MULTILEVEL AND LATENT VARIABLE MODELING
WITH COMPOSITE LINKS AND EXPLODED LIKELIHOODS

SOPHIA RABE-HESKETH

UNIVERSITY OF CALIFORNIA AT BERKELEY AND UNIVERSITY OF LONDON

ANDERS SKRONDAL

LONDON SCHOOL OF ECONOMICS AND NORWEGIAN INSTITUTE OF PUBLIC HEALTH

Composite links and exploded likelihoods are powerful yet simple tools for specifying a wide range of latent variable models. Applications considered include survival or duration models, models for rankings, small area estimation with census information, models for ordinal responses, item response models with guessing, randomized response models, unfolding models, latent class models with random effects, multilevel latent class models, models with log-normal latent variables, and zero-inflated Poisson models with random effects. Some of the ideas are illustrated by estimating an unfolding model for attitudes to female work participation.

Key words: composite link, exploded likelihood, unfolding, multilevel model, generalized linear mixed model, latent variable model, item response model, factor model, frailty, zero-inflated Poisson model, gllamm.

Introduction

Latent variable models are becoming increasingly general. An important advance is the accommodation of a wide range of response processes using a generalized linear model formulation (e.g., Bartholomew & Knott, 1999; Skrondal & Rabe-Hesketh, 2004b). Other recent developments include combining continuous and discrete latent variables (e.g., Muthén, 2002; Vermunt, 2003), allowing for interactions and nonlinear effects of latent variables (e.g., Klein & Moosbrugger, 2000; Lee & Song, 2004) and including latent variables varying at different levels (e.g., Goldstein & McDonald, 1988; Muthén, 1989; Fox & Glas, 2001; Vermunt, 2003; Rabe-Hesketh, Skrondal, & Pickles, 2004a).

Instead of treating different model types as separate it is conceptually appealing to consider a unifying model framework. This encourages specification of models tailormade to research problems by making it easy to combine features of different model types. Furthermore, a framework facilitates a unified approach to estimation that can be implemented in a single software program.

When responses are noncontinuous, all main approaches to latent variable modeling such as item response modeling, structural equation modeling (including factor analysis), and multilevel modeling use either a generalized linear ('response function') formulation or a latent response formulation for the relationship between latent variables and responses. Models defined via a latent response formulation, as well as linear models, can equivalently be expressed using the generalized linear model formulation (e.g., Takane & de Leeuw, 1987; Bartholomew & Knott, 1999; Skrondal & Rabe-Hesketh, 2004b).

In this paper we show that frameworks based on a generalized linear model formulation can be extended considerably by using *composite links* (e.g., Thompson & Baker, 1981; Cox, 1984;

Rindskopf, 1992). Here, the conditional expectations of the responses are linked not just to a single inverse link function of a linear predictor but instead to a linear combination of several. The papers cited above provide impressive catalogs of model types that can be generated using composite links without latent variables. Here we extend their work by proposing new uses of composite links; some are novel even without latent variables whereas others can only be used with latent variables. The idea of combining composite links and latent variables has been outlined in a keynote lecture (Skrondal & Rabe-Hesketh, 2004a) but to our knowledge not fully explored before.

We also discuss the use of what we call *exploded likelihoods*, where the likelihood is expressed as a product of contributions of standard form so that estimation can be accomplished by expanding the data and generating artificial responses. This trick is well known in some areas such as discrete time survival modeling (e.g., Allison, 1982). Here we also describe some less familiar applications and show the utility of the approach when combined with composite links. In particular, we demonstrate that a zero-inflated Poisson model can be written as a generalized linear model with composite link and exploded likelihood. It then becomes straightforward to include latent variables, thus producing novel models for counts.

A large number of interesting models can be specified using composite links and/or exploded likelihoods. It is important to note that we do not claim that these models cannot be specified and estimated without these devices. Indeed, a separate program or module (and paper) could be written for each model considered. However, an important merit of the approach suggested here is that all models are expressed within a single modeling framework. Unifying different models in this way invariably leads to model extensions because features of different model types can be combined. For instance, by embedding unfolding models within a framework that also includes MIMIC models, it becomes obvious that latent variables can be regressed on covariates in unfolding models. In addition to this conceptual advantage of a single framework, the practical advantage of having a single software package for a general framework should not be underestimated. This allows researchers to tailormake models addressing their research question and to accommodate unusual features of the data, such as inadvertently merged categories, without having to develop new software. Indeed, it will be possible to specify and estimate models that the developers of the software never anticipated.

The plan of the paper is as follows. We first outline generalized linear models and a framework for latent variable modeling to introduce notation and model features to be combined with composite links and exploded likelihoods. Second, we describe composite links and explore different kinds of models which they can produce. Third, we discuss exploded likelihoods and describe various uses of this approach. Fourth, we consider the combined use of composite links and exploded likelihoods to specify zero-inflated Poisson models with latent variables. We also provide a brief illustration of some of these ideas to specify unfolding models for attitudes to work participation among American women. Finally, we close the paper with a brief conclusion.

Generalized Linear Models

In generalized linear models (e.g., McCullagh & Nelder, 1989) the conditional expectation $\mu_i \equiv \mathrm{E}(y_i|\mathbf{x}_i)$ of the response y_i for unit i given the covariates $\mathbf{x}_i$ is linked to a linear predictor $\nu_i = \mathbf{x}_i'\boldsymbol{\beta}$ through a link function $g(\cdot)$,

$$g(\mu_i) = \nu_i,$$

or, equivalently,

$$\mu_i = g^{-1}(\nu_i),$$

where $g^{-1}(\cdot)$ is the inverse link function.

Conditional distributions for the responses y_i, given their linear predictors, are specified as

$$f(y_i|v_i) = \exp\left\{ \frac{y_i\theta_i - b(\theta_i)}{\phi_i} + c(y_i, \phi_i) \right\}.$$

Here, θ_i is the canonical or natural parameter (a function of μ_i), ϕ_i is the scale or dispersion parameter, and $b(\cdot)$ and $c(\cdot)$ are functions depending on the member of the exponential family.

A Framework for Latent Variable Modeling Based on Generalized Linear Models

We briefly describe a particular framework called 'generalized linear latent and mixed models' or GLLAMMs (Rabe-Hesketh et al., 2004a; Skrondal & Rabe-Hesketh, 2004b). Other frameworks can of course also be combined with composite links and exploded likelihoods.

Response Model

In a multilevel setting, with for instance students i (level-1) nested in classes j (level-2) nested in schools k (level-3), the response model is a generalized linear model with linear predictor specified as

$$v_i = \mathbf{x}_i'\boldsymbol{\beta} + \sum_{l=2}^{L} \sum_{m=1}^{M_l} \eta_m^{(l)} \mathbf{z}_{mi}^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}. \tag{1}$$

Here, $\eta_m^{(l)}$ represents the mth latent variable at hierarchical level l, $\mathbf{z}_{mi}^{(l)}$ is a vector of covariates, and $\boldsymbol{\lambda}_m^{(l)}$ a corresponding vector of parameters. For notational simplicity we have omitted unit indices $j, k, \ldots$ for levels 2 to L. Note that multivariate responses, such as sets of items in a factor model, are treated as if the items (variables) were level-1 units and the original units were level-2 clusters (see also Rijmen, Tuerlinckx, De Boeck, & Kuppens, 2003; De Boeck & Wilson, 2004). The purpose of $\mathbf{z}_{mi}^{(l)\prime} \boldsymbol{\lambda}_m^{(l)}$ is to allow for item-specific factor loadings.

Items measuring a latent variable are sometimes of mixed types, for instance, continuous and dichotomous. To accommodate such models, the link functions $g_i(\mu_i)$ and conditional response distributions $f_i(y_i|v_i)$ can differ for different units or variables.

Structural Equations

For continuous latent variables we specify multilevel structural equations of the form

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\mathbf{w}_j + \boldsymbol{\zeta}_j, \tag{2}$$

where $\boldsymbol{\eta}_j$ is the vector of all $M = \sum_{l=2}^{L} M_l$ latent variables at different hierarchical levels for cluster j, $\mathbf{B}$ a strictly upper-triangular matrix of regression parameters, $\mathbf{w}_j$ a vector of covariates, $\boldsymbol{\Gamma}$ a matrix of regression parameters, and $\boldsymbol{\zeta}_j$ a vector of disturbances. The disturbances are multivariate normal with zero means and are uncorrelated across levels. Each element of $\boldsymbol{\zeta}_j$ varies at the same level as the corresponding element of $\boldsymbol{\eta}_j$. Importantly, this specification allows cross-level effects from higher- to lower-level latent variables.

Latent variables can be discrete either for latent class modeling or for nonparametric maximum likelihood estimation (NPMLE) where the latent variable distribution is left unspecified (e.g., Laird, 1978; Heckman & Singer, 1984). For latent class models a multinomial logit model can be used to allow class probabilities to depend on covariates.

Composite Links

Thompson and Baker (1981) suggested linking the expectation μ_i with a composite function of several linear predictors instead of a function of a single linear predictor as in generalized linear models. They considered three kinds of composite links:

1. *Simple composite links.* The expectation μ_i is given by a weighted sum of inverse links with known weights w_{qi},

$$\mu_i = \sum_q w_{qi}\, g_q^{-1}(\nu_{qi}), \qquad (3)$$

where ν_{qi} is the qth linear predictor for unit i and $g_q^{-1}(\cdot)$ is an inverse link function. Note that w_{qi} corresponds to the element in the ith row and qth column of the C matrix used by Thompson and Baker (1981).

2. *Bilinear composite links.* These links extend the simple composite links in two ways. First, the known constants w_{qi} are replaced by linear combinations $\boldsymbol{\alpha}'\mathbf{w}_{qi}$ of known constants with unknown parameters $\boldsymbol{\alpha}$, giving

$$\mu_i = \sum_q \boldsymbol{\alpha}'\mathbf{w}_{qi}\, g_q^{-1}(\nu_{qi}). \qquad (4)$$

Second, the expectation may be some (not necessarily linear) function $h\{\cdot\}$ of the above sum,

$$\mu_i = h\left\{ \sum_q \boldsymbol{\alpha}'\mathbf{w}_{qi}\, g_q^{-1}(\nu_{qi}) \right\}.$$

3. *General composite links.* These links arise if general functions $f_{qi}[g_q^{-1}(\nu_{qi})]$ replace $w_{qi} g_q^{-1}(\nu_{qi})$ in (3).

The focus of this paper is on simple composite links although we will also mention applications of bilinear composite links. We use quadrature methods to evaluate the marginal likelihood of multilevel and latent variable models with simple composite links and Newton–Raphson with numerical derivatives to maximize the likelihood (Rabe-Hesketh, Skrondal, & Pickles, 2005). For algorithms using analytical derivatives, the modifications necessary for including simple composite links are relatively minor. It may also be possible to modify algorithms based on linearization such as penalized quasilikelihood (e.g., Breslow & Clayton, 1993) along the lines outlined for the iterative weighted least squares algorithm in Thompson and Baker (1981).

We will now explore some applications of composite links, emphasizing their combination with multilevel and latent variable models.

Missing Categorical Data

In categorical data, groups of units with missing values on some of the variables produce what are sometimes called supplemental margins. Such marginal information can be combined with the complete data using log-linear models with composite links (e.g., Rindskopf, 1992). The expected count for a cell in a marginal table is equated with the sum of the expected counts of all cells in the complete table that have contributed to it, giving a composite link.

Typical applications are studies where an expensive or invasive 'gold standard' measurement of a response or covariate is collected only on a subset of units (validation sample or phase II sample) (see Espeland & Hui, 1987; Clayton, Spiegelhalter, Dunn, & Pickles, 1998). In the extreme case, the gold standard is missing for everyone, leading to latent class models which can also be formulated through composite links (see the subsection on 'Latent Class Models with Conditional Dependence').

In small area estimation, supplemental margins, in the form of one-way tables from census data can add valuable information for estimating associations. Here random effects can be used to model heterogeneity between areas (e.g., Tranmer et al., 2005).

Another application is for fused cells. This will be discussed in the illustrative example where a bug in the computer-assisted telephone interviewing software caused two of the response categories to be merged for some attitude items.

Three-Parameter Item Response Models

The conventional two-parameter logistic item response model can be specified as

$$\Pr(y_{ij} = 1 | v_{ij}) = \frac{\exp(v_{ij})}{1 + \exp(v_{ij})}, \qquad v_{ij} = \beta_i + \lambda_i \eta_j,$$

where η_j represents the ability of respondent j, $-\beta_i/\lambda_i$ is the difficulty and λ_i the discrimination parameter of item i. Note that the probability of a correct response approaches zero as ability tends to minus infinity.

When multiple choice questions are used in ability testing it is possible for respondents to guess the right answer. Thus, even very unable examinees will have a nonzero probability of getting a correct answer to an item which is at odds with the above model. The two-parameter model is therefore sometimes replaced by the three-parameter logistic item response model (e.g., Birnbaum, 1968),

$$\Pr(y_{ij} = 1 | v_{ij}) = c_i + (1 - c_i) \frac{\exp(v_{ij})}{1 + \exp(v_{ij})},$$

where c_i, often called a 'guessing parameter', is interpreted as the probability of a correct answer on item i for an examinee with ability equal to minus infinity.

If we treat the guessing parameters c_i as known, the response model can be expressed using a simple composite link

$$\Pr(y_{ij} = 1 | v_{ij}) = w_{1i} g_1^{-1}(v_{1ij}) + w_{2i} g_2^{-1}(v_{2ij}),$$

where $g_1(\cdot)$ is the identity link and $g_2(\cdot)$ is the logit link, $w_{1i} = c_i$, $w_{2i} = 1 - c_i$, $v_{1ij} = 1$, and $v_{2ij} = v_{ij}$ (see also Skrondal & Rabe-Hesketh, 2004b, section 9.4). Instead of treating c_i as known, we can estimate the guessing parameters by using the bilinear link (4) with $\boldsymbol{\alpha}' = (1, c_1, \ldots, c_I)$, $\mathbf{w}'_{1i} = (0, \mathbf{d}'_i)$, and $\mathbf{w}'_{2i} = (1, -\mathbf{d}'_i)$, where $\mathbf{d}_i$ is a vector of dimension I (where I is the number of items) with one in the ith position and zeros elsewhere. Special caution should be exercised in ensuring that this model is identified.

It is worth noting that the above kind of model (without latent variables) is said to have 'natural responsiveness' or 'nonzero background' in quantal response bioassay (e.g., Finney, 1971).

Randomized Response Models

It is hard to obtain trustworthy information from respondents regarding sensitive issues such as whether they have ever used illegal drugs. Improved prevalence estimates of sensitive issues can potentially be obtained by guaranteeing the privacy of the respondents. One way of doing this is by means of the Warner design (Warner, 1965) where the respondent answers a question selected at random from two possibilities, one question positively worded ('Have you *ever* used illegal drugs?') and the other negatively worded ('Have you *never* used illegal drugs?'). The privacy is secured by letting the realization of the random allocation be known to the respondent but unknown to the interviewer. The probability of giving a positive response is then given by

$$\Pr(y = 1) = p\pi + (1 - p)(1 - \pi),$$

where π is the unknown probability of a 'yes' status for the sensitive issue (e.g., having used illegal drugs) and p the known probability of selecting the positively worded question.

The Warner model can be combined with a model for π_i with linear predictor ν_i and link function $g(\cdot)$, for instance a logit link,

$$\Pr(y_i = 1|\nu_i) = p\frac{\exp(\nu_i)}{1 + \exp(\nu_i)} + (1 - p)\left(1 - \frac{\exp(\nu_i)}{1 + \exp(\nu_i)}\right),$$

or a probit link.

Since the probability of presenting the sensitive question p is a known constant, the response model can be expressed using a simple composite link

$$\Pr(y_i = 1|\nu_i) = w_{1i}g_1^{-1}(\nu_{1i}) + w_{2i}g_2^{-1}(\nu_{2i}),$$

where $g_1(\cdot)$ and $g_2(\cdot)$ are logit (or probit) links, $w_{1i} = p$, $w_{2i} = 1 - p$, and $\nu_{1i} = \nu_i = -\nu_{2i}$. In a similar manner, the unrelated-question and forced-alternative designs for sensitive questions can also be modeled using composite links.

When each respondent j is responding to a set of items i, it has been suggested to combine randomized response models with item response models by using linear predictors of the form $\nu_{ij} = \beta_i + \lambda_i\eta_j$ (e.g., Fox, 2005; Böckenholt & van der Heijden, 2007) or with latent class models (e.g., Böckenholt & van der Heijden, 2007). Fox (2005) also suggests using multilevel item response models. Such extensions are straightforward using a generalized latent variable framework with composite links.

Cumulative Models for Ordinal Responses

Cumulative models for ordinal responses, with S ordered response categories $s = 0, \dots, S - 1$, can be written as

$$\Pr(y_i \geq s|\nu_i) = g^{-1}(\nu_i - \kappa_s), \tag{5}$$

where ν_i is a linear predictor, the κ_s are threshold parameters ($\kappa_0 = -\infty, \kappa_S = \infty$), and the inverse link function is a cumulative distribution function such as the standard normal, logistic, or extreme value distribution.

The response probabilities can be written as composite links with $w_{1i} = 1$ and $w_{2i} = -1$,

$$\Pr(y_i = s|\nu_i) = g^{-1}(\nu_i - \kappa_s) - g^{-1}(\nu_i - \kappa_{s+1}).$$

Thissen and Steinberg (1986) call the models 'difference models' due to the above form. An advantage of the composite link formulation as opposed to custom-made software for ordinal responses is that left- and right-censoring, or even interval censoring of an ordinal response are easily accommodated by using appropriate values of w_{qi}. Another advantage is that it is easy to allow the thresholds to depend on covariates, $\kappa_{si} = \mathbf{z}_i'\boldsymbol{\delta}$, as suggested by Terza (1985) or to include category-specific covariates in the linear predictor. The latter is useful if a cumulative model is used for discrete time survival data where the categories s correspond to time intervals and where time-varying covariates, such as time itself, are included in the model. An example of such a model is given in Rabe-Hesketh, Yang, and Pickles (2001) and Skrondal and Rabe-Hesketh (2004b, p. 381).

Letting the linear predictor be $\nu_{ij} = \beta_i + \lambda_i\eta_j$ as in the two-parameter item response model and the thresholds be item-specific κ_{is}, we obtain Samejima's (1969) graded response model for ordinal items. Multilevel versions of cumulative models have been discussed by, for instance, Jansen (1992) and Fielding (2003) and latent class versions by Vermunt (2001).
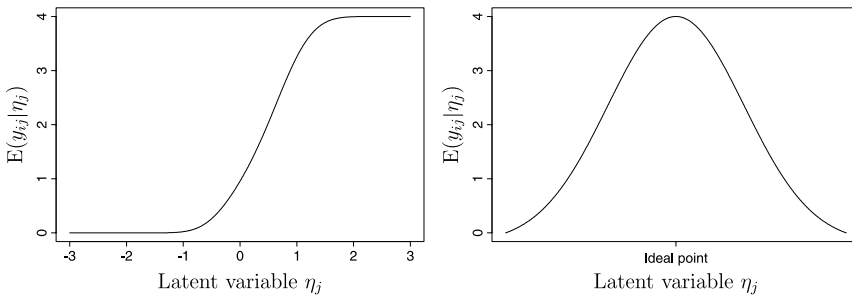
FIGURE 1.
Conditional response expectation for ordinal 'monotonic' item (left panel) and ordinal 'unfolding' or 'ideal point' item (right panel).

Unfolding or Ideal Point Models

In all item response models considered so far, the probability of a positive response or the conditional response expectation for an item is a monotonic function of the latent trait η_j. This assumption may be violated for attitude items where respondents are asked to rate their agreement as 'disagree' or 'agree', or more generally, in terms of ordered categories $s = 0, \ldots, S - 1$.

The classical example of Thurstone (1928) concerns attitudes to capital punishment measured as agree/disagree responses to statements such as 'capital punishment seems wrong but is sometimes necessary'. As sentiments favoring capital punishment increase from negative infinity, it seems reasonable that the probability of endorsing the statement initially increases from 0, reaches a maximum, and then declines as the latent trait increases toward infinity. In other words, hardline conservatives will disagree because capital punishment is not regarded as wrong whereas hardline liberals will disagree with the statement because capital punishment is deemed never to be justified. Respondents somewhere between these extremes, at the 'ideal point', will have the greatest probability of agreeing with the statement.

An unfolding or ideal point process is illustrated for an ordinal five-category item in the right panel of Figure 1, where the conditional response expectation

$$\mathrm{E}(y_{ij}|\eta_j) = \sum_{s=0}^{S-1} s \, \Pr(y_{ij} = s|\eta_j)$$

is shown as a function of a latent variable such as sentiment in favor of capital punishment. Note that this expected response function has a peak at the 'ideal point' in contrast to the monotonic function shown in the left panel of Figure 1 for Samejima's graded response model.

It has been argued (e.g., Andrich & Luo, 1993; Verhelst & Verstralen, 1993; Andrich, 1996; Roberts & Laughlin, 1996) that a respondent may give a particular rating s of an attitude item for two reasons. Considering 'disagree', he can 'disagree from below' because his latent trait is below the position of the item or 'disagree from above' because it exceeds the position. These two possibilities can be expressed in terms of $2S$ 'subjective ratings' z_{ij} such that $z_{ij} = s$ if the respondent 'disagrees from below' and $z_{ij} = 2S - s - 1$ if he 'disagrees from above'. The relationship between objective (or observed) and subjective ratings is illustrated in Figure 2. For instance, both the subjective rating 'strongly disagree from below' (0) and 'strongly disagree from above' (9) will produce an objective or observed rating 'strongly disagree'. Andrich (e.g., Andrich, 1996) considers a slight variation of this setup where there is only one subjective 'strongly agree' rating, whereas the other subjective ratings could either be 'from below' or 'from above'.
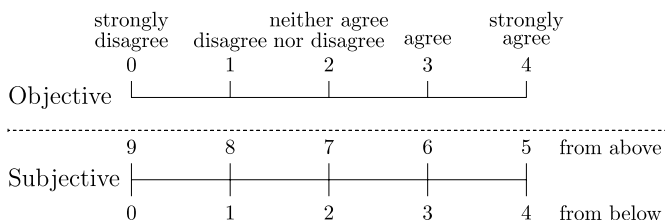
FIGURE 2.
Objective and subjective ratings.

Since the z_{ij} are not observed, the probability of the observed rating y_{ij}, given the latent trait η_j, can be written as the sum of the probabilities of the two disjunct 'subjective ratings' corresponding to the observed rating. We propose using a cumulative model (5) for the subjective ratings z_{ij} which gives

$$\Pr(y_{ij} = s|\eta_j) = \Pr(z_{ij} = s|\eta_j) + \Pr(z_{ij} = 2S - s - 1|\eta_j)$$

$$= [g^{-1}(v_{ij} - \kappa_s) - g^{-1}(v_{ij} - \kappa_{s+1})] + [g^{-1}(v_{ij} - \kappa_{2S-s-1}) - g^{-1}(v_{ij} - \kappa_{2S-s})],$$

where $v_{ij} = \beta_i + \lambda_i \eta_j$ as in a two-parameter item response model. Note that we are using a composite link with four components to model the objective responses. For identification, the thresholds must be constrained as, for instance, $\kappa_s = -\kappa_{2S-s}(s = 1, \ldots, S - 1)$, $\kappa_S = 0$, and $\kappa_{2S} = \infty$.

Unfolding models have been advocated not only for attitude items such as those considered here but also for preference items (e.g., Coombs, 1964; DeSarbo & Hoffman, 1986; Hoijtink, 1990; Andrich, 1995), pairwise comparisons (e.g., Andrich, 1989), and stages of development (e.g., Noël, 1999). Nonparametric unfolding response models are discussed by Johnson (2006).

Importantly, embedding the unfolding models in a general latent variable modeling framework produces a wide range of novel models. We can specify models with multidimensional continuous latent traits, possibly with unspecified distributions using nonparametric maximum likelihood estimation. Alternatively, latent variables can be taken as discrete using latent class specifications. Furthermore, we can regress latent variables on covariates as will be shown in the illustrative example and on the same or higher-level latent variables. This appears not to have been done before.

Latent Class Models with Conditional Dependence

Latent class models can be specified by modeling the 'complete' data (including latent class membership) using log-linear models (e.g., Haberman, 1979). In the standard shorthand notation for hierarchical log-linear models, a latent class model with three observed responses Y_1 to Y_3 can be written as $\{CY_1, CY_2, CY_3\}$, where C denotes latent class membership. Since latent class membership is unknown, we must sum over the latent classes to obtain expected counts for the observable response patterns. For simplicity and without loss of generality, we consider a two-class model with three dichotomous observed responses y_i ($i = 1, 2, 3$) which are conditionally independent given class membership. Using dummy coding, the log-linear model can be written as

$$\log \mu_{y_1 y_2 y_3 c} = v_{y_1 y_2 y_3 c} = \beta_0 + \alpha_0 c + \sum_{i=1}^{3} \beta_i y_i + \sum_{i=1}^{3} \alpha_i c y_i, \tag{6}$$

where $c = 0, 1$ is the latent class indicator, $\mu_{y_1 y_2 y_3 c}$ is the expected count for response pattern y_1, y_2, y_3 and latent class c, and β_p and α_p ($p = 0, 1, 2, 3$) are parameters for the main effects of class and item and their interactions. The expected values $\mu_{y_1 y_2 y_3}$ of the observed counts are

modeled as the sum of the class-specific expected counts

$$\mu_{y_1 y_2 y_3} = \exp(\nu_{y_1 y_2 y_3 0}) + \exp(\nu_{y_1 y_2 y_3 1}),$$

giving a composite link with two log links. As pointed out by Rindskopf (1992) and others, this formulation makes it easy to relax conditional independence by including interactions among pairs of items. For instance, dependence between items i and i' can be specified by including an interaction term $\gamma_{ii'} y_i y_{i'}$ in (6).

An alternative approach to relaxing conditional independence is to include a continuous subject-specific random intercept η_j within a latent class model (e.g., Qu, Tan, & Kutner, 1996). To accomplish this we expand the data to obtain counts (0 or 1) for each response and latent class pattern for each subject j. The linear predictor for the corresponding expected count can then be written as

$$\nu_{y_1 y_2 y_3 cj} = \beta_0 + \alpha_0 c + \sum_{i=1}^{3} \beta_i y_i + \sum_{i=1}^{3} \alpha_i y_i c + \eta_j \left(\sum_{i=1}^{3} y_i (1-c) \lambda_{i0} + \sum_{i=1}^{3} y_i c \lambda_{i1} \right), \quad (7)$$

where η_j can be interpreted as subject j's propensity to have a '1' (e.g., score positively on a diagnostic test), with item-specific effects λ_{i0} for class 0 (e.g., healthy subjects) and λ_{i1} for class 1 (e.g., ill subjects). Since the total count for each person j is fixed at 1, we must estimate the multinomial logit version of this model

$$\Pr(y_1, y_2, y_3, c) = \frac{\exp(\nu_{y_1 y_2 y_3 cj})}{\sum_{y_1 y_2 y_3 c} \exp(\nu_{y_1 y_2 y_3 cj})}.$$

Again, we do not know c, so the likelihood contribution for subject j becomes

$$\Pr(y_1, y_2, y_3) = \frac{\exp(\nu_{y_1 y_2 y_3 0 j}) + \exp(\nu_{y_1 y_2 y_3 1 j})}{\sum_{y_1 y_2 y_3 c} \exp(\nu_{y_1 y_2 y_3 cj})}.$$

This is a composite link model if each multinomial logit term is viewed as an inverse link. An alternative to the multinomial logit approach is to use a log-linear model with a subject-specific constant δ_j in the linear predictor in (7) (e.g., Palmgren, 1981; Chen & Kuo, 2001). The advantage of the latter approach is that software for ordinary generalized linear mixed models can be used if extended to allow composite links. A disadvantage is that a large number of nuisance parameters must be estimated.

Multilevel Latent Class Models

Latent class membership may depend on observed covariates (e.g., Dayton & MacReady, 1988). For the case of a single observed covariate x the corresponding log-linear model becomes $\{XC, CY_1, CY_2, CY_3\}$. If all response variables are dichotomous, the model can be written as

$$\nu_{y_1 y_2 y_3 cx} = \beta_0 + \alpha_0 c + \delta_0 x + \delta_1 xc + \sum_{i=1}^{3} \beta_i y_i + \sum_{i=1}^{3} \alpha_i c y_i,$$

and can be estimated as described in the previous section.

When the subjects j are nested in clusters k, we may want to allow class membership to depend on a cluster-level random intercept $\zeta_k^{(3)}$ instead (Vermunt, 2003, 2007). An example would be disease mapping or small area estimation of the prevalence of clinical depression in different geographical areas k when true depression of subjects j (nested in areas k) has been diagnosed using several imperfect methods.

The linear predictor must include a fixed intercept for each cluster since the number of subjects per cluster is treated as fixed (Palmgren, 1981) and can be written as

$$\nu_{y_1 y_2 y_3 ck} = \beta_0 + \alpha_0 c + \delta_k + \zeta_k^{(3)} c + \sum_{i=1}^{3} \beta_i y_i + \sum_{i=1}^{3} \alpha_i c y_i,$$

where, typically, $\zeta_k^{(3)} \sim N(0, \psi^{(3)})$ or, alternatively, $\zeta_k^{(3)}$ is specified as discrete.

This setup requires that there is a response for each cell in the five-way contingency table which can result in a large data set with many zero counts if the clusters are small. Furthermore, many nuisance parameters δ_k must be estimated if there are many clusters. Nevertheless, both these computational disadvantages are likely to be less severe than for the models with subject-specific random effects discussed in the previous section.

Models with Log-Normal Latent Variables

In a linear mixed model we may want to specify skewed random effect distributions, for instance, for sensitivity analysis. A convenient choice of distribution is the log-normal, a simple example being

$$\mu_{ij} = \mathbf{x}'_{ij}\beta + \exp(\eta_{1j}) + \exp(\eta_{2j})z_{ij},$$

where η_{1j} and η_{2j} have a bivariate normal distribution. This model can be specified using a composite link with $w_{1i} = 1$, $w_{2i} = 1$, and $w_{3ij} = z_{ij}$, with corresponding identity link and two log links. It can easily be extended to L levels.

If we use a bilinear composite link, we can include log-normal latent variables in factor models as well,

$$\mu_{ij} = \mathbf{d}'_i \beta + \exp(\eta_{1j}) \mathbf{d}'_i \lambda = \beta_i + \exp(\eta_{1j}) \lambda_i.$$

This uses the feature that the weights in the composite link can be linear combinations $\alpha' \mathbf{w}_{qi}$ with $\alpha' = (1, \lambda')$, $\mathbf{w}'_{1i} = (1, 0, \ldots, 0)$ and $\mathbf{w}'_{2i} = (1, \mathbf{d}_i)$. This is combined with an identity link $g_1^{-1}(\nu_{1i}) = \mathbf{d}_i \beta$ and a log link $g_2^{-1}(\nu_{2ij}) = \exp(\eta_{1j})$.

We can use the additional feature of bilinear composite links that the expectation can be equated to a function $h\{\cdot\}$ of the right-hand sides in the equations above. This produces generalized linear mixed models and item response models, respectively. Two-parameter item response models with a log-normal latent variable can therefore be specified as generalized linear mixed models with bilinear composite links.

Some Other Applications of Composite Links

Candy (1997) introduced so-called additive generalized linear mixed models,

$$\mu_{ij} = g^{-1}(\mathbf{x}'_{ij}\beta) + \mathbf{z}'_{ij}\eta_j,$$

where the conditional expectation is linear in the random effects. The first link is one of the standard links for generalized linear models, whereas the second (for the random part) is the identity link. The random part is thus 'external' to the link function as discussed by Allison (1987).

Heisterkamp, van Houwelingen, and Downs (1999) used composite links for back-calculation, where the aim is to estimate the incidence of infections over time from a time series of counts of new diagnosed cases, taking into account the time delay between infection and diagnosis. A log-linear model with linear predictor ν_q is specified for the number of new infections in period q and the corresponding inverse link functions are combined as in (3) where the weight w_{qi} represents the probability (assumed known) that an infection that occurred in period q is diagnosed in a later period i.

Similarly, Eilers and Borgdorff (2004) described a transfer model for digit preference where a log-linear model is specified for the true frequency distribution (if there were no digit-preference). The weight w_{qi} is then the probability that a true response q is misclassified as i. If the misclassification probabilities are not known a bilinear composite link (4) can be used.

Consider a probit or logit model for a true binary response t_i that is imperfectly observed as y_i with known probabilities of misclassification (or sensitivity and specificity) that are independent of the covariates. A model for the binary observed response can be written using a simple composite link of the form

$$\Pr(y_i = 1|v_i) = \Pr(y_i = 1|t_i = 1)\Pr(t_i = 1|v_i) + \Pr(y_i = 1|t_i = 0)\Pr(t_i = 0|v_i)$$

$$= w_1 g^{-1}(v_i) + w_2 g^{-1}(-v_i).$$

Although this model has been estimated by Magder and Hughes (1997) and many others (see Neuhaus, 1999), it has apparently not been pointed out that it is a model with a composite link. The model can easily be extended to include random effects or latent variables.

Exploded Likelihoods

A single response y_i is sometimes 'exploded' into a set of artificial responses a_{ri}, $r = 1, \ldots, R_i$, with corresponding explosion of the explanatory variables. The point of doing this is to write the conditional response probability (density) for y_i as a product,

$$\Pr(y_i|v_{1i}, \ldots, v_{Ri}) = \prod_{r=1}^{R_i} f_{ri}(a_{ri}|v_{ri}).$$

Here, $f_{ri}(\cdot)$ are distributions from the exponential family (possibly different distributions for different ri), with conditional means given by

$$\mu_{ri} = g_{ri}^{-1}(v_{ri}),$$

where $g_{ri}^{-1}(\cdot)$ are link functions (possibly different for different ri) and v_{ri} is a linear predictor based on the exploded explanatory variables.

Some Previous Applications of Exploded Likelihoods

Consider an unordered polytomous response with S categories, $s = 1, \ldots, S$. Such data arise for instance if subjects select their most preferred among S alternatives. It is quite well known that the multinomial logit model for such discrete choice data can be estimated by expanding each original observation to S artificial observations corresponding to the available alternatives. We will refer to the groups of observations representing each original observation as alternative sets. A choice indicator is created that is equal to one for the alternative that was chosen and zero otherwise. Treating this choice indicator as the artificial response, the model can then be estimated using conditional logistic regression, conditioning on the sums of responses for the alternative sets, or using Poisson regression with a constant for each alternative set (e.g., Chen & Kuo, 2001). Stratified Cox regression can also be used, treating the choice dummy variable as the failure indicator and the alternative sets as strata (e.g., Allison & Christakis, 1994; Chen & Kuo, 2001).

The term 'exploded logit' (Chapman & Staelin, 1982) is often used to describe the Luce–Plackett model for rankings (Luce, 1959; Plackett, 1975). Let r_i^{ℓ} be the alternative given rank ℓ among $s = 1, \ldots, S$ alternatives for subject i and let $\mathbf{R}_i \equiv (r_i^1, r_i^2, \ldots, r_i^S)$ be the complete ranking of alternatives. The likelihood contribution of a ranking becomes the product of successive

multinomial logit choice probabilities among remaining alternatives

$$\Pr\big(\mathbf{R}_i|v_i^1,\ldots,v_i^S\big) = \prod_{\ell=1}^{S-1} \frac{\exp\big(v_i^{r_i^\ell}\big)}{\sum_{m=\ell}^{S} \exp\big(v_i^{r_i^m}\big)}.$$

Thus, $S-1$ artificial polytomous responses are generated for each ranking. Each of these polytomous responses could be expanded further as explained for first choice data above. Skrondal and Rabe-Hesketh (2003) and Böckenholt (2001) provide overviews of multilevel and latent variable models for discrete choices and rankings.

It has been shown that the partial likelihood of the Cox proportional hazards model for survival data is proportional to the Poisson likelihood of an appropriately expanded data set (e.g., Holford, 1976; Whitehead, 1980). Clayton (1988), Goldstein (2003), and Skrondal and Rabe-Hesketh (2004b) used this exploded likelihood idea to estimate models with random effects, known as frailties in this context.

Similar data expansion can be used to estimate discrete time survival models by binary regression (e.g., Allison, 1982). Using a complementary log-log link, this corresponds to assuming a Cox proportional hazards model for the unobserved continuous survival times whereas a logit link produces a continuation ratio model (Cox, 1972). Läärä and Matthews (1985) show that the continuation ratio model is equivalent to a cumulative model with a complementary log-log link. As Engel (1993) points out, this model can therefore be estimated via an exploded likelihood as shown above or via a composite link as shown previously. Frailty versions of discrete time hazard models have been discussed by, for instance, Rabe-Hesketh et al. (2001) for continuous frailties and by Vermunt (1997) and Muthén and Masyn (2005) for discrete frailties.

Consider a linear regression model with normal errors where the response is subject to censoring. For the noncensored responses the likelihood contribution is $\phi((y_i - v_i)/\sigma)/\sigma$, corresponding to an identity link and a normal distribution with standard deviation σ. For responses that are right-censored at t_i, the likelihood contribution is $\Phi((v_i - t_i)/\sigma)$, corresponding to a scaled probit link (with $-t_i$ as an offset), and a Bernoulli distribution. This tobit model can be produced by creating an artificial response equal to 1 for each right-censored response (and equal to 0 for responses that are left-censored at t_i). If y_i represents log survival time, we obtain the log-normal survival model (e.g., Cox, 1984) which has been extended to multilevel or frailty models by, for instance, Goldstein (2003).

Combining Composite Links and Exploded Likelihoods: Zero-Inflated Poisson (ZIP) Models with Latent Variables

The ZIP model (e.g., Lambert, 1992) is a finite mixture model for counts where the population is assumed to consist of two components; a component $c = 0$ where the count can only be zero and a component $c = 1$ where the count has a Poisson distribution.

A logistic regression model is specified for the probability π_{0i} of belonging to the zero-count component $c = 0$,

$$\pi_{0i} = \frac{\exp(v_{0i})}{1 + \exp(v_{0i})}, \tag{8}$$

and a Poisson distribution is assumed for the count observed in the other component $c = 1$,

$$\Pr(y_i = k|v_{1i}) = \exp(-\mu_{1i})\mu_{1i}^k/k!, \qquad \mu_{1i} = \exp(v_{1i}). \tag{9}$$

The probability of a count that is nonzero is

$$\Pr(y_i = k > 0 | v_{0i}, v_{1i}) = \Pr(y_i = k > 0, c = 1 | v_{0i}, v_{1i}) = (1 - \pi_{0i}) \exp(-\mu_i) \mu_i^k / k!$$

$$= \left(\frac{1}{1 + \exp(v_{0i})} \right) \left[\exp(-\mu_i) \mu_i^k / k! \right]$$

and the probability of a zero count is

$$\Pr(y_i = 0 | v_{0i}, v_{1i}) = \Pr(y_i = 0, c = 0 | v_{0i}, v_{1i}) + \Pr(y_i = 0, c = 1 | v_{0i}, v_{1i})$$

$$= \pi_{0i} + (1 - \pi_{0i}) \exp(-\mu_i)$$

$$= \left(\frac{1}{1 + \exp(v_{0i})} \right) [\exp(v_{0i}) + \exp(-\exp(v_{1i}))].$$

For a nonzero count, the probability is the product of the probability of an artificial response $a_{1i} = 0$ in a logistic regression model with linear predictor v_{0i} and the Poisson probability of a count response $a_{2i} = k$ with a log link and linear predictor v_{1i}. Therefore, for nonzero counts, we obtain the correct likelihood by creating two responses, 0 and k, and specifying a mixed response (logistic/Bernoulli and log/Poisson) model.

For a zero count, we again create a $a_{1i} = 0$ response, modeled as a logistic regression, for the first term. For the second term, we specify a composite link,

$$[\exp(v_{0i}) + \exp(-\exp(v_{1i}))] = g_1^{-1}(v_{0i}) + g_2^{-1}(v_{1i}),$$

where $g_1(\cdot)$ is the log link and $g_2(\cdot)$ the log-log link. If we create a $a_{2i} = 1$ response and specify a Bernoulli distribution with this composite link, we obtain the required term.

Within this setup it is straightforward to include random effects or latent variables. A potential application would be modeling the number of alcoholic drinks consumed by respondents nested in regions. Region-specific random effects can be used in both (8) and (9) to model variations in the prevalence of nondrinking and in the amount consumed among drinkers, with possible correlations between these random effects. Hall (2000) considers the restrictive special case with a random intercept only for $c = 1$ in (9), without using the idea of combining composite links and exploded likelihoods. We can also specify other ZIP models with latent variables, such as random coefficient models, covariate measurement error models, etc.

Illustrative Example: Unfolding Attitudes to Female Work Participation

In 1988 and 2002 the respondents of the US General Social Survey (Davis, Smith, & Marsden, 2003) were presented with the following attitude statements regarding female work participation:

  [famhapp] A woman and her family will all be happier if she goes to work
  [twoincs] Both the husband and wife should contribute to the family income
  [warmrel] A working mother can establish just as warm and secure a relationship with her children as a woman who does not work
  [jobindep] Having a job is the best way for a woman to be an independent person
  [housewrk] Being a housewife is just as fulfilling as working for pay
  [homekid] A job is alright, but what most women really want is a home and children
  [famsuff] All in all, family life suffers when the woman has a full-time job
  [kidsuff] A pre-school child is likely to suffer if his or her mother works
  [hubbywrk] A husband's job is to earn money; a wife's job is to look after the home

The respondents rated each statement as either 'disagree completely' (0), 'disagree' (1), 'agree somewhat' (2), 'agree' (3), or 'agree completely' (4).

We use an unfolding model of the type proposed earlier in the paper to investigate sentiments in favor of female work participation η_j among female respondents. We specify $g(\cdot)$ as scaled probit links with item-specific scale parameters σ_i (estimated on the log-scale),

$$g^{-1}(\nu_{ij} - \kappa_s) = \Phi\left(\frac{\beta_i + \lambda_i \eta_j - \kappa_s}{\sigma_i}\right).$$

A special complication arises in the analysis of these data. The 'disagree completely' and 'disagree' responses were inadvertently collapsed into a single 'disagree' response in 2002 due to a bug in the computer-assisted telephone interviewing (CATI) software. Thus, in 2002 the composite link for 'disagree' is the sum of the composite links for 'disagree' and 'disagree completely'. Note that this use of composite links is useful whether or not latent variables are involved in the model. Another challenge of these data is that three of the items ([famhapp], [twoincs], and [hubbywrk]) were not included in the 2002 questionnaire. We treat these items as missing at random which is reasonable since they are missing by design.

To investigate if sentiments in favor of female work participation η_j have changed from 1988 to 2002, we specify the structural model

$$\eta_j = \gamma_1 w_j + \zeta_j, \qquad \zeta_j \sim N(0, \psi),$$

where w_j is a dummy variable for year being [2002]. Estimation of the mean change in sentiment γ_1 is possible here because the item-specific parameters of items included both in 1988 and 2002 are set equal across time.

TABLE 1.
Maximum likelihood estimates for scaled probit unfolding model.

Item i	Item parameters					
	β_i		λ_i		$\ln \sigma_i$	
	Est	SE	Est	SE	Est	SE
[famhapp]	−2.32	0.08	0.30	0.04	−0.24	0.05
[twoincs]	−1.60	0.07	0.29	0.05	−0.06	0.05
[warmrel]	−0.99	0.07	1		0	
[jobindep]	−0.27	0.14	1.15	0.15	0.64	0.05
[housewrk]	1.29	0.08	0.54	0.08	0.22	0.06
[homekid]	2.11	0.07	0.76	0.06	−0.06	0.04
[famsuff]	2.19	0.08	1.43	0.09	−0.29	0.05
[kidsuff]	2.24	0.08	1.49	0.09	−0.46	0.06
[hubbywrk]	2.42	0.09	1.14	0.09	−0.11	0.05

Threshold parameters ($\kappa_{2S-s} = -\kappa_s$)		
	Est	SE
κ_1 ('disagree completely'/'disagree')	3.43	0.11
κ_2 ('disagree'/'agree somewhat')	2.36	0.08
κ_3 ('agree somewhat'/'agree')	1.67	0.06
κ_4 ('agree/'agree completely')	0.72	0.03
κ_5	0	

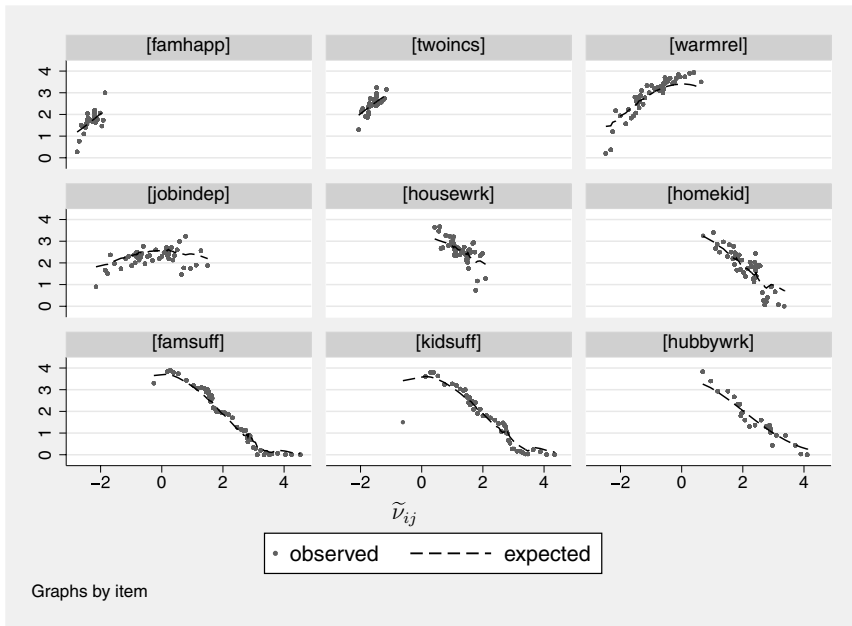| Latent trait regression parameters		
	Est	SE
[2002] γ_1	−0.04	0.04
Residual variance ψ	0.62	0.08

FIGURE 3.
Mean expected and observed responses as a function of linear predictor $\widetilde{\nu}_{ij}$.

Maximum likelihood estimates from gllamm (Rabe-Hesketh, Skrondal, & Pickles, 2004b; Rabe-Hesketh & Skrondal, 2005) based on data from 1462 female respondents are given in Table 1. Here the items have been ordered from the most positive to the most negative according to their estimated $\widehat{\beta}_i$. Since the magnitude of $\widehat{\gamma}_1$ is negligible, mean sentiment in favor of female work participation does not appear to have changed.

Following Roberts and Laughlin (1996) we assess model fit graphically. First, we estimate the linear predictor $\widetilde{\nu}_{ij}$ of respondent j and item i by plugging in the parameter estimates and the empirical Bayes prediction $\widetilde{\eta}_j$ of the latent trait into the linear predictor. Substituting this into the unfolding model, we obtain the expected response category for each person–item pair. Grouping the $\widetilde{\nu}_{ij}$ into approximately homogeneous groups of size 30 for each item and plotting the corresponding average observed and expected frequencies versus the average $\widetilde{\nu}_{ij}$ for each item gives Figure 3. Our unfolding model appears to fit quite well.

Although the expected response takes the form of a single-peaked function consistent with an unfolding process when all items are considered together, none of the individual items exhibit single-peaked behavior with the possible exception of [jobindep]. Using conventional item response models that assume monotonicity might therefore be appropriate if either: (1) reversing the coding of the appropriate items can be based on a priori information; or (2) the model accommodates negative factor loadings.

Conclusion

Composite links and exploded likelihoods are remarkably powerful tools for specifying novel latent variable models while remaining in a unified modeling framework. Indeed, we do not in any way purport to exhaust potential applications in this paper. For instance, composite links and exploded likelihoods could fruitfully be combined with covariate measurement error models, multilevel structural equation models, and models including both random effects and common factors.

It should be emphasized that a general latent variable framework is not required to exploit the ideas discussed here. Each of the uses of composite links can be combined with ordinary generalized linear mixed models to produce novel models, albeit less general than the ones considered here. For instance, in the extensions to item response models the discrimination parameters can be set to one. The devices described to specify randomized response, latent class, tobit, and zero-inflated Poisson models can also be used with ordinary generalized linear models.

Implementation of composite links and exploded likelihoods is quite straightforward in software for multilevel and latent variable modeling based on generalized linear models. For instance, maximum likelihood estimation, as well as pseudo maximum likelihood estimation with probability weights (Rabe-Hesketh & Skrondal, 2006), is available in the gllamm software (Rabe-Hesketh et al., 2004b; Rabe-Hesketh & Skrondal, 2005). All link functions in gllamm can be used as building blocks in simple composite links.

References

Allison, P.D. (1982). Discrete time methods for the analysis of event histories. In S. Leinhardt (Ed.), *Sociological methodology 1982*. San Francisco: Jossey-Bass.

Allison, P.D. (1987). Introducing a disturbance into logit and probit regression models. *Sociological Methods & Research*, *15*, 355–374.

Allison, P.D., & Christakis, N.A. (1994). Logit models for sets of ranked items. In P.V. Marsden (Ed.), *Sociological methodology 1994* (pp. 199–228). Oxford: Blackwell.

Andrich, D. (1989). A probabilistic item response theory model for unfolding preference data. *Applied Psychological Measurement*, *13*, 193–216.

Andrich, D. (1995). Hyperbolic cosine latent trait models for unfolding direct-responses and pairwise preferences. *Applied Psychological Measurement*, *19*, 269–290.

Andrich, D. (1996). A hyperbolic cosine latent trait model for unfolding polytomous responses: Reconciling Thurstone and Likert methodologies. *British Journal of Mathematical and Statistical Psychology*, *49*, 347–365.

Andrich, D., & Luo, G. (1993). A hyperbolic cosine latent trait model for unfolding dichotomous single stimulus responses. *Applied Psychological Measurement*, *17*, 253–276.

Bartholomew, D.J., & Knott, M. (1999). *Latent variable models and factor analysis*. London: Arnold.

Birnbaum, A. (1968). Test scores, sufficient statistics, and the information structures of tests. In F.M. Lord, & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 425–435). Reading, MA: Addison-Wesley.

Böckenholt, U. (2001). Mixed-effects analyses of rank-ordered data. *Psychometrika*, *66*, 45–62.

Böckenholt, U., & van der Heijden, P.G.M. (2007). Item randomized-response models for measuring noncompliance: Risk-return perceptions, social influences, and self-protective responses. *Psychometrika*, *72*, DOI: 10.1007/s11336-005-1495-y.

Breslow, N.E., & Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, *88*, 9–25.

Candy, S.G. (1997). Estimation in forest yield models using composite link functions with random effects. *Biometrics*, *53*, 146–160.

Chapman, R.G., & Staelin, R. (1982). Exploiting rank ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, *14*, 288–301.

Chen, Z., & Kuo, L. (2001). A note on the estimation of the multinomial logit model with random effects. *The American Statistician*, *55*, 89–95.

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

Clayton, D.G., Spiegelhalter, D., Dunn, G., & Pickles, A. (1998). Analysis of longitudinal binary data from multiphase sampling. *Journal of the Royal Statistical Society, Series B*, *60*, 71–87.

Coombs, C.H. (1964). *A theory of data*. New York: Wiley.

Cox, C. (1984). Generalized linear models—The missing link. *Journal of the Royal Statistical Society, Series C*, *33*, 18–24.

Cox, D.R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, *34*, 187–203.

Davis, J.A., Smith, T.W., & Marsden, P.V. (2003). *General social surveys, 1972–2002 [Cumulative File]*. Ann Arbor, MI: ICPSR [Distributor].

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

De Boeck, P., & Wilson, M. (Eds.) (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

DeSarbo, W.S., & Hoffman, D.L. (1986). Simple unweighted unfolding threshold models for the spatial representation of binary choice data. *Applied Psychological Measurement*, *10*, 247–264.

Eilers, P.H.C., & Borgdorff, M.W. (2004). Modeling and correction of digit preference in tuberculin surveys. *The International Journal of Tuberculosis and Lung Disease*, *8*, 232–239.

Engel, R. (1993). On the analysis of grouped extreme value data with GLIM. *Journal of the Royal Statistical Society, Series C*, *42*, 633–640.

Espeland, M.A., & Hui, S.L. (1987). A general approach to analyzing epidemiologic data that contain misclassification errors. *Biometrics*, *43*, 1001–1012.

Fielding, A. (2003). Ordered category responses and random effects in multilevel and other complex structures. In S.P. Reise, & N. Duan (Eds.), *Multilevel modeling. Methodological advances, issues, and applications* (pp. 181–208). Mahwah, NJ: Erlbaum.

Finney, D.J. (1971). *Probit analysis*. Cambridge: Cambridge University Press.

Fox, J.P. (2005). Randomized item response theory models. *Journal of Educational and Behavioral Statistics*, *30*, 1–24.

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

Goldstein, H. (2003). *Multilevel statistical models* (3rd ed.). London: Arnold.

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

Haberman, S.J. (1979). *Analysis of qualitative data: New developments* (Vol. 2). New York: Academic Press.

Hall, D.B. (2000). Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, *56*, 1030–1039.

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.

Heisterkamp, S.H., van Houwelingen, J.C., & Downs, A.M. (1999). Empirical Bayesian estimators for a Poisson process propagated in time. *Biometrical Journal*, *41*, 385–400.

Hoijtink, H. (1990). A latent trait model for dichotomous choice data. *Psychometrika*, *55*, 641–656.

Holford, T.R. (1976). Life tables with concomitant variables. *Biometrics*, *32*, 587–597.

Jansen, J. (1992). Statistical analysis of threshold data from experiments with nested errors. *Computational Statistics & Data Analysis*, *13*, 319–330.

Johnson, M.S. (2006). Nonparametric estimation of item and respondent locations from unfolding-type items. *Psychometrika*, *71*, 257–279.

Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.

Läärä, E., & Matthews, J.N.S. (1985). The equivalence of two models for ordinal data. *Biometrika*, *72*, 206–207.

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

Lambert, D. (1992). Zero-inflated Poisson-regression with an application to defects in manufacturing. *Technometrics*, *34*, 1–14.

Lee, S.-Y., & Song, X.-Y. (2004). Maximum likelihood analysis of a general latent variable model with hierarchically mixed data. *Biometrics*, *60*, 624–636.

Luce, R.D. (1959). *Individual choice behavior*. New York: Wiley.

Magder, S.M., & Hughes, J.P. (1997). Logistic regression when the outcome is measured with uncertainty. *American Journal of Epidemiology*, *146*, 195–203.

McCullagh, P., & Nelder, J.A. (1989). *Generalized linear models* (2nd ed.). London: Chapman & Hall.

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

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

Muthén, B.O., & Masyn, K. (2005). Dicrete-time survival mixture analysis. *Journal of Educational and Behavioral Statistics*, *30*, 27–58.

Neuhaus, J.M. (1999). Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, *86*, 843–855.

Noël, Y. (1999). Recovering unimodal latent patterns of change by unfolding analysis: Application to smoking cessation. *Psychological Methods*, *4*, 173–191.

Palmgren, J. (1981). The Fisher information matrix for log linear models arguing conditionally on observed explanatory variables. *Biometrika*, *68*, 563–566.

Plackett, R.L. (1975). The analysis of permutations. *Journal of the Royal Statistical Society, Series C*, *24*, 193–202.

Qu, Y., Tan, M., & Kutner, M.H. (1996). Random effects models in latent class analysis for evaluating accuracy of diagnostic tests. *Biometrics*, *52*, 797–810.

Rabe-Hesketh, S., & Skrondal, A. (2005). *Multilevel and longitudinal modeling using Stata*. College Station, TX: Stata Press.

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A*, *116*, 805–827.

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., Skrondal, A., & Pickles, A. (2004a). Generalized multilevel structural equation modeling. *Psychometrika*, *69*, 167–190.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004b). *GLLAMM manual*. Technical report 160. U.C. Berkeley Division of Biostatistics Working Paper Series. Downloadable from http://www.bepress.com/ucbbiostat/paper160/.

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, *128*, 301–323.

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

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*, 185–205.

Rindskopf, D. (1992). A general approach to categorical data analysis with missing data, using generalized linear models with composite links. *Psychometrika*, *57*, 29–42.

Roberts, J.S., & Laughlin, J.E. (1996). A unidimensional item response model for unfolding responses from a graded disagree–agree response scale. *Applied Psychological Measurement*, *20*, 231–255.

Samejima, F. (1969). *Estimation of latent trait ability using a response pattern of graded scores*. Psychometric Monograph, Vol. 17. Bowling Green, OH: The Psychometric Society.

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

Skrondal, A., & Rabe-Hesketh, S. (2004a). Generalised linear latent and mixed models with composite links and exploded likelihoods. In A. Biggeri, E. Dreassi, C. Lagazio, & M. Marchi (Eds.), *19th International workshop on statistical modeling* (pp. 27–39). Florence: Firenze University Press.

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

Takane, Y., & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, *52*, 393–408.

Terza, J.V. (1985). Ordinal probit: A generalization. *Communications in Statistics, Theory and Methods*, *14*, 1–11.

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*, *51*, 567–577.

Thompson, R., & Baker, R.J. (1981). Composite link functions in generalized linear models. *Journal of the Royal Statistical Society, Series C*, *30*, 125–131.

Thurstone, L.L. (1928). Attitudes can be measured. *American Journal of Sociology*, *33*, 529–554.

Tranmer, M., Pickles, A., Fieldhouse, E., Elliot, M., Dale, A., Brown, M., Martin, D., Steel, D., & Gardiner, C. (2005). The case for small area microdata. *Journal of the Royal Statistical Society, Series A*, *168*, 29–49.

Verhelst, N.D., & Verstralen, H.H.F.M. (1993). A stochastic unfolding model derived from the partial credit model. *Kwantitative Methoden*, *42*, 73–92.

Vermunt, J.K. 1997. *Log-linear models for event histories*. Thousand Oaks, CA: Sage.

Vermunt, J.K. (2001). The use of restricted latent class models for defining and testing nonparametric item response theory models. *Applied Psychological Measurement*, *25*, 283–294.

Vermunt, J.K. (2003). Multilevel latent class models. In R.M. Stolzenberg (Ed.), *Sociological methodology 2003* (Vol. 33, pp. 213–239). Oxford: Blackwell.

Vermunt, J.K. (in press). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*.

Warner, S.L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, *60*, 63–69.

Whitehead, J. (1980). Fitting Cox's regression model to survival data using GLIM. *Journal of the Royal Statistical Society, Series C*, *29*, 269–275.