are randomized to several dose groups and are given a functional observational battery of tests of motor activity. Animal i's response is denoted $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$, with higher values of y_{ij} indicating greater activity of type j. One could potentially fit a linear mixed effect model (1), with

$$y_{ij} = \beta_{j1} + \beta_{j2} dose_i + b_{i1} + b_{i2} dose_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \tag{4}$$

with β_{j1}, β_{j2} a test-specific intercept and slope and b_{i1}, b_{i2} a random intercept and slope. Here, b_{i1} allows for baseline heterogeneity among animals in motor activity, which induces dependency in the elements of $\mathbf{y}_i$, while b_{i2} allows heterogeneity in the dose effect.

These same data could be analyzed with an SEM. For example, let

$$\begin{aligned} y_{ij} &= \mu_j + \lambda_j \eta_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_j^2) \\ \eta_i &= \gamma_1 dose_i + \delta_i, \quad \delta_i \sim N(0, 1), \end{aligned} \tag{5}$$

where $\lambda_j \geq 0$ and we consider the predictors as fixed. Here, we incorporate dose on the latent variable level to reduce dimensionality in assessing the effect of dose on motor activity. This dimensionality reduction is one advantage of the SEM. Another is flexibility in characterizing the covariance. In particular, note that we allow for differences in scale between the test items through an item-specific factor loading, λ_j, and residual variance, σ_j^2. The resulting correlation between y_{ij} and $y_{ij'}$ is

$$\text{corr}(y_{ij}, y_{ij'}) = \frac{\lambda_j \lambda_{j'}}{(\lambda_j^2 + \sigma_j^2)^{1/2} (\lambda_{j'}^2 + \sigma_{j'}^2)^{1/2}}, \tag{6}$$

which can differ for different combinations of tests, j and j'. In contrast, the correlation under model (4) is

$$\text{corr}(y_{ij}, y_{ij'}) = \frac{\mathbf{w}_{ij}' \boldsymbol{\Omega} \mathbf{w}_{ij'}}{(\mathbf{w}_{ij}' \boldsymbol{\Omega} \mathbf{w}_{ij} + \sigma^2)^{1/2} (\mathbf{w}_{ij'}' \boldsymbol{\Omega} \mathbf{w}_{ij'} + \sigma^2)^{1/2}}, \tag{7}$$

where $\mathbf{w}_{ij} = (1, dose_i)'$. In comparison with (6), the mixed model-based correlation (7) is more flexible in allowing the correlation to vary with dose, which is often biologically-motivated in toxicology studies. However, for animals in the control group, (7) simplifies to $\text{corr}(y_{ij}, y_{ij'}) = \omega_{11}/(\omega_{11} + \sigma^2)$, so that the observations are treated as exchangeable. Although residual dependency could be incorporated to relax this restrictive assumption, the SEM induces a more flexible baseline dependency structure without such complications.

It is interesting to consider combined models that share the beneficial features of the SEM and linear mixed modeling frameworks. In general, this can be accomplished by replacing $\mathbf{\Lambda}_y\boldsymbol{\eta}_i$ and $\mathbf{\Lambda}_x\boldsymbol{\xi}_i$ in expression (2) with $\mathbf{\Lambda}_y\mathbf{W}_i^y\boldsymbol{\eta}_i$ and $\mathbf{\Lambda}_x\mathbf{W}_i^x\boldsymbol{\xi}_i$, with $\mathbf{W}_i^y, \mathbf{W}_i^x$ diagonal matrices having elements equal to known predictor values, which play the role of the $\mathbf{w}_{ij}$ vector in expression (1). For example, in the rat neurological health example, one could let

$$\begin{aligned} y_{ij} &= \mu_j + \lambda_{j1}\eta_{i1} + \lambda_{j2} dose_i \eta_{i2} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma_j^2) \\ \eta_{i1} &= \gamma_1 dose_i + \delta_{i1}, \quad \delta_{i1} \sim N(0,1), \\ \eta_{i2} &\sim \text{N}(0,1), \end{aligned} \tag{8}$$

where the second latent variable, η_{i2}, is introduced to allow heterogeneity among mice in the dose effect. This combination of SEMs and linear mixed effects models remains to be fully evaluated, though we have obtained promising results in preliminary simulation studies using a Bayesian approach implemented with Gibbs sampling (joint with Jesus Palomo). Alternative approaches for allowing differences in scale in modeling multiple continuous outcomes have been considered by Lin et al. (2000) and Roy, Lin and Ryan (2003).

A clear limitation of measurement model (2) is the reliance on a normal linear model, which requires the different measured items to be continuous and normally distributed. Fortunately, it is straightforward to modify the model to allow mixed ordered categorical and continuous items, which is typically what one encounters in practice. In particular, following the suggestion of Muthén (1984), one replaces $\mathbf{y}_i$ and $\mathbf{x}_i$ in expression (2) with $\mathbf{y}_i^*$ and

$\mathbf{x}_i^*$, respectively, with the $*$ superscript denoting that the variables are *underlying variables* which are only observed directly in certain special cases. Then, one links the observed variables to these underlying variables using $z_{ik} = g_j(z_{ik}^*; \boldsymbol{\tau}_k)$, with $g_j(\cdot)$ the identity link for continuous z_{ik} and a threshold link mapping from $\Re \rightarrow [1, \ldots, c_j]$ for ordered categorical z_{ik} with c_j categories. The $\boldsymbol{\tau}_k$ parameters represent unknown thresholds. This specification results in a sort of multivariate probit model, which is conceptually-related to approaches proposed for joint modeling of categorical and continuous outcomes (Catalano and Ryan, 1992; Gueorguieva and Agresti, 2001; Dunson, Chen and Harry, 2003).

2.3 Longitudinal Data

Until this point, I have focused on the case in which all the data are collected at a single time point. However, the SEM modeling framework is flexible enough to accommodate longitudinal data. In particular, suppose that data $\mathbf{z}_{ij} = (\mathbf{y}_{ij}', \mathbf{x}_{ij}')'$ are collected for subject i ($i = 1, \ldots, n$) at n_i follow-up times, with $\mathbf{y}_{ij}$ denoting a $p \times 1$ vector of measured response variables and $\mathbf{x}_{ij}$ denoting a $q \times 1$ vector of measured predictors. Then, one could simply let $\mathbf{y}_i = (\mathbf{y}_{i1}', \ldots, \mathbf{y}_{i,n_i}')'$ and $\mathbf{x}_i = (\mathbf{x}_{i1}', \ldots, \mathbf{x}_{i,n_i})'$ and define an SEM according to expressions (2) and (3).

However, because the SEM framework is so broad, it is useful to consider a simplified structure, which is natural for longitudinal data. In particular, one can simply add a j subscript to $\mathbf{y}_i, \boldsymbol{\eta}_i, \boldsymbol{\epsilon}_i^y$ and $\mathbf{x}_i, \boldsymbol{\xi}_i, \boldsymbol{\epsilon}_i^x$ in expression (2), possibly also replacing $\boldsymbol{\mu}_y, \boldsymbol{\mu}_x$ with a regression term having time-dependent covariates impacting measurement of the latent variables. One may also consider incorporating matrices of multipliers $\mathbf{W}_{ij}^y, \mathbf{W}_{ij}^x$ in the factor loading terms, but we focus on the simpler case. Then, it remains to specify a dynamic linear structural relations model. In particular, one needs to allow correlation among the latent variables at the current and previous time points. Following Dunson (2003), to avoid overfitting, it may be useful to consider a simple structure in which the latent variable η_{ijl} depends only on the other latent variables at time point j and on $\eta_{i,j-1,l}$.

To make this approach more concrete, let's revisit the stress and depression example described in Section 2.1. However, now suppose that the study enrolled patients with a history of depression to determine how life events-induced stress impacts depression. At baseline, the patients are each given a depression assessment and asked to complete a detailed questionnaire on life events-induced stress in the previous six months. The patients are then followed for 3 years, with the assessments repeated every six months.

For simplicity, suppose that there is a single time-varying depression severity variable η_{ij} and a single time-varying life events-induced stress severity variable ξ_{ij}. These variables are measured through the different questionnaire and test items, which are a mixture of categorical and continuous variables. However, using the underlying normal framework described above, we can consider the measurement model:

$$\begin{aligned} y^*_{ijk} &= \mu_{y,k} + \lambda_{y,k}\eta_{ij} + \epsilon_{y,ijk}, \quad \epsilon_{y,ijk} \sim N(0,\sigma^2_{y,k}),\ k = 1,\ldots,p, \\ x^*_{ijl} &= \mu_{x,l} + \lambda_{x,l}\xi_{ij} + \epsilon_{x,ijl}, \quad \epsilon_{x,ijl} \sim N(0,\sigma^2_{x,l}),\ l = 1,\ldots,q, \end{aligned} \tag{9}$$

with $\boldsymbol{\lambda}_y = (\lambda_{y,1},\ldots,\lambda_{y,p})'$ and $\boldsymbol{\lambda}_x = (\lambda_{x,1},\ldots,\lambda_{x,p})'$ vectors of factor loadings having one or more elements restricted to be positive for identifiability and to define the direction of η_{ij} and ξ_{ij}, respectively. In addition, for categorical measured items, the residual error variances are fixed at one for identifiability. We assume that the accuracy of the test items in measuring the latent variables does not vary over time, so that we can use a fixed intercept, factor loadings and residual variance.

Note that model (9) assumes that the correlation in the different measured items within a time point and across time points is explained entirely by the two time-varying latent traits. In many applications, one may be concerned that an individual may tend to respond high (or low) to certain items across time due to item-specific factors (e.g., the individual interprets the question wording in a different manner than other subjects). For this reason, one may want to consider item-specific random effects as described by Dunson (2003). I do not consider that approach further here.

It remains to specify the latent variable model. A simple model for this application is

$$\begin{aligned} \eta_{ij} &= \mathbf{u}_{ij}'\boldsymbol{\beta} + \kappa\eta_{i,j-1} + \gamma_j\xi_{ij} + \delta_{ij}, \quad \delta_{ij} \sim N(0,1), \\ \xi_{ij} &= \mathbf{v}_{ij}'\boldsymbol{\alpha} + \nu\xi_{i,j-1} + \zeta_{ij}, \quad \zeta_{ij} \sim N(0,1), \end{aligned} \tag{10}$$

where $\mathbf{u}_{ij}$ and $\mathbf{v}_{ij}$ are vectors of known predictors, including treatment, demographics, gender and patient history. In addition, κ and ν are autocorrelation parameters allowing dependency within a latent trait over time. For example, the depression score for an individual should correlate over time. Finally, the parameters of most interest are $\{\gamma_j\}$, which characterize the effect of life events-induced stress on severity of depression, adjusting for patient characteristics and previous depression score. Potentially, one could consider a model with $\gamma_j = \gamma$ for parsimony and simplicity in interpretation.

The specification in (10) uses a discrete-time formulation, which is often appropriate in longitudinal studies that collect data at regular follow-up times. The discrete-time assumption greatly simplifies flexible model fitting and inferences without restrictive parametric assumptions about the trajectories with time. However, an interesting area of ongoing research would replace the latent variables and residuals by time-varying stochastic processes. For example, one could let $\eta_i(t)$, $\xi_i(t)$, $\delta_i(t)$, and $\zeta_i(t)$ be continuous time random functions, which are assigned a Gaussian process (refer to Rasmussen and Williams, 2006, for a recent book on Gaussian processes). This has the advantage of allowing the subjects to have varying observation times without needing to collapse the data into bins. For a recent application of Bayesian latent variable modeling of longitudinal data, refer to Daniels and Normand (2006). Roy and Lin (2000) and Muthén et al. (2002) propose alternative latent variable models for longitudinal data.

2.4 Joint Modeling

It is interesting to note that the framework described above can easily be modified to allow joint modeling of multivariate longitudinal data with an event time, allowing for informatively missing data, as described by Dunson and Perreault (2001) and Dunson, Chen and Harry (2003). First consider the case in which subjects may have informatively missing observations in a longitudinal study. For example, in studying the relationship between life events and depression, it is necessary for the patient to show up for their regularly scheduled clinic visit in order for life events stress and depression to be measured at that time. Patients with severe depression may be more likely to miss visits, leading to problems with informative missingness. To deal with this type of problem, one can introduce a time-specific missingness indicator, with the probability of this indicator being equal to one dependent on the individual's time-dependent latent traits through a probit regression model.

This allows for informative missingness through a shared parameter approach conceptually related to widely-used shared random effects models (Follmann and Wu, 1995). However, naive use of the shared random effects approach can lead to misleading results (Fieuws and Verbeke, 2004). A more robust approach is to include both outcome- and missingness-specific latent variables, which are correlated (Dunson and Perreault, 2001). Potentially, one can also allow the latent variable distributions to be unknown, as discussed in Section 4. Roy and Lin (2002) applied a parametric latent variable analysis to account for nonignorable dropout and missing covariates in a study collecting multivariate longitudinal data. An alternative latent class pattern mixture modeling approach for informatively intermittent missing data was proposed by Lin, McCulloch and Rosenheck (2004).

Note that this same type of approach can be used for joint modeling of longitudinal and event time data within a discrete-time framework. One can simply introduce repeated binary indicators of whether failure has yet occurred and model these indicators as another measurement of the latent traits. By allowing for a time-varying intercept in these models, one obtains a flexible model for the event time. Such an approach can easily accommodate intermittent missingness, informative censoring, and joint modeling within one framework.

A related area is the use of latent variables in modeling of progressive disease processes. For many diseases, it is useful to consider a multistate model in which an individual progresses stochastically through latent health states over time, with the transition times between states unknown. For example, for chronic health conditions, such as uterine fibroids, one can consider a three state model in which an individual progresses between (1) no disease, (2) preclinical disease, and (3) clinical disease. As it is not known whether an individual is in state 1 or 2 prior to screening, the presence of disease represents a time-varying latent class variable. Based on cross-sectional screening data, one can potentially obtain the current status for each individual under study, along with several indicators of disease severity. Dunson and Baird (2002) proposed to use these disease severity indicators as measurements of a severity latent variable, which progresses stochastically based on waiting time in state 2. Such models are useful for inferences on incidence and progression based on cross-sectional data.

2.5 Model Fitting and Inferences

Before considering alternative models and methods, it is important to discuss the important issue of model fitting and inferences. Most of the vast literature on this problem has focused on frequentist approaches, and there are a number of software packages available for routine fitting, which can accommodate the models considered in Sections 2.1-2.3. Options are more limited in the case in which there are mixed polytomous and continuous data, with two stage least square procedures often used, as maximum likelihood (ML) estimation can be difficult due to the high-dimensional integration involved. To solve this problem, Shi and Lee (2000) developed an Monte Carlo-EM algorithm, which has since been generalized to accommodate more complex cases involving multi-level data and nonlinear latent variable structures (Lee and Song, 2005).

As noted in a recent review of Bayesian SEMs by Palomo et al (2006), the Bayesian approach has a number of advantages. First, Markov chain Monte Carlo (MCMC) algorithms can be used to estimate exact posterior distributions of the parameters and latent variables, while ML estimation only produces a point estimate of the parameters, with asymptotic standard errors. Often, even in moderate to large sample sizes, there may be quite limited information available about certain parameters, so asymptotic assumptions may not be justified. Lee and Song (2004) demonstrated better performance of a Bayesian approach in small samples compared with ML estimation. In addition, it is very useful to have posterior distributions of the latent variables available. Estimation of the latent variables is of interest in many applications, and one can use latent variable residual plots to assess goodness of fit (Albert and Chib, 1995; Gelman et al., 2005). An additional advantage is the ability to incorporate prior information, which is particularly important in SEMs given the many constraints that are necessary. Finally, the flexibility of the Bayesian approach implemented with MCMC permits an easier extension to deal with more complex models, such as the cases considered in the subsequent sections of this article. Perhaps most importantly, given the typical uncertainty one has in specifying the model, is the ability to use Bayesian methods to select and average across plausible models in performing inferences and predictions.

For the models considered in Sections 2.1-2.3, posterior computation can proceed via a straightforward Gibbs sampling algorithm, relying on the data augmentation trick of Albert and Chib (1993) in the case in which measured variables have mixed polytomous and continuous scales. Refer to Arminger (1998), Scheines, Hoijtink and Boomsma (1999) and Palomo et al. (2006) for descriptions of the Gibbs sampler for SEMs. When one chooses conditionally-conjugate normal or truncated normal priors for the free intercepts, factor loadings, and structural parameters and inverse-gamma priors for the variances, the resulting full conditional posterior distributions have simple forms. However, care should be taken to avoid choosing diffuse priors, since an improper posterior can result in the limiting case as the prior variance goes to ∞. In addition, as in other hierarchical models, diffuse but proper priors can lead to poor computational performance. Gelman (2006) recently addressed this issue eloquently in the context of variance component models, and his recommendations can be applied also to SEMs. Lee and Press (1998) study robustness in Bayesian factor analysis with respect to changes in the hyperparameters.

3 Generalized Latent Variable Models

3.1 Brief Overview

Although the underlying normal SEM framework described in the previous Section provides a useful approach for accommodating a very broad class of data structures, with mixed categorical and continuous variables observed over time, one is limited in flexibility due to the hierarchical normal linear structure. A number of authors have proposed approaches that allow parametric non-linear regression structures (Arminger, 1998; Lee and Zhu, 2000, 2002, among others). There is also an expanding literature on LV models, which allow measured and latent variables to have a much broader class of distributions.

Sammel, Ryan and Legler (1997) proposed an approach, which characterized each of the measured outcomes with a generalized linear model, with the different outcomes modelled jointly through the introduction of normal latent variables common to the different models. Moustaki and Knott (2000) extended this framework to a class of generalized latent trait models. Both of these articles relied on the EM algorithm for model fitting, and computational hurdles arise as the number of latent variables increases. One of the primary difficulties is in marginalizing over the latent variable distributions when outcomes have a mixture of distributions in the exponential family. Although standard approximations can be used, the accuracy may decrease with latent variable dimension.

Following a Bayesian approach with MCMC used for computation, Dunson (2000) proposed a broader class of models, which allows observed, underlying and latent variables to have distributions in the exponential family, while also accommodating multilevel data structures. Dunson (2003) applied a similar framework in the context of dynamic modeling of multivariate longitudinal data.

Extending the multivariate spatial latent variable model of Wang and Wall (2003), Zhu, Eickhoff and Yan (2005) proposed a class of generalized linear latent variable models for repeated measurements of spatially correlated multivariate data. A Monte Carlo EM algorithm was used for estimation, applying a novel method for accelerating convergence. Liu, Wall and Hodges (2005) proposed an alternative generalized spatial structural equation model using a Bayesian approach implemented with MCMC. Hogan and Tchernis (2004) applied Bayesian latent factor analysis for spatially correlated data to a material deprivation application.

3.2 Some Dangers

There are some potential problems and pitfalls that can arise in fitting generalized latent variable models. As an illustration, consider the simple case in which p different types of binary indicators, $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})$, are measured at a single time point. For example, $\mathbf{y}_i$ may consist of indicators of the occurrence of tumors in different organ sites in an animal bioassay study in which the exposure dose for animal i is denoted x_i. Then, it is natural to consider the following model:

$$\begin{aligned} \Pr(y_{ij} = 1 \,|\, x_i, \eta_i) &= \frac{\exp(\alpha_j + \lambda_j \eta_i)}{1 + \exp(\beta_0 + \lambda_j \eta_i)} \\ \eta_i &= \beta x_i + \delta_i, \quad \delta_i \sim N(0,1), \end{aligned} \tag{11}$$

where $\lambda_j \geq 0$ for identifiability. Similar models were considered by Sammel et al. (1997).

In expression (11) one may interpret η_i as a tumor score for animal i, with animals having a higher score more likely to develop tumors in multiple organ sites. By including dose at the latent variable level, one reduces dimensionality in assessing the overall effect of dose on the risk of developing tumors. The incorporation of a single slope parameter, β, also simplifies interpretation and aids in quantitative risk assessment. However, it is important to note that the λ_j factor loadings parameters control both within-animal dependency in the occurrence of tumors in different sites and site-specific dose effects. Hence, if there is low within-animal correlation but a large effect of dose, one may underestimate the dose effect and overestimate the correlation. One may also obtain misleading results if the correlation changes with dose, which is a common occurrence in toxicologic studies.

The fact that the same parameters control the mean and variance is also true when considering normal or underlying normal latent factor regression models. However, in generalized latent variable models, one faces an additional problem due to the fact that one is effectively modeling the response distribution as a Gaussian mixture of exponential family distributions. If the exponential family distribution is Gaussian, the mixture is Gaussian. However, for other distributions in the exponential family, this is not the case and one obtains a different distributional form in marginalizing over the latent variable distributions. Hence, inferences are potentially more sensitive to lack of fit then in the normal linear case. Dunson and Herring (2005) discuss this problem in the context of modeling mixed discrete outcome data using a novel underlying Poisson log-linear framework. For multivariate counts, they note that the latent variables account for both over-dispersion relative to the Poisson distribution and dependency among the multiple outcomes. For this reason, one must be careful to introduce latent variables having both roles to avoid bias. This is also the case for models in the Sammel et al. (1997), Moustaki and Knott (2000) and Dunson (2000, 2003) frameworks.

4 Semiparametric Methods

A concern limiting the more widespread use of latent variable models outside of traditional social science applications is the sensitivity to parametric assumptions. In latent variable models, such assumptions are often difficult to check and there is an absence of simple approaches for dealing with model violations. For example, one cannot simply apply a post-hoc transformation of a latent variable, so that the normality assumption is more appropriate. For this reason, the development of semiparametric methods is a very interesting area in need of additional development.

Although technically parametric, one promising approach is to use a finite mixture of latent variable models. For multivariate continuous data, finite mixtures of factor analysis models, deemed mixtures of factor analyzers (MFA), have been considered by several authors (Utsugi and Kumagai, 2001; Fokoue and Titterington, 2003). McLachlan, Peel and Bean (2003) applied MFA as a flexible approach to model-based density estimation and clustering of high-dimensional data. In addition, Zhu and Lee (2001) proposed a Bayesian approach to finite mixtures of linear structural relations models.

A disadvantage of finite mixture models is that one may be sensitive to the choice of the number of mixture components, k, and standard methods for model selection (AIC, BIC, etc) are not appropriate tools for choosing k (Richardson and Green, 1997). Hence, an appealing alternative is to consider a semiparametric approach. For example, following a Bayesian perspective, one could allow a latent variable distribution, G, to be unknown through the use of a Dirichlet process (DP) prior (Ferguson, 1973, 1974). Such an approach has been applied to allow an unknown random effects distribution by Bush and MacEachern (1996), Kleinman and Ibrahim (1998) and Ishwaran and Takahara (2002) among others.

As mentioned previously, random effects models may be insufficiently flexible when measured outcomes have different scales. In addition, in longitudinal data applications, it may not be appropriate to assume that random effects do not vary over time. To allow an unknown latent trait distribution to vary dynamically with time, Dunson (2006) proposed a dynamic mixture of Dirichlet processes. This approach can easily accommodate multiple latent variables. In the setting of multiple event time analysis, Pennell and Dunson (2006) proposed a dynamic nonparametric frailty model. Dunson, Watson and Taylor (2003) proposed an alternative strategy, which models the residuals nonparametrically instead of the latent variable distribution.

An alternative direction, which has been considered by a number of authors, is to avoid a full likelihood specification. For example, Reboussin and Liang (1998) proposed an estimating equations approach for the LISCOMP model, which is a general framework for structural equation modeling of mixed categorical and continuous data. Sammel and Ryan (2002) developed a robust score test of exposure effects in a latent variable model for multiple continuous outcomes. There has also been a considerable amount of work on robust frequentist estimation. For example, Yuan, Bentler and Chan (2004) propose a robust procedure for structural equation modeling with heavy tailed distributions.

5 Model Uncertainty

Another hurdle to routine use of latent variable models is the presence of uncertainty in the number of latent variables and their structural relationship. In particular, in many (or most) applications, there are many competing latent variable models that are consistent with prior knowledge. For this reason, it would be appealing to be able to select better models from amongst these models. In considering more than a few candidate models, there may be many models that are difficult to distinguish based on the data and on prior knowledge, so it is not recommended to base inferences on a single selected model.

A useful approach is instead to estimate posterior model probabilities for each of the competing models (Press and Shigemasu, 1999). These posterior probabilities could be used as a basis for inferences and as weights for model averaging. Lopes and West (2004) propose a Bayesian approach to allow uncertainty in the number of factors in factor analysis, relying on reversible jump MCMC (Green, 1995) for posterior computation. Unfortunately, this approach is quite computationally intensive, requiring one to run an MCMC analysis for each model in the list prior to implementing the reversible jump algorithm. A more efficient approach to handle factor analysis model selection in high dimensions was recently proposed by Carvalho et al. (2005). Unfortunately, such approaches are difficult to extend to more general structural equation models.

In the setting of generalized linear mixed models with uncertainty in the predictors having random effects, Cai and Dunson (2006) recently proposed a Bayesian model averaging approach, relying on a stochastic search variable selection algorithm. Lee and Song (2003) used path sampling (Gelman and Meng, 1998) to calculate Bayes factors for comparing nonlinear structural equation models. There is a need for innovative approaches, which allow model selection and averaging in applications involving large numbers of competing models; a common scenario in latent variable modeling. In addition, as approaches based on the Bayes factor have well-known sensitivity to the prior, it is important to develop default methods that avoid reliance on subjectively-chosen hyperparameters.

6 Discussion

Latent variable models are extremely useful as a general modeling strategy that can be applied to an extremely broad class of problems. This article has provided a brief overview of some recent work and ongoing areas of interest in latent variable modeling, focused on longitudinal data applications. This is not meant to be a comprehensive overview, so I apologize for excluding a lot of very interesting work in this area.

With the promise of latent variable models comes a number of challenges, which need to be carefully addressed when considering applications. First, one needs to think carefully about whether the model is identified. This can be accomplished by marginalizing over the latent variable distributions to obtain a marginal likelihood, available in closed form for normal hierarchical latent variable models. In examining this marginal likelihood, one can assess whether parameters are redundant and what features of the data inform about the parameter values. For non-normal models, the normal model may provide a useful reference. Second, one should not assume that inferences are robust to the assumed specification, including the model structure and (for Bayesians) the prior. Although formal model averaging would be preferred, at the very least, one should repeat the analysis under a variety of reasonable alternative models and priors to judge sensitivity.

These problems will be reduced as improvements in the available methodology continue. Some of the areas in greatest need of additional research include the development of (1) routine approaches for model selection and averaging, made challenging by constraints on the parameters; (2) methods for fitting of latent variable models in large sample sizes; (3) more flexible modeling frameworks, allowing unknown latent variable distributions and regression structures; (4) automated approaches for judging model identifiability; and (5) default or noninformative priors that permit routine fitting of Bayesian methods without reliance on subjective choices.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, and NIEHS. The author thanks Michael Pennell and Gregg Dinse for their helpful comments.

References

Albert, J.H. and Chib, S. (1993). Bayesian-analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.

Albert, J. and Chib, S. (1995). Bayesian residual analysis for binary response regression models. *Biometrika* **82**, 747-759.

Arminger, G. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika* **63**, 271-300.

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

Bush, C.A. and MacEachern, S.N. (1996). A semiparametric Bayesian model for randomised block designs. *Biometrika* **83**, 275-285.

Byrne, B.M. (1998). *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, Applications and Programming.* Lawrence Erlbaum Associates.

Cai, B. and Dunson, D.B. (2006). Bayesian covariance selection in generalized linear mixed models. *Biometrics*, in press.

Carvalho, C., Lucas, J., Wang, Q., Nevins, J. and West, M. (2005). High-dimensional sparse factor models and latent factor regression. *ISDS Discussion Paper* **2005-15**, Duke University, Durham, NC.

Catalano, P.J. and Ryan, L.M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* **87**, 651-658.

Chen, Z. and Dunson, D.B. (2003). Random effects selection in linear mixed models. *Biometrics* **59**, 762-769.

Daniels, M.J. and Normand, S-L.T. (2006). Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics* **7**, 1-15.

Diggle, P.J., Heagerty, P., Liang, K.Y. and Zeger, S.L. (2002). *The Analysis of Longitudinal Data*. Oxford, England: 2nd Edition. Oxford University Press.

Dunson, D.B. (2000). Bayesian latent variable models for clustered mixed outcomes. *Journal of the Royal Statistical Society B* **62**, 355-366.

Dunson, D.B. (2003). Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association* **98**, 555-563.

Dunson, D.B. (2006). Bayesian dynamic modeling of latent trait distributions. *Biostatistics*, in press.

Dunson, D.B. and Baird, D.D. (2002). Bayesian modeling of incidence and progression of disease from cross-sectional data. *Biometrics* **58**, 813-822.

Dunson, D.B., Chen, Z. and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59**, 521-530.

Dunson, D.B. and Herring, A.H. (2005). Bayesian latent variable models for mixed discrete outcomes. *Biostatistics* **6**, 11-25.

Dunson, D.B. and Perrealt, S.D. (2001). Factor analytic models of clustered multivariate data with informative censoring. *Biometrics* **57**, 302-308.

Dunson, D.B., Watson, M., and Taylor, J.A. (2003). Bayesian latent variable models for median regression on multiple outcomes. *Biometrics* **59**, 296-304.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Annals of Statistics* **2**, 615-629.

Fieuws, S. and Verbeke, G. (2004). Joint modelling of multivariate longitudinal profiles: pitfalls of the random-effects approach. *Statistics in Medicine* **23**, 3093-3104.

Fokoue, E. and Titterington, D.M. (2003). Mixtures of factor analysers: Bayesian estimation and inference by stochastic simultion. *Machine Learning* **50**, 73-94.

Follmann, D. and Wu, M. (1995). An approximate generalized linear-model with random effects for informatively missing data. *Biometrics* **51**, 151-168.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, to appear.

Gelman, A. and Meng, X.L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* **13**, 163-185.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D.F. and Meulders, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* **61**, 74-85.

Gray, S.M. and Brookmeyer, R. (1998). Estimating a treatment effect from multidimensional longitudinal data. *Biometrics* **54**, 976-988.

Gray, S.M. and Brookmeyer, R. (2000). Multidimensional longitudinal data: Estimating a treatment effect from continuous, discrete or time-to-event response variables. *Journal of the American Statistical Association* **95**, 396-406.

Green, P.J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.

Gueorguieva, R.V. and Agresti, A. (2001). A correlated probit model for joint modeling of clustered binary and continuous responses. *Journal of the American Statistical Association* **96**, 1102-1112.

Hogan, J.W. and Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-level material deprivation from census data. *Journal of the American Statistical Association* **99**, 314-324.

Ishwaran, H. and Takahara, G. (2002). Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models. *Journal of the American Statistical Association* **97**, 1154-1166.

Kleinman, K.P. and Ibrahim, J.G. (1998). A semiparametric Bayesian approach to the random effects model. *Biometrics* **54**, 921-938.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Lee, S.E. and Press, S.J. (1998). Robustness of Bayesian factor analysis estimates. *Communications in Statistics - Theory and Methods* **27**, 1871-1893.

Lee, S.Y. and Song, X.Y. (2003). Model comparison of nonlinear structural equation models with fixed covariates. *Psychometrika* **68**, 27-47.

Lee, S.Y. and Song, X.Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research* **39**, 653-686.

Lee, S.Y. and Song, X.Y. (2005). Maximum likelihood analysis of a two-level nonlinear structural equation model with fixed covariates. *Journal of Educational and Behavioral Statistics* **30**, 1-26.

Lee, S.Y. and Zhu, H.T. (2000). Statistical analysis of nonlinear structural equation models with continuous and polytomous data. *British Journal of Mathematical & Statistical Psychology* **53**, 209-232.

Lee, S.Y. and Zhu, H.T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika* **67**, 189-210.

Lin, H.Q., McCulloch, C.E. and Rosenheck, R.A. (2004). Latent pattern mixture models for informative intermittent missing data in longitudinal studies. *Biometrics* **60**, 295-305.

Lin, X., Ryan, L., Sammel, M., Zhang, D., Padungtod, C. and Xu, X. (2000). A scaled linear mixed model for multiple outcomes. *Biometrics* **56**, 593-601.

Liu, X., Wall, M.M. and Hodges, J.S. (2005). Generalized spatial structural equation models. *Biostatistics* **6**, 539-557.

Lopes, H.F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41-67.

McLachlan, G.J., Peel, D. and Bean, R.W. (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statistics & Data Analysis* **41**, 379-388.

Moustaki, I. and Knott, M. (2000). Generalized latent trait models. *Psychometrika* **65**, 391-411.

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

Muthén, B., Brown, C.H., Masyn, K., Jo, B., Khoo, S.T., Yang, C.C., Wang, C.P., Kellam, S.G., Carlin, J.B. and Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459-475.

Palomo, J., Dunson, D.B. and Bollen, K.A. (2006). Bayesian structural equation modeling. *Handbook on Structural Equation Models*, in press.

Pennell, M.L. and Dunson, D.B. (2006). Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*, in press.

Press, S.J. and Shigemasu, K. (1999). A note on choosing the number of factors. *Communications in Statistics - Theory and Methods* **28**, 1653-1670.

Rasmussen, C.E. and Williams, C.K.I. (2006). Gaussian Processes for Machine Learning. London, UK: The MIT Press.

Reboussin, B.A. and Liang, K.Y. (1998). An estimating equations approach for the LISCOMP model. *Psychometrika* **63**, 165-182.

Reboussin, B.A., Liang, K.Y., and Reboussin, D.M. (1999). Estimating equations for a latent transition model with multiple discrete indicators. *Biometrics* **55**, 839-845.

Richardson, S. and Green, P.J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society B* **59**, 731-758.

Roy, J. and Lin, X.H. (2000). Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics* **56**, 1047-1054.

Roy, J. and Lin, X.H. (2002). Analysis of multivariate longitudinal outcomes with nonignorable dropout and missing covariates: Changes in methadone treatment practices. *Journal of the American Statistical Association* **97**, 40-52.

Roy, J., Lin, X.H., and Ryan, L.M. (2003). Scaled marginal models for multiple continuous outcomes. *Biostatistics* **4**, 371-383.

Sammel, M.D. and Ryan, L.M. (2002). Effects of covariance misspecification in a latent variable model for multiple outcomes. *Statistica Sinica* **12**, 1207-1222.

Sammel, M.D., Ryan, L.M., and Legler, J.M. (1997). Latent variable models for mixed discrete and continuous outcomes. *Journal of the Royal Statistical Society B* **59**, 667-678.

Scheines, R., Hoijtink, H. and Boomsma, A. (1999). Bayesian estimation and testing of structural equation models. *Psychometrika* **64**, 37-52.

Sanchez, B.N., Budtz-Jorgensen, E., Ryan, L.M. and Hu, H. (2005). Structural equation models: A review with applications to environmental epidemiology. *Journal of the American Statistical Association* **100**, 1443-1455.

Shi, J.Q. and Lee, S.Y. (2000). Latent variable models with mixed continuous and polytomous data. *Journal of the Royal Statistical Society B* **62**, 77-87.

Utsugi, A. and Kumagai, T. (2001). Bayesian analysis of mixtures of factor analyzers. *Neural Computation* **13**, 993-1002.

Wang, F.J. and Wall, M.M. (2003). Generalized common spatial factor model. *Biostatistics* **4**, 569-582.

Yuan, K.H., Bentler, P.M. and Chan, W. (2004). Structural equation modeling with heavy tailed distributions. *Psychometrika* **69**, 421-436.

Zhu, J., Eickhoff, J.C. and Yan, P. (2005). Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics* **61**, 674-683.

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