doi: 10.1111/j.1467-9469.2007.00573.x

Latent Variable Modelling: A Survey*

ANDERS SKRONDAL

Department of Statistics and The Methodology Institute, London School of Economics and Division of Epidemiology, Norwegian Institute of Public Health

SOPHIA RABE-HESKETH

Graduate School of Education and Graduate Group in Biostatistics, University of California, Berkeley and Institute of Education, University of London

ABSTRACT. Latent variable modelling has gradually become an integral part of mainstream statistics and is currently used for a multitude of applications in different subject areas. Examples of 'traditional' latent variable models include latent class models, item–response models, common factor models, structural equation models, mixed or random effects models and covariate measurement error models. Although latent variables have widely different interpretations in different settings, the models have a very similar mathematical structure. This has been the impetus for the formulation of general modelling frameworks which accommodate a wide range of models. Recent developments include multilevel structural equation models with both continuous and discrete latent variables, multiprocess models and nonlinear latent variable models.

Key words: factor analysis, GLLAMM, item–response theory, latent class, latent trait, latent variable, measurement error, mixed effects model, multilevel model, random effect, structural equation model

1. Introduction

Latent variables are random variables whose realized values are hidden. Their properties must thus be inferred indirectly using a statistical model connecting the latent (unobserved) variables to observed variables. Somewhat unfortunately, latent variables are referred to by different names in different parts of statistics, examples including 'random effects', 'common factors', 'latent classes', 'underlying variables' and 'frailties'.

Latent variable modelling has gradually become an integral part of mainstream statistics and is currently used for a multitude of applications in different subject areas. Examples include, to name a few, longitudinal analysis (e.g. Verbeke & Molenberghs, 2000), covariate measurement error (e.g. Carroll *et al.*, 2006), multivariate survival (e.g. Hougaard, 2000), market segmentation (e.g. Wedel & Kamakura, 2000), psychometric measurement (e.g. McDonald, 1999), meta-analysis (e.g. Sutton *et al.*, 2000), capture–recapture (e.g. Coull & Agresti, 1999), discrete choice (e.g. Train, 2003), biometrical genetics (e.g. Neale & Cardon, 1992) and spatial statistics (e.g. Rue & Held, 2005).

Twenty-five years ago, the Danish statistician Erling B. Andersen published an important survey of latent variable modelling in the *Scandinavian Journal of Statistics* (Andersen, 1982). Andersen called his paper 'Latent Structure Analysis: A Survey', an aptly chosen title as his survey was confined to the type of models discussed by Lazarsfeld & Henry (1968) in their seminal book 'Latent Structure Analysis'. Specifically, Andersen focused on latent trait models (where the latent variable is continuous, whereas observed variables are categorical) popular in educational testing and latent class models (where both the latent variable and the observed variables are categorical) stemming from sociology, but some space was also

*This paper was presented at the 21st Nordic Conference on Mathematical Statistics, Rebild, Denmark, June 2006 (NORDSTAT 2006).

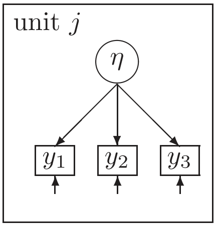

Fig. 1. Path diagram of simple latent variable model.

devoted to the rather esoteric latent profile models (where the latent variable is categorical, whereas the observed variables are continuous). Lazarsfeld & Henry (1968) mentioned factor models (where both latent and observed variables are continuous) only very briefly and, in line with this, Andersen excluded these models from his survey. This is remarkable as factor models are the most popular latent variable models in psychology.

The simple type of latent variable model considered by Andersen connects a single latent variable η_j for unit (e.g. person) j to observed variables $\mathbf{y}_j = (y_{1j}, y_{2j}, \ldots, y_{nj})'$. A basic construction principle of latent variable models (Arminger & Küsters, 1989) is the specification of conditional independence of the observed variables given the latent variable (see also McDonald, 1967; Lazarsfeld & Henry, 1968):

$$\Pr(\mathbf{y}_j \mid \eta_j) = \prod_{i=1}^{n} \Pr(y_{ij} \mid \eta_j).$$

This particular form of conditional independence is often called 'local independence'.

A very succinct representation of such a latent variable model is in terms of a path diagram as shown for three observed variables in Fig. 1. Circles represent latent variables, rectangles represent observed variables, arrows connecting circles and/or rectangles represent (linear or nonlinear) regressions, and short arrows pointing at circles or rectangles represent (not necessarily additive) residual variability.

A limitation of Andersen's survey was that only simple latent variable models with a single latent variable were considered. Nothing was said about multidimensional factor models or the important synthesis of common factor models and structural equation models pioneered by the fellow Scandinavian statistician Karl G. Jöreskog. Furthermore, mixed or random effects models, and covariate measurement error models were omitted from the survey. In the present survey, we fill these gaps and discuss some major subsequent developments in latent variable modelling.

The plan of the paper is as follows. We start by surveying more or less traditional latent variable models. We then discuss how different kinds of latent variable models have gradually converged by borrowing features from other models. Recognizing the similar mathematical structure of latent variable models, unifying frameworks for latent variable modelling have been developed and we describe one such framework. Finally, we survey some extensions of latent variable modelling such as multilevel structural equation models with both continuous and discrete latent variables, multiprocess models and nonlinear latent variable models.

2. Traditional latent variable models

The classification scheme of traditional latent variable models presented in Table 1 is useful to keep in mind in the sequel. We discuss latent class, item–response and factor models in the same section on measurement models because in all the three models the latent variables

Table 1. *Traditional latent variable models*

	Latent variable(s)	
Observed variable(s)	Continuous	Categorical
Continuous	Common factor model Structural equation model Linear mixed model Covariate measurement error model	Latent profile model
Categorical	Latent trait model/IRT	Latent class model

can be thought of as representing 'true' variables or constructs, and the observed variables as indirect or fallible measures. In the subsequent sections, we discuss structural equation models, linear mixed models and covariate measurement error models.

2.1. *Measurement models*

2.1.1. *Latent class models*

Consider $j=1,\ldots,N$ independent units. In latent class models, a latent categorical variable is measured with error by a set of categorical variables y_{ij}, $i=1,\ldots,n$. The categories of the latent variable represent labels for C subpopulations or latent classes, $c=1,\ldots,C$, with class membership probabilities π_c. To simplify the notation, we consider binary response variables here.

In the exploratory latent class model, the conditional response probability for measure i, given latent class membership c, is specified as:

$$\Pr(y_{ij}=1\,|\,c)=\pi_{i\,|\,c}, \tag{1}$$

where $\pi_{i\,|\,c}$ are free parameters. The responses y_{ij} and $y_{i'j}$ are conditionally independent given class membership.

As a function of the parameters $\boldsymbol{\pi}=(\pi_1,\pi_{1\,|\,1},\ldots,\pi_{n|1},\ldots,\pi_C,\pi_{1|C},\ldots,\pi_{n|C})'$, the marginal likelihood becomes:

$$l^{\mathrm{M}}(\boldsymbol{\pi})=\prod_{j=1}^{N}\Pr(\mathbf{y}_j;\boldsymbol{\pi})=\prod_{j=1}^{N}\sum_{c=1}^{C}\pi_c\prod_{i=1}^{n}\Pr(y_{ij}\,|\,c)=\prod_{j=1}^{N}\sum_{c=1}^{C}\pi_c\prod_{i=1}^{n}\pi_{i|c}^{y_{ij}}(1-\pi_{i|c})^{1-y_{ij}}.$$

It is evident that the latent class model is a multivariate finite mixture model with C components.

An important application of latent class models is in medical diagnosis where both latent classes (disease versus no disease) and the sets of measurements (diagnostic test results) are dichotomous (e.g. Rindskopf & Rindskopf, 1986).

A latent profile model for continuous responses has the same structure as an exploratory latent class model, but with a different conditional response distribution, $y_{ij}\,|\,c\sim\mathrm{N}(\mu_{i|c},\sigma_i^2)$.

Historical notes. The formalization of latent class models is due to Lazarsfeld (e.g. Lazarsfeld, 1950; Lazarsfeld & Henry, 1968), although models used as early as the nineteenth century (e.g. Peirce, 1884) can be viewed as special cases. Lazarsfeld also appears to have introduced the terms 'manifest' and 'latent' variables for observed and unobserved variables, respectively. A rigorous statistical treatment of latent class modelling was given by Goodman (1974a,b) who specified the models as log-linear models and considered maximum-likelihood estimation. The latent profile model was introduced by Green (1952), but the term was coined by Gibson (1959). An early reference to finite mixture models is Pearson (1894), even though earlier writings by Quetelet and other nineteenth century statisticians mention such approaches.

2.1.2. Item–response theory (IRT) models

Consider now the case where a latent continuous variable or 'latent trait' θ_j is measured with error by a set of categorical variables usually called items. The canonical example is from educational testing where the items are exam questions, y_{ij} is '1' if examinee j answered item i correctly and '0' otherwise, and θ_j represents the ability of the examinee.

One-parameter IRT model. In the simplest IRT model, the one-parameter logistic (1-PL) model, the conditional response probability for item i, given ability θ_j, is specified as:

$$\Pr(y_{ij}=1 \mid \theta_j)=\frac{\exp(\theta_j-b_i)}{1+\exp(\theta_j-b_i)}.$$

This model is called a 'one-parameter model' because there is one parameter, the 'item difficulty' b_i, for each item.

If it is assumed that ability is a random variable with a normal distribution, $\theta_j \sim \mathrm{N}(0,\psi)$, integrating out the latent variable produces the marginal likelihood:

$$\begin{aligned} l^{\mathrm{M}}(\mathbf{b},\psi) &= \prod_{j=1}^{N} \Pr(\mathbf{y}_j;\mathbf{b},\psi) \\ &= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \Pr(y_{ij} \mid \theta_j) g(\theta_j;\psi)\,\mathrm{d}\theta_j \\ &= \prod_{j=1}^{N} \int_{-\infty}^{\infty} \prod_{i=1}^{n} \frac{\exp(\theta_j-b_i)^{y_{ij}}}{1+\exp(\theta_j-b_i)} g(\theta_j;\psi)\,\mathrm{d}\theta_j, \end{aligned}$$

where $\mathbf{b}$ is the vector of item difficulties and $g(\cdot;\psi)$ is the normal density with zero mean and variance ψ.

Alternatively, the θ_j can be treated as unknown fixed parameters giving the so-called Rasch model. An incidental parameter problem (Neyman & Scott, 1948) occurs if 'incidental parameters' θ_j are estimated jointly with the 'structural parameters' $\mathbf{b}$, producing inconsistent estimators for $\mathbf{b}$. Inference can instead be based on a conditional likelihood $l^{\mathrm{C}}(\mathbf{b})$ constructed by conditioning on the sufficient statistic for θ_j; the sum score $t_j=\sum_{i=1}^{n} y_{ij}$. The conditional likelihood can be written as

$$l^{\mathrm{C}}(\mathbf{b}) = \prod_{j=1}^{N} \Pr(\mathbf{y}_j;\mathbf{b} \mid t_j) = \prod_{j=1}^{N} \frac{\prod_{i=1}^{n} \exp(-b_i y_{ij})}{\sum_{\mathbf{d}_j \in \mathcal{B}(t_j)} \prod_{i=1}^{n} \exp(-b_i d_{ij})},$$

where

$$\mathcal{B}(t_j)=\left\{ \mathbf{d}_j=(d_{1j},\ldots,d_{nj}) : d_{ij}=0 \text{ or } 1, \sum_{i=1}^{n} d_{ij}=t_j \right\}$$

is the set of all distinct sequences $\mathbf{d}_j$ of zeroes and ones with sum t_j. Clusters with $t_j=0$ or $t_j=n$ do not contribute to the likelihood as their conditional probabilities become 1.

An appealing feature of one-parameter models is that items and examinees can be placed on a common scale (according to the difficulty and ability parameters) so that the probability of a correct response depends only on the amount θ_j-b_i by which the examinee's position exceeds the item's position. Differences in difficulty between items are the same for all

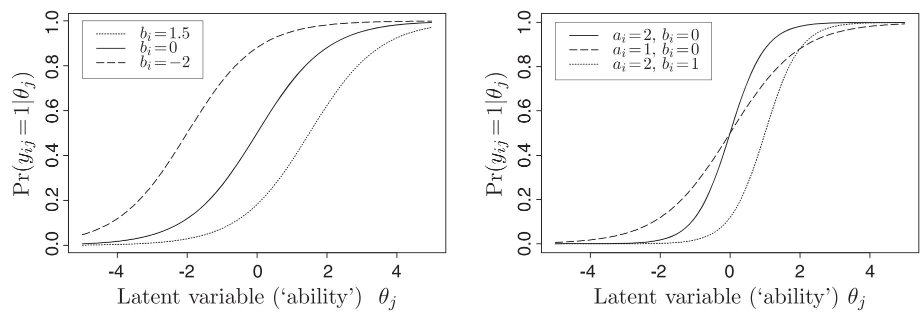

Fig. 2. Item characteristic curves for 1-PL (left) and 2-PL (right).

examinees, and differences in abilities of two examinees are the same for all items, a property called 'specific objectivity' by Rasch. For a given item, the probability of a correct response increases monotonically with ability as shown in the left panel of Fig. 2. Additionally, we see that for each ability, performance decreases with difficulty. The model is therefore said to exhibit 'double monotonicity'.

Two-parameter IRT model. Although the 1-PL model is simple and elegant, it is often deemed to be unrealistic in practice. A more general model, incorporating the 1-PL model as a special case, is the two-parameter logistic (2-PL) model specified as

$$\Pr(y_{ij}=1 \mid \theta_j)=\frac{\exp[a_i(\theta_j-b_i)]}{1+\exp[a_i(\theta_j-b_i)]}.$$

The model is called a 'two-parameter model' because there are two parameters for each item, a discrimination parameter a_i and a difficulty parameter b_i. The latent variable or ability is usually assumed to have a normal distribution $\theta_j \sim \mathrm{N}(0,1)$. It should be noted that a conditional likelihood can no longer be constructed as there is no sufficient statistic for θ_j and the marginal likelihood is hence used.

Item characteristic curves for the 2-PL model are shown for different difficulties and discrimination parameters in the right panel of Fig. 2. Note that the 2-PL model does not share the double monotonicity property of the 1-PL model. An item can be easier than another item for low abilities but more difficult than the other item for higher abilities due to the item–examinee interaction $a_i\theta_j$.

An alternative, and very similar model, is the normal ogive

$$\Pr(y_{ij}=1 \mid \theta_j)=\Phi(a_i(\theta_j-b_i)),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function.

More complex IRT models include the three-parameter logistic model (Birnbaum, 1968) which accommodates guessing, the partial credit (Masters, 1982), rating scale (Andrich, 1978) and graded response (Samejima, 1969) models for ordinal responses and the nominal response model (e.g. Rasch, 1961; Bock, 1972; Andersen, 1973).

Historical notes. The 'latent trait' terminology is due to Lazarsfeld (1954) whereas the term 'item–response theory' (IRT) was coined by Lord (1980). Lord was instrumental in developing statistical models for ability testing (e.g. Lord, 1952; Lord & Novick, 1968), mainly considering the now obsolete approach of joint maximum-likelihood estimation of abilities and item parameters (a_i and b_i). Two approaches are used to circumvent the incidental parameter problem. The Dane Georg Rasch (Rasch, 1960) suggested conditional maximum-likelihood estimation for the 1-PL model and Andersen (1970, 1972, 1973, 1980) provided a rigorous statistical treatment of this approach. Bock & Lieberman (1970) introduced marginal maximum-likelihood estimation and Gauss–Hermite quadrature in latent variable modelling, focusing on the two-parameter normal ogive model introduced by Lawley (1943) and Lord (1952).

2.1.3. Common factor models

A vector of continuous latent variables or 'common factors' ξ is indirectly observed via a vector of continuous observed variables, $\mathbf{x} = (x_1, x_2, \dots, x_q)'$. Note that the unit indices are usually suppressed in factor models.

The common factor model is usually written as

$$\mathbf{x} = \mathbf{\Lambda}_x \xi + \boldsymbol{\delta}. \tag{2}$$

The matrix $\mathbf{\Lambda}_x\,(q \times n)$ is a so-called factor loading matrix and the element of $\mathbf{\Lambda}_x$ pertaining to item i and latent variable l is denoted $\lambda_{il}^{(x)}$. The common factors have zero means $\mathrm{E}(\boldsymbol{\xi}) = \mathbf{0}$ and covariance matrix $\mathbf{\Phi}$. $\boldsymbol{\delta}$ are vectors of unique factors (specific factors and/or measurement errors) pertaining to the elements of $\mathbf{x}$, for which it is assumed that $\mathrm{E}(\boldsymbol{\delta}) = \mathbf{0}$. $\mathbf{\Theta}_\delta$ is the covariance matrix of $\boldsymbol{\delta}$, typically specified as diagonal with positive elements $\theta_{kk}^{(\delta)}$. It is finally assumed that $\mathrm{cov}(\boldsymbol{\delta}, \xi) = \mathbf{0}$.

It should be noted that it would be preferable to include a vector $\boldsymbol{\mu}$ of parameters for the expectations of $\mathbf{x}$ in the common factor model. This vector is usually omitted because, with complete data, it can be profiled out of the likelihood by substituting the sample means. $\mathbf{x}$ is in this case taken as mean centred.

The covariance structure of $\mathbf{x}$, called the 'factor structure', becomes

$$\mathbf{\Sigma} = \mathrm{cov}(\mathbf{x}) = \mathbf{\Lambda}_x \mathbf{\Phi} \mathbf{\Lambda}_x' + \mathbf{\Theta}_\delta. \tag{3}$$

Assuming multivariate normality for ξ and $\boldsymbol{\delta}$, and hence for $\mathbf{x}$, produces a marginal likelihood that can be expressed in closed form

$$l^{\mathrm{M}}(\mathbf{\Lambda}_x, \mathbf{\Phi}, \mathbf{\Theta}_\delta) = |2\pi\mathbf{\Sigma}|^{-n/2} \exp\left(-\frac{1}{2}\sum_{j=1}^{N} \mathbf{x}_j' \mathbf{\Sigma}^{-1} \mathbf{x}_j\right).$$

In the complete data case, the empirical covariance matrix $\mathbf{S}$ of $\mathbf{x}$ is the sufficient statistic for the parameters structuring $\mathbf{\Sigma}$ and has a Wishart distribution. Maximum-likelihood estimates can be obtained (e.g. Jöreskog, 1967) by minimizing the fitting function

$$F_{\mathrm{ML}} = \log|\mathbf{\Sigma}| + \mathrm{tr}(\mathbf{S}\mathbf{\Sigma}^{-1}) - \log|\mathbf{S}| - n,$$

with respect to the unknown free parameters. F_{ML} is non-negative and zero only if there is a perfect fit $\mathbf{\Sigma} = \mathbf{S}$.

The common factors are often interpreted as 'hypothetical constructs' measured indirectly by the observed variables $\mathbf{x}$. Hypothetical constructs are legion in the social sciences, an important example from psychology being the 'big-five theory of personality' (Costa & McCrae, 1985) comprising the constructs 'openness to experience', 'conscientiousness', 'extraversion', 'agreeableness' and 'neuroticism'. Friedman's (1957) notion of 'permanent income' is an example from economics.

It is important to distinguish between two approaches to common factor modelling: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is an inductive approach for 'discovering' the number of common factors and estimating the model parameters, imposing a minimal number of constraints for identification. The standard

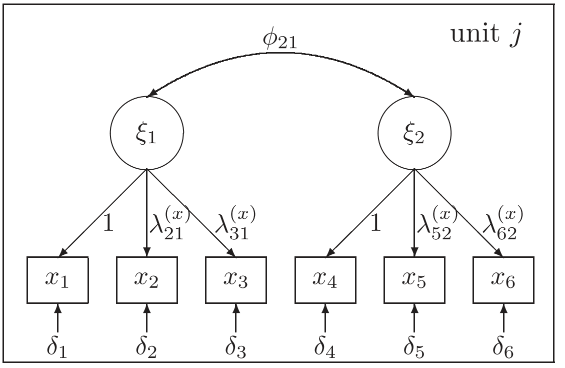

Fig. 3. Independent clusters factor model.

identifying constraints are that $\mathbf{\Phi} = \mathbf{I}$, and that $\mathbf{\Theta}_\delta$ and $\mathbf{\Lambda}_x' \mathbf{\Theta}_\delta^{-1} \mathbf{\Lambda}_x$ are both diagonal. Although mathematically convenient, the one-size-fits-all parameter restrictions imposed in EFA, particularly the specification of uncorrelated common factors, are often not meaningful from a subject matter point of view.

In CFA, restrictions are imposed based on substantive theory or research design. An important example is the 'independent clusters model' where $\mathbf{\Lambda}_x$ has many elements set to zero such that each variable measures one and only one factor. A path diagram of a two-factor independent clusters model with three variables measuring each factor is given in Fig. 3. Here the double-headed, curved error denotes a covariance between the two common factors.

The scales of the common factors are fixed either by 'factor standardization' where variances of the common factors are fixed to 1 (as in EFA) or by 'anchoring' where one factor loading for each factor is fixed to 1 as in the figure. Elements of $\boldsymbol{\delta}$ may be correlated, but special care must be exercised in this case to ensure that the model is identified.

In a unidimensional factor model, the common factor may represent a true variable measured with error, such as a person's true blood pressure fallibly measured by several measures. When measure-specific intercepts $\beta_i^{(x)}$ are included, we obtain the 'congeneric measurement model',

$$x_i = \beta_i^{(x)} + \lambda_i^{(x)} \xi + \delta_i.$$

The reliability ρ_i, the fraction of true score variance to total variance, for a particular measure x_i becomes

$$\rho_i = \left[(\lambda_i^{(x)})^2 \phi \right] \Big/ \left[(\lambda_i^{(x)})^2 \phi + \theta_{ii}^{(\delta)} \right].$$

The congeneric measurement model has several important nested special cases. The 'essentially tau-equivalent measurement model' prescribes that the measures are on the same scale

$$(\lambda_i^{(x)} = \lambda^{(x)}),$$

the 'tau-equivalent measurement model' that the measures also have the same means

$$(\lambda_i^{(x)} = \lambda^{(x)}, \beta_i^{(x)} = \beta^{(x)})$$

and the 'parallel measurement model' that the measurement error variances are also equal for all measures

$$(\lambda_i^{(x)} = \lambda^{(x)}, \beta_i^{(x)} = \beta^{(x)}, \theta_{ii}^{(\delta)} = \theta^{(\delta)}).$$

The parallel measurement model is also the model from classical (psychometric) test theory (e.g. Gulliksen, 1950).

Historical notes. The invention of factor analysis is attributed to Spearman (1904) who argued that intelligence was composed of a general factor, common to all subdomains such as mathematics, music, etc., and specific factors for each of the subdomains. However, the term factor analysis was introduced by Thurstone (1931), whose name, incidentally, is an anglicization of the Swedish name Torsten. Thurstone (1935, 1947) and Thomson (1938) introduced multidimensional EFA. Lawley (1939), Rao (1955), Anderson & Rubin (1956) and Lawley & Maxwell (1971) presented factor analysis as a statistical method. Anderson & Rubin (1956) and Bock & Bargmann (1966) anticipated CFA which was developed in a series of important papers by Jöreskog (e.g. 1969, 1971b).

2.2. Structural equation models (SEM) with latent variables

In the full structural equation or LInear Structural RELations (LISREL) model, the following relations are specified among continuous latent dependent variables $\boldsymbol{\eta}=(\eta_1, \eta_2, \ldots, \eta_m)'$ and continuous latent explanatory variables $\boldsymbol{\xi}=(\xi_1, \xi_2, \ldots, \xi_n)'$:

$$\boldsymbol{\eta}=\mathbf{B}\boldsymbol{\eta}+\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta}. \tag{4}$$

Here $\mathbf{B}$ is a matrix of structural parameters relating the latent dependent variables to each other, $\boldsymbol{\Gamma}$ is a matrix of structural parameters relating latent dependent variables to latent explanatory variables and $\boldsymbol{\zeta}$ is a vector of disturbances. We define $\boldsymbol{\Phi}=\text{cov}(\boldsymbol{\xi})$ and $\boldsymbol{\Psi}=\text{cov}(\boldsymbol{\zeta})$ and assume that $\text{E}(\boldsymbol{\xi})=\mathbf{0}$, $\text{E}(\boldsymbol{\zeta})=\mathbf{0}$ and $\text{cov}(\boldsymbol{\xi},\boldsymbol{\zeta})=\mathbf{0}$.

Confirmatory factor models are specified for $\boldsymbol{\xi}$ as in (2) and for $\boldsymbol{\eta}$ as

$$\mathbf{y}=\boldsymbol{\Lambda}_y\boldsymbol{\eta}+\boldsymbol{\epsilon}, \tag{5}$$

where $\mathbf{y}$ are continuous measures, $\boldsymbol{\Lambda}_y$ is the factor loading matrix and $\boldsymbol{\Theta}_\epsilon$ is the covariance matrix of $\boldsymbol{\epsilon}$. It is assumed that $\text{E}(\boldsymbol{\epsilon})=\mathbf{0}$, $\text{cov}(\boldsymbol{\epsilon},\boldsymbol{\eta})=\mathbf{0}$, $\text{cov}(\boldsymbol{\epsilon},\boldsymbol{\xi})=\mathbf{0}$, $\text{cov}(\boldsymbol{\delta},\boldsymbol{\xi})=\mathbf{0}$, $\text{cov}(\boldsymbol{\epsilon},\boldsymbol{\zeta})=\mathbf{0}$ and $\text{cov}(\boldsymbol{\epsilon},\boldsymbol{\delta})=\mathbf{0}$.

Substituting from the structural model in (5), and assuming that $\mathbf{I}-\mathbf{B}$ is of full rank, we obtain the reduced form for $\mathbf{y}$,

$$\mathbf{y}=\boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi}+\boldsymbol{\zeta})+\boldsymbol{\epsilon}, \tag{6}$$

whereas the reduced form for $\mathbf{x}$ is simply the common factor model in (2).

We denote the vector of all measures by $\mathbf{z}=(\mathbf{x}',\mathbf{y}')'$ and let $\boldsymbol{\Sigma}=\text{cov}(\mathbf{z})$. The covariance structure of $\boldsymbol{\Sigma}$ becomes

$$\boldsymbol{\Sigma}=\begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}'_{yx} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_y \end{pmatrix}, \tag{7}$$

where the submatrices are structured as

$$\boldsymbol{\Sigma}_x=\boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x+\boldsymbol{\Theta}_\delta, \tag{8}$$

$$\boldsymbol{\Sigma}_{yx}=\boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x, \tag{9}$$

$$\boldsymbol{\Sigma}_y=\boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}'+\boldsymbol{\Psi})[(\mathbf{I}-\mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_y+\boldsymbol{\Theta}_\epsilon. \tag{10}$$

Note that the common factor model (2) is a special case of the LISREL model.

An example of a LISREL model with two latent explanatory variables and two latent dependent variables is shown in Fig. 4. Research questions often concern the decomposition of the

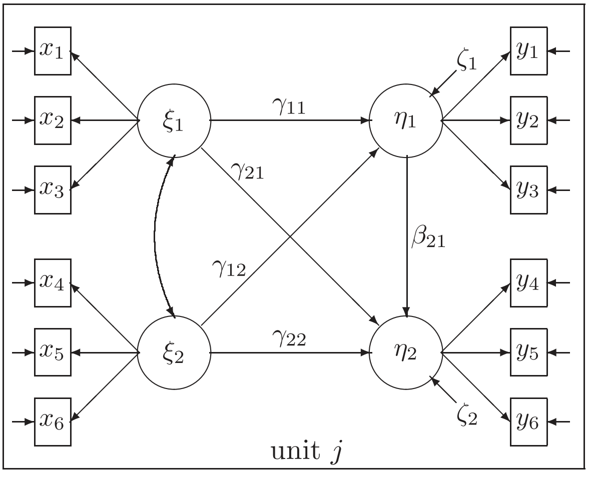

Fig. 4. Path diagram for structural equation model.

total effect of a latent variable on another latent variable. In the recursive structural equation model (without feedback effects) given in the figure, the total effect of ξ_2 on η_2 becomes the sum of a direct effect γ_{22} and an indirect effect $\gamma_{12}\beta_{21}$ (via the mediating latent variable η_1).

An important special case of the LISREL model is the Multiple-Indicator–MultIple-Cause (MIMIC) model where the latent variables are regressed on observed covariates **x**, whereas there are no regressions among the latent variables,

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\mathbf{x} + \boldsymbol{\zeta}. \qquad (11)$$

This model results if $\boldsymbol{\Lambda}_x = \mathbf{I}$ and $\boldsymbol{\Theta}_\delta = \mathbf{0}$ in the measurement model for the explanatory common factors in (2) so that $\mathbf{x} = \boldsymbol{\xi}$, and the restriction $\mathbf{B} = \mathbf{I}$ is imposed in the structural model (4).

The marginal likelihood of the LISREL model with multivariate normal $\boldsymbol{\xi}$, $\boldsymbol{\zeta}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$, and consequently multivariate normal **z**, can be expressed in closed form and a fitting function constructed analogously to the one for common factor models.

Historical notes. Structural equation modelling without latent variables, often called 'path analysis', was introduced by Wright (1918) and further developed as 'simultaneous equation modelling' by the Cowles Commission econometricians during World War II (e.g. Haavelmo, 1944). Early treatments of the MIMIC model include, for instance Zellner (1970) and Hauser & Goldberger (1971). The general LISREL framework was developed in a series of important contributions by Jöreskog (e.g. 1973a,b, 1977). This framework is undoubtedly the most popular, but Bentler & Weeks (1980) and McArdle & McDonald (1984), among others, have proposed alternative frameworks that essentially cover the same range of models.

2.3. Linear mixed models

Units are often nested in clusters, examples including pupils nested in schools or repeated measurements nested within individuals. In these situations, regression models can be extended to handle between-cluster variability, both in the overall level of the response and in the effects of covariates, that is not accounted for by observed covariates. Although the random effects in these models are clearly latent, these models have usually not been recognized as latent variable models.

Let $i=1,\ldots,n_j$ be the number of units nested in cluster $j=1,\ldots,N$. A linear mixed model can be written as

$$\mathbf{y}_j=\mathbf{X}_j\boldsymbol{\beta}+\mathbf{Z}_j\mathbf{b}_j+\mathbf{e}_j, \tag{12}$$

where $\mathbf{y}_j$ is an n_j-dimensional vector of continuous responses y_{ij} for cluster j, $\mathbf{X}_j$ an $n_j\times p$ matrix of covariates with fixed effects $\boldsymbol{\beta}$, $\mathbf{Z}_j$ an $n_j\times q$ matrix of covariates with random effects $\mathbf{b}_j$ and $\mathbf{e}_j$ a vector of residual errors.

The random effects and residuals are assumed to have multinormal distributions $\mathbf{b}_j\sim \mathrm{N}(\mathbf{0},\mathbf{G})$, $\mathbf{e}_j\sim \mathrm{N}(\mathbf{0},\mathbf{V}_j)$, both independent across clusters given the covariates and independent of each other. It is furthermore usually assumed that $\mathbf{V}_j=\sigma_e^2\mathbf{I}_{n_j}$. The covariates are assumed to be independent of the random effects and residuals.

The marginal expectation of the responses for cluster j, given the observed covariates but not the random effects, is

$$\boldsymbol{\mu}_j=\mathrm{E}(\mathbf{y}_j\,|\,\mathbf{X}_j,\mathbf{Z}_j)=\mathbf{X}_j\boldsymbol{\beta},$$

and the marginal covariance matrix becomes

$$\boldsymbol{\Sigma}_j=\mathrm{cov}(\mathbf{y}_j\,|\,\mathbf{X}_j,\mathbf{Z}_j)=\mathbf{Z}_j\mathbf{G}\mathbf{Z}_j'+\sigma_e^2\mathbf{I}_{n_j}.$$

As for common factor and structural equation models with continuous responses, the marginal likelihood can be expressed in closed form, giving

$$l^{\mathrm{M}}(\boldsymbol{\beta},\mathbf{G},\sigma_e^2)=|2\pi\boldsymbol{\Sigma}|^{-n_j/2}\exp\left\{-\frac{1}{2}\sum_{j=1}^{N}(\mathbf{y}_j-\boldsymbol{\mu}_j)'\boldsymbol{\Sigma}^{-1}(\mathbf{y}_j-\boldsymbol{\mu}_j)\right\}.$$

The most common linear mixed model is the random intercept model where there is only one random effect, a random intercept $b_j\sim \mathrm{N}(0,g)$,

$$\mathbf{y}_j=\mathbf{X}_j\boldsymbol{\beta}+\mathbf{1}_jb_j+\mathbf{e}_j. \tag{13}$$

This model induces a very simple dependence among responses within clusters, with constant residual intra-class correlation $g/(g+\sigma_e^2)$.

Removing the fixed part $\mathbf{X}_j\boldsymbol{\beta}$ in (13) gives the measurement model from classical test theory. This model is also known as a 'variance components model' or a 'one-way random effects model'. The latter term comes from analysis of variance where the clusters are viewed as levels of a factor (not to be confused with factors in factor analysis) which is treated as random.

Models with several clustering variables or factors can be specified where some factors are treated as fixed and others as random. Such models are used in 'generalizability theory' (Cronbach *et al.*, 1972) where the factors represent 'facets' (or aspects) of the measurement situation such as raters, temperatures, etc.

The linear mixed model in (12) can be viewed as a two-level model. Higher level models arise when the clusters j are themselves nested in superclusters k, etc., and when random effects are included at the corresponding levels.

Historical notes. Variance components models can be traced back to the works of astronomers such as Airy (1861) or even earlier. A milestone in the statistical literature was Fisher (1918) who implicitly employed variance components models. Eisenhart (1947) coined the term 'mixed model' for models with both random and fixed effects. Swamy (1970, 1971) introduced the linear mixed model under the name 'random coefficient model'. The work of Harville (1976, 1977) and Laird & Ware (1982) was important in introducing linear mixed models in statistics and biostatistics. Aitkin *et al.* (1981), Mason *et al.* (1984), Goldstein

(1987) and Bryk & Raudenbush (1992) popularized the models under the names 'multilevel' or 'hierarchical' models in social science. Generalizability theory was introduced by Cronbach *et al.* (1972).

2.4. *Covariate measurement error models*

It is often unrealistic to assume that the covariates in regression models are measured without error. Thus, it is useful to consider the regression model

$$y_j = \mathbf{x}_j'\boldsymbol{\beta}_x + \mathbf{z}_j'\boldsymbol{\beta}_z + \epsilon_j, \tag{14}$$

where y_j is the continuous response for unit j, $\mathbf{x}_j$ a vector of continuous covariates measured with error having regression parameters $\boldsymbol{\beta}_x$, $\mathbf{z}_j$ a vector of observed continuous covariates measured without error having regression parameters $\boldsymbol{\beta}_z$ and ϵ_j a residual.

The latent covariates $\mathbf{x}_j$ are measured with error according to the 'classical measurement model'

$$\mathbf{w}_j = \mathbf{x}_j + \mathbf{u}_j, \tag{15}$$

where $\mathbf{w}_j$ is a continuous vector of fallible (but observed) measures and $\mathbf{u}_j$ is a vector of measurement errors. It is assumed that $\mathrm{E}(\mathbf{u}_j) = \mathbf{0}$, $\mathrm{cov}(\mathbf{x}_j, \epsilon_j) = \mathbf{0}$, $\mathrm{cov}(\mathbf{z}_j, \epsilon_j) = \mathbf{0}$, $\mathrm{cov}(\mathbf{x}_j, \mathbf{u}_j) = \mathbf{0}$ and $\mathrm{cov}(\mathbf{u}_j, \epsilon_j) = \mathbf{0}$. It is also typically assumed that $\mathrm{cov}(\mathbf{u}_j)$ is diagonal, giving uncorrelated measurement errors, but that the diagonal elements may differ, allowing for different measurement error variances for the covariates.

Importantly, estimates of *both* $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_z$ are generally inconsistent if estimation is based on (14), with fallible regressors $\mathbf{w}_j$ substituted for $\mathbf{x}_j$. In the case of a single covariate x_j, the ordinary least squares estimate $\hat{\beta}_x$ is attenuated, but can be corrected by simply dividing it by the reliability of w_j. More generally, the regression parameters of interest $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_z$ can be estimated by maximizing the likelihood implied by (14) and (15) if multivariate normality is assumed for $\mathbf{x}_j$, $\mathbf{u}_j$ and ϵ_{ij}. This assumption is often relaxed by using instrumental variable estimation or specifying alternative, possibly non-parametric, distributions.

The classical measurement error model (15) must be distinguished from the 'Berkson measurement error model' where

$$\mathbf{x}_j = \mathbf{w}_j + \mathbf{u}_j, \tag{16}$$

with measurement errors $\mathbf{u}_j$ assumed to be independent of the true covariates $\mathbf{w}_j$. This model makes sense if the variables $\mathbf{w}_j$ are controlled variables, for instance, an experimenter may aim to administer given doses of some drugs $\mathbf{w}_j$, but the actual doses $\mathbf{x}_j$ differ due to measurement error. In this case, consistent estimates for the regression parameters can simply be obtained by estimating the model with fallible regressors.

Historical notes. The literature on covariate measurement error models, or 'errors in variables models' as they are often called in econometrics, can be traced back to Adcock (1877, 1878). Important contributions to the statistical literature include Durbin (1954) and Cochran (1968). The Berkson measurement error model was introduced by Berkson (1950). Instrumental variable estimation is due to Reiersøl (1945) and the term 'instrumental variables' was coined by the fellow Norwegian Ragnar Frisch. Carroll *et al.* (2006) distinguish between 'structural modelling' where parametric models are used for the distribution of the true covariates and 'functional modelling' where parametric models are specified for the response but no assumptions made regarding the distribution of the unobserved covariates.

3. Convergence of latent variable models

Some latent variable models have been extended by borrowing features of other models, leading to a convergence of models. For instance, starting from an item–response model and borrowing the idea of multidimensional latent variables from factor analysis yields the same model as starting from a factor model and modifying it for dichotomous responses.

In fact, the general structural equation or LISREL model discussed in section 2.2, can be viewed as a convergence or synthesis of CFA from psychometrics and simultaneous equation models from econometrics. This was quite a radical accomplishment at a time when, according to the econometrician Goldberger (1971, p. 83):

> Economists and psychologists have been developing their statistical techniques quite independently for many years. From time to time, a hardy soul strays across the frontier but is not met with cheers when he returns home.

Unfortunately, this statement applies equally well today and also to researchers from the biometric, econometric, psychometric and other statistical communities. For instance, biostatisticians typically attribute the invention of linear mixed models to Laird & Ware (1982), although Laird and Ware themselves referred to the work of the mathematical statistician Harville (1977). Interestingly, neither Harville nor Laird and Ware appeared to be aware of equivalent models introduced by the econometrician Swamy (1970, 1971) in an *Econometrica* paper and a Springer book with the title 'Statistical Inference in Random Coefficient Regression Models'. Even more remarkably, such a lack of communication is also evident *within* specific statistical communities. For instance, factor analysts and item–response theorists rarely cite each other, although their work is closely related and often published in the same journal, *Psychometrika*.

Probably due to this compartmentalization of statistics, model extensions that could have been 'borrowed' from other model types are often reinvented instead. In this section, we briefly review developments that we view as convergences, although the original authors did not necessarily view them in this way.

3.1. Common factor models and item–response models

The relationship between a unidimensional factor model and a 'normal ogive' (probit link) two-parameter item–response model was discussed by Lord & Novick (1968), but the convergence of these models gained momentum when Christofferson (1975) and Muthén (1978) extended common factor models to dichotomous responses.

They specified a traditional common factor model for an underlying (latent) continuous response y^*. In the unidimensional case, the model is typically written as

$$y^*_{ij} = \beta_i + \lambda_i \eta_j + \epsilon_{ij}, \quad \eta_j \sim \mathrm{N}(0, \psi), \quad \epsilon_{ij} \sim \mathrm{N}(0, 1), \quad \mathrm{cov}(\eta_j, \epsilon_{ij}) = 0, \quad \lambda_1 = 1, \tag{17}$$

where

$$y_{ij} = \begin{cases} 1, & \text{if } y^*_{ij} > 0, \\ 0, & \text{otherwise.} \end{cases}$$

This model is equivalent to a 'normal ogive' item response model as demonstrated by Bartholomew (1987) and Takane & de Leeuw (1987). To see this, consider the probability that y_{ij} equals 1, given the common factor,

$$\Pr(y_{ij} = 1 \mid \eta_j) = \Pr(y^*_{ij} > 0 \mid \eta_j) = \Phi(\beta_i + \lambda_i \eta_j) = \Phi(a_i(\theta_j - b_i)).$$

This is an item–response model where $b_i = -\beta_i/\lambda_i$, the factor loading λ_i corresponds to the discrimination parameter a_i and the common factor η_j to the 'ability' θ_j.

3.2. Structural equation models and item–response models

Muthén (1979) used the latent response formulation (17) to specify unidimensional MIMIC models with dichotomous indicators and estimated the models by maximum likelihood. Rather remarkably, equivalent models, albeit with a logit instead of a probit link, were only slowly introduced in the IRT literature.

Andersen & Madsen (1977), Andersen (1980b) and Mislevy (1983) proposed multigroup unidimensional item–response models where the ability mean (and variance) can differ between groups. Mislevy (1987) and Zwinderman (1991) specified what we call MIMIC models but without referring to Muthén's (1979) ground-breaking work, or to structural equation modelling.

Muthén (1983, 1984) specified a full LISREL-type structural equation model for dichotomous, ordinal and censored responses. Arminger & Küsters (1988, 1989) extended the approach further to include other response types such as counts and unordered categorical responses.

3.3. Structural equation models and linear mixed models

The similarity of linear mixed models and structural equation models with latent variables became obvious when both kinds of models were used to specify equivalent linear growth curve models.

Let the factor loading matrix $\boldsymbol{\Lambda}$ consist of a column of ones and a column of time-points t_i, $i=1,\ldots,n$, which must be the same for all subjects. Substituting the MIMIC model (11) in the confirmatory factor model (5), we obtain

$$\mathbf{y}_j = \boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x}_j + \boldsymbol{\Lambda}\boldsymbol{\zeta}_j + \boldsymbol{\epsilon}_j = \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j + \boldsymbol{\epsilon}_j,$$

where $\boldsymbol{\Lambda}\boldsymbol{\Gamma}\mathbf{x}_j$ corresponds to $\mathbf{X}_j\boldsymbol{\beta}$, both producing a vector of linear combinations of the covariates $\mathbf{x}_j$, $\boldsymbol{\Lambda}$ can be equated to $\mathbf{Z}_j$ because the elements are known constants, and $\boldsymbol{\zeta}_j$ corresponds to $\mathbf{b}_j$ in the linear mixed model (12).

Rao (1958) showed that growth models with subject-specific coefficients can be written as factor models where the factor loadings are known functions of time. McArdle (1988) and Meredith & Tisak (1990) extend this model by allowing the factor loadings to be estimated. The limitation that the time-points t_i must be the same for all subjects in the structural equation model formulation has recently been largely overcome by allowing for 'definition variables' which define the factor loadings (e.g. Mehta & Neale, 2005).

3.4. Other convergences

If multivariate normality is assumed for $\mathbf{x}$, the classical covariate measurement error model in (14) and (15) is a LISREL model with $\boldsymbol{\Lambda}_y = 1$ and $\boldsymbol{\Theta}_\epsilon = \mathbf{0}$ in the common factor model for the response factors, and $\boldsymbol{\Lambda}_x = \mathbf{I}$ in the model for the explanatory common factors.

Dayton & MacReady (1988) and Formann (1992) have combined traditional latent class models with a multinomial logit regression model for class membership, given observed covariates or so-called 'concomitant variables'. While this extension is analogous to a MIMIC model, the structural model is nonlinear, and there is no simple reduced form. Hagenaars (1993, 2002) and Vermunt (1997) extended the models further to allow latent class membership to depend on other discrete latent variables, coming close to a general structural equation model with categorical latent variables.

Linear mixed models have been extended to handle non-continuous responses (e.g. Heckman & Willis, 1976), giving so-called generalized linear mixed models (Gilmour *et al.*, 1985; Breslow & Clayton, 1993). One-parameter item–response models are just logistic random intercept models with a random intercept for subjects, separate fixed intercepts for each item, but no covariates. Fischer (1973) extended this model to include item covariates with fixed effects, and Rijmen *et al.* (2002) also include item covariates with random coefficients. De Boeck & Wilson (2004) consider a wide range of item–response models viewed from a mixed model perspective.

The connection between the Rasch model and log-linear models, which includes many latent class models as special cases, was demonstrated by Tjur (1982). Subsequently, de Leeuw & Verhelst (1986), Follmann (1988) and Lindsay *et al.* (1991) discussed semi-parametric maximum-likelihood estimation of the one-parameter logistic item–response model by treating the latent variable or ability distribution as discrete. The model with a discrete ability distribution can be viewed as a latent class model where all individuals within the same latent class have the same ability, and the latent classes are ordered along the ability dimension. Magidson & Vermunt (2001) discuss latent class factor models with several independent binary latent variables. The response model is that of a multidimensional common factor or item–response model.

'Mixture regression models' (e.g. Quandt, 1972) or 'latent class growth models' (e.g. Nagin & Land, 1993) are linear or generalized linear mixed models with categorical random effects. In a longitudinal setting, such models can be used to discover different 'latent trajectory classes'.

4. Modern frameworks for latent variable modelling

The move towards unifying latent variable models began with McDonald (1967) and Lazarsfeld & Henry (1968) who noted the common construction principle of conditional independence for the traditional measurement models discussed in section 2. A milestone in the development of frameworks for latent variable modelling was the advent of the general structural equation or LISREL model (e.g. Jöreskog, 1973) discussed in section 2.2, which can be viewed as a synthesis of CFA from psychometrics and simultaneous equation models from econometrics.

The LISREL model was subsequently generalized to handle other types of responses in addition to continuous responses. Probit measurement models for dichotomous, ordinal and censored responses were included in a series of important papers by Muthén (1983, 1984), and Arminger & Küsters (1988, 1989) extended the approach further to include other response types such as counts and unordered categorical responses. However, two important limitations of the LISREL framework and its generalizations are that all latent variables are continuous (and cannot be discrete) and that hierarchical or multilevel data can be handled only in the balanced case where each cluster includes the same number of units with the same covariate values.

Recognizing the mathematical similarity of a wide range of latent variable models, Muthén (2002), Rabe-Hesketh *et al.* (2004), Skrondal & Rabe-Hesketh (2004), among others, have developed very general frameworks for latent variable modelling which accommodate and extend the models mentioned in section 2. In this survey, we will focus on 'Generalized Linear Latent and Mixed Models' or GLLAMMs (Rabe-Hesketh *et al.*, 2004; Skrondal & Rabe-Hesketh, 2004) for two reasons. First, it is the framework with which we are most familiar. Secondly, and perhaps more importantly, because the GLLAMM model can be written down explicitly in its full generality just like the unifying LISREL model.

GLLAMMs consist of two building blocks: a response model and a structural model.

4.1. Response model

It has been recognized by Bartholomew (1980), Mellenbergh (1994) and others that, conditional on the latent variables, the response model of many latent variable models is a generalized linear model (McCullagh & Nelder, 1989). As for such models, the response model of GLLAMMs has three components: a link, a distribution and a linear predictor.

4.1.1. Links and conditional distributions

Let $\boldsymbol{\eta}$ be the vector of all latent variables in the model and let $\mathbf{x}$ and $\mathbf{z}$ denote vectors of covariates. The conditional expectation of the response y, given $\mathbf{x}$, $\mathbf{z}$ and $\boldsymbol{\eta}$, is 'linked' to the linear predictor ν (see section 4.1.2) via a link function $g(\cdot)$,

$$g(E[y \mid \mathbf{x}, \mathbf{z}, \boldsymbol{\eta}]) = \nu. \tag{18}$$

The specification is completed by choosing a conditional distribution for the response variable, given the latent variables and observed covariates.

Common combinations of links and distributions include: (i) the identity link and normal distribution for continuous responses; (ii) the logit, probit or complementary log–log link and Bernoulli distribution for dichotomous responses; (iii) the cumulative version of these links and multinomial distribution for ordinal responses (e.g. McCullagh, 1980); and (iv) the log link and Poisson distribution for counts.

For some response types, such as discrete or continuous time survival data, we can use the conventional links and distributions of generalized linear models in slightly non-conventional ways (e.g. Allison, 1982; Clayton, 1988). For polytomous responses or discrete choices and rankings, a multinomial logit link is used (Skrondal & Rabe-Hesketh, 2003) which is an extension of a generalized linear model. Very flexible links can be specified with a composite link function (e.g. Thompson & Baker, 1981; Rabe-Hesketh & Skrondal, 2007). Different links and conditional distributions can, furthermore, be specified for different responses. We will see applications of such mixed response models in section 5.3.

4.1.2. Linear predictor

We first consider a simplified version of the linear predictor and show in section 4.1.3 how this can be used to specify some of the models discussed in section 2. The linear predictor for item or unit i within cluster j can be written as

$$\nu_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj}\mathbf{z}'_{mij}\boldsymbol{\lambda}_m. \tag{19}$$

The elements of $\mathbf{x}_{ij}$ are covariates with 'fixed' effects $\boldsymbol{\beta}$. The mth latent variable η_{mj} is multiplied by a linear combination $\mathbf{z}'_{mij}\boldsymbol{\lambda}_m$ of covariates $\mathbf{z}_{mij}$ where $\boldsymbol{\lambda}_m$ are parameters (usually factor loadings – but see section 5.3).

For multilevel settings where clusters are nested in superclusters, etc., the linear predictor contains latent variables varying at the different levels $l = 2, \ldots, L$. To simplify notation, we will not use subscripts for the units of observation at the various levels. For a model with M_l latent variables at level l, the linear predictor has the form

$$\nu = \mathbf{x}'\boldsymbol{\beta} + \sum_{l=2}^{L}\sum_{m=1}^{M_l} \eta_m^{(l)}\mathbf{z}_m^{(l)\prime}\boldsymbol{\lambda}_m^{(l)}. \tag{20}$$

4.1.3. Some traditional latent variable models as special cases

Latent class models. The exploratory latent class model can be specified using (19) with $\mathbf{x}'_{ij}\boldsymbol{\beta} = 0$, $M = n$ latent variables η_{mj}, scalar dummy covariates

$$z_{mij} = \begin{cases} 1, & \text{if } i = m, \\ 0, & \text{otherwise,} \end{cases}$$

and scalar $\lambda_m = 1$:

$$\nu_{ij} = \sum_{m=1}^{n} \eta_{mj} z_{mij} = \mathbf{s}_i' \boldsymbol{\eta}_j = \eta_{ij},$$

where $\mathbf{s}_i$ is a selection vector containing a 1 in position i and 0 elsewhere. The latent variables are discrete with locations $\eta_{mj} = e_{mc}$ $(m = 1, \ldots, n; c = 1, \ldots, C)$ and masses π_c. Using a logit link gives the model in (1) with $\text{logit}(\pi_{i|c}) = e_{ic}$.

Item–response or common factor models. For unidimensional models $M = 1$, we drop the m subscript for η_{mj} and $\boldsymbol{\lambda}_m$ in (19) and the linear predictor for item i and unit j becomes

$$\nu_{ij} = \mathbf{s}_i' \boldsymbol{\beta} + \eta_j \mathbf{s}_i' \boldsymbol{\lambda} = \beta_i + \eta_j \lambda_i, \quad \lambda_1 = 1,$$

where λ_i is the ith element of $\boldsymbol{\lambda}$. Combined with an identity link and normal conditional distribution, this is a common factor model. Heteroscedastic unique factor standard deviations are specified as

$$\ln(\sqrt{\theta_{ii}}) = \mathbf{s}_i' \boldsymbol{\alpha},$$

so that $\theta_{ii} = \exp(2\alpha_i)$. Combined with a logit link and a Bernoulli conditional distribution, we get a 2-PL item-response model.

The multidimensional version of these models is produced by including $M > 1$ latent variables in the linear predictor. Let $\mathbf{a}_m$ be a vector containing the indices i of the items that measure the mth common factor. For instance, for the independent clusters two-factor model in Fig. 3, we specify $\mathbf{a}_1 = (1, 2, 3)'$ and $\mathbf{a}_2 = (4, 5, 6)'$. Also let $\mathbf{s}_i[\mathbf{a}_m]$ denote the corresponding elements of the selection vector $\mathbf{s}_i$. Then the linear predictor can be written as

$$\nu_{ij} = \mathbf{s}_i' \boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj} \mathbf{s}_i[\mathbf{a}_m]' \boldsymbol{\lambda}_m, \quad \lambda_{m1} = 1,$$

where λ_{m1} is the first element of $\boldsymbol{\lambda}_m$.

Mixed effects models. A mixed model with M random effects can be specified using scalar z_{mij} with $\lambda_m = 1$:

$$\nu_{ij} = \mathbf{x}_{ij}' \boldsymbol{\beta} + \sum_{m=1}^{M} \eta_{mj} z_{mij} = \mathbf{x}_{ij}' \boldsymbol{\beta} + \mathbf{z}_{ij}' \boldsymbol{\eta}_j,$$

where $\mathbf{z}_{ij} = (z_{1ij}, \ldots, z_{Mij})'$. This is linear mixed model if an identity link is combined with a normal conditional distribution and a generalized linear mixed model if other links and/or distributions are chosen.

4.2. Structural model for the latent variables

4.2.1. Continuous latent variables

The structural model for continuous latent variables $\boldsymbol{\eta} = (\boldsymbol{\eta}^{(2)\prime}, \cdots, \boldsymbol{\eta}^{(L)\prime})'$ has the form

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\mathbf{w} + \boldsymbol{\zeta}, \tag{21}$$

where $\mathbf{B}$ is an $M \times M$ regression parameter matrix ($M = \sum_l M_l$), $\mathbf{w}$ is a vector of observed covariates, $\boldsymbol{\Gamma}$ is a regression parameter matrix and $\boldsymbol{\zeta}$ is a vector of latent disturbances.

The structural model (21) resembles the structural part of the LISREL model in (4). A notational difference is that all latent variables are now denoted $\boldsymbol{\eta}$, with relations among all latent variables governed by the parameter matrix $\mathbf{B}$ and regressions of the latent variables on observed covariates governed by $\boldsymbol{\Gamma}$, a formulation previously used by Muthén (1984). Importantly, model (21) is a *multilevel* structural model where latent variables can vary at different hierarchical levels, thus including the conventional single-level structural model of Muthén as a special case.

Latent variables cannot be regressed on latent or observed variables varying at a lower level because such specifications do not make sense. Furthermore, the model is currently confined to recursive relations, not permitting feedback effects among the latent variables. The two restrictions together imply that the matrix $\mathbf{B}$ is strictly upper triangular if the elements of $\boldsymbol{\eta}^{(l)}$ are permuted appropriately, with elements of $\boldsymbol{\eta}=(\boldsymbol{\eta}^{(2)\prime}, \cdots, \boldsymbol{\eta}^{(L)\prime})'$ arranged in increasing order of l. Each element of $\boldsymbol{\zeta}$ varies at the same level as the corresponding element of $\boldsymbol{\eta}$.

Latent disturbances at the same level may be dependent, whereas latent disturbances at different levels are independent. We specify a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Psi}_l$ for the latent disturbances at level l.

Substituting the structural model into the response model, we obtain a reduced form model (which is nonlinear in the parameters). The reason for explicitly specifying a structural model is that for many models we are interested in the structural parameters and not in the reduced form parameters. For instance, in the LISREL model, $\boldsymbol{\Lambda}_y$, $\mathbf{B}$ and $\boldsymbol{\Gamma}$ have substantive interpretations, over and above the reduced form parameter matrix $\boldsymbol{\Lambda}_y(\mathbf{I}-\mathbf{B})^{-1}\boldsymbol{\Gamma}$ in (6). For models without a $\mathbf{B}$ or $\boldsymbol{\Lambda}$ matrix, such as linear mixed models or one-parameter item–response MIMIC models, the model can be written directly in reduced form.

4.2.2. *Categorical latent variables*

For discrete latent variables, the structural model is the model for the (prior) probabilities that the units belong to the different latent classes. For a unit j, let the probability of belonging to class c be denoted as $\pi_{jc}=\Pr(\boldsymbol{\eta}_j=\mathbf{e}_c)$. This probability may depend on covariates $\mathbf{w}_j$ through a multinomial logit model

$$\pi_{jc}=\frac{\exp(\mathbf{w}_j'\boldsymbol{\varrho}^c)}{\sum_d \exp(\mathbf{w}_j'\boldsymbol{\varrho}^d)}, \tag{22}$$

where $\boldsymbol{\varrho}^c$ are regression parameters with $\boldsymbol{\varrho}^1=\mathbf{0}$ imposed for identification.

5. Some model extensions

Having demonstrated the convergence of different kinds of traditional latent variable models and described a framework that unifies them, we now turn to latent variable models which are more complex than those discussed in section 2.

Many of the recent developments in latent variable modelling involve extending the models in at least one of the following ways: (i) allowing latent variables to vary at different hierarchical levels; (ii) combining continuous and discrete latent variables in the same model; (iii) accommodating multiple processes and mixed responses; and (iv) specifying nonlinear latent variable models.

5.1. *Multilevel latent variable models*

5.1.1. *Multilevel structural equation models*

In multilevel data, the units are nested in clusters, leading to within-cluster dependence. The traditional approach to extending structural equation models to the multilevel setting is to formulate separate within-cluster and between-cluster models (e.g. Longford & Muthén, 1992; Poon & Lee, 1992). Let $\mathbf{y}_{ik}$ be the vector of continuous responses for unit i in cluster k. Then the within-cluster model is

$$\mathbf{y}_{ik} \sim \mathrm{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_W(\boldsymbol{\vartheta}_W)), \tag{23}$$

where $\boldsymbol{\mu}_k$ is the mean vector for cluster k and the within-cluster covariance structure $\boldsymbol{\Sigma}_W(\boldsymbol{\vartheta}_W)$ is structured by a within-cluster structural equation model with parameters $\boldsymbol{\vartheta}_W$. The between-cluster model for the cluster means is

$$\boldsymbol{\mu}_k \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_B(\boldsymbol{\vartheta}_B)), \tag{24}$$

where $\boldsymbol{\mu}$ is the population mean vector and $\boldsymbol{\Sigma}_B(\boldsymbol{\vartheta}_B)$ is the between-cluster covariance structure structured by a between-cluster structural equation model with parameters $\boldsymbol{\vartheta}_B$.

An advantage of this set-up is that it allows software for conventional structural equation models to be 'tricked' into estimating the model. In particular, Muthén (1989) suggests an approach which corresponds to maximum likelihood for balanced data where all clusters have the same size n. This approach as well as an *ad hoc* solution for the unbalanced case are described in detail in Muthén (1994). Goldstein (2003) describes another *ad hoc* approach, using multivariate multilevel modelling to estimate $\boldsymbol{\Sigma}_W$ and $\boldsymbol{\Sigma}_B$ consistently by either maximum likelihood or restricted maximum likelihood. Structural equation models can then be fitted separately to each estimated matrix.

Unfortunately, the standard implementation of the within- and between-model approaches is limited in several ways. First, it is confined to continuous responses. Multilevel factor models for dichotomous and ordinal measures, or equivalently, multilevel item–response models, have been used by, for instance Ansari & Jedidi (2000), Rabe-Hesketh *et al.* (2004), Goldstein & Browne (2005) and Steele & Goldstein (2006). Raudenbush & Sampson (1999) and Maier (2001) estimate a multilevel 1-PL model which is equivalent to an ordinary multilevel logistic regression model. A second limitation of the standard within- and between-model is that it does not permit cross-level paths from latent or observed variables at a higher level to latent or observed variables at a lower level, although such effects will often be of primary interest. Thirdly, it does not allow for indicators varying at different levels. Fox & Glas (2001) and Rabe-Hesketh *et al.* (2004) overcome these limitations. For instance, the GLLAMM framework described in section 4 includes cross-level effects among latent variables via the **B** matrix in the structural model and can handle indicators and latent variables varying at an arbitrary number of levels in the response model.

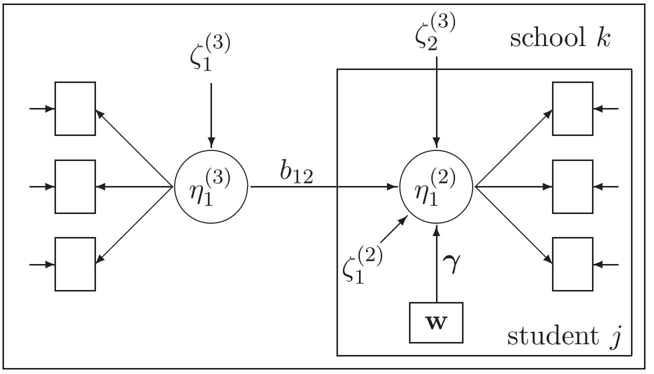

Fig. 5. Path diagram of multilevel structural equation model.

As an illustration of a model requiring these features, Fig. 5 shows a multilevel structural equation model considered by Rabe-Hesketh *et al.* (2007c). Student-specific variables are enclosed by the inner frame labelled 'student j', whereas variables outside this frame only vary over schools k. The ability of student j in school k is denoted $\eta_{1jk}^{(2)}$ and is measured by ordinal items. It is regressed on observed student-level covariates $\mathbf{w}_{jk}$ with regression parameter vector $\boldsymbol{\gamma}$ (the vector version of $\boldsymbol{\Gamma}$). Student-level ability is also regressed on a school-level latent variable 'teacher excellence', measured by school-level ordinal items, with cross-level regression parameter b_{12}. The random intercept $\zeta_{2k}^{(3)}$ represents the effects of unobserved school-level covariates on student ability.

5.1.2. Multilevel latent class models

Vermunt (2003) extends the traditional latent class model by generalizing (22) to the multilevel setting.

The item-level model is a conditional response model for item i, unit j, cluster k, given class membership c. For dichotomous responses, the model can be written as

$$\Pr(y_{ijk}=1 \mid \boldsymbol{\eta}_{jk}^{(2)}=\mathbf{e}_c)=\frac{\exp(\mathbf{e}_c)}{1+\exp(\mathbf{e}_c)}. \tag{25}$$

The unit-level model is a multinomial logit model for class membership,

$$\Pr(\boldsymbol{\eta}_{jk}^{(2)}=\mathbf{e}_c)=\frac{\exp(\alpha_{jk}^c)}{\sum_b \exp(\alpha_{jk}^b)},$$

where the linear predictor α_{jk}^c of the structural model includes a cluster-level random intercept $\eta_k^{(3)}$,

$$\alpha_{jk}^c=\mathbf{v}_{jk}'\boldsymbol{\varrho}^c+\gamma_c\eta_k^{(3)}.$$

Here, $\mathbf{v}_{jk}$ are unit- and cluster-specific covariates with fixed class-specific coefficients $\boldsymbol{\varrho}^c$ and γ_c are class-specific scaling constants. (An alternative specification uses mutually correlated class-specific random intercepts for each class but one.) A normal distribution is specified for the cluster-level random intercept.

Vermunt (2008) discusses an alternative model where the cluster-level random intercept is discrete.

5.2. Models with discrete and continuous latent variables

In section 5.2.1, we discuss latent variable models having a standard form but including both continuous and discrete latent variables. By contrast, the models surveyed in section 5.2.2 are finite mixtures of traditional latent variable models with parameters depending on class membership.

5.2.1. Continuous and discrete latent variables within otherwise traditional models

Discrete and continuous latent variables in response model. In a model for rankings, Böckenholt (2001) includes a discrete alternative-specific random intercept as well as continuous common factors and random coefficients. Similarly, McCulloch *et al.* (2002) specify a 'latent class mixed model' with both discrete and continuous random coefficients for joint modelling of continuous longitudinal responses and survival. The latent classes are interpreted as subpopulations differing both in their mean trajectories of (log) prostate-specific antigen and in their time to onset of prostate cancer. Variability among men within the same latent class is accommodated by the (continuous) random effects. Both Böckenholt (2001) and McCulloch *et al.* (2002) treat the discrete and continuous latent variables as independent of each other.

Discrete latent variables in response model and continuous latent variables in structural model. One of the models proposed by Vermunt (2003) which was described in section 5.1 includes a continuous cluster-level random intercept in the structural model for a discrete latent variable.

Continuous latent variables in response model and discrete latent variables in structural model. The response model could contain only continuous latent variables, whereas discrete latent variables appear in the structural model. A simple structural model would have the form

$$\boldsymbol{\eta}_{jc} = \mathbf{e}_c + \boldsymbol{\zeta}_j, \quad \boldsymbol{\zeta}_j \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Psi}), \tag{26}$$

whereas more complex structural models could have class-specific residual covariance matrices $\boldsymbol{\Psi}_c$. These structural models are just finite mixtures of normal distributions. Such a model was used by Verbeke & Lesaffre (1996), Allenby *et al.* (1998) and Lenk & DeSarbo (2000) for random coefficient models, by Magder & Zeger (1996), Carroll *et al.* (1999) and Richardson *et al.* (2002) for covariate measurement error models, and Uebersax (1993) and Uebersax & Grove (1993) for measurement models with dichotomous and ordinal responses.

5.2.2. *Mixtures of traditional latent variable models*

Mixture structural equation models with continuous responses. In these models, any parameter of a conventional structural equation model with continuous responses can depend on discrete latent variables. Both the response model and the structural model can therefore differ between latent classes, giving a multiple-group structural equation model of the kind proposed by Jöreskog (1971a), with the crucial difference that group membership is unknown.

Yung (1997) and Fokoué & Titterington (2003), among others, consider the special case of finite mixture factor models. Yung's model can be written as

$$\mathbf{y}_{jc} = \boldsymbol{\beta}_c + \boldsymbol{\Lambda}_c \boldsymbol{\eta}_{jc} + \boldsymbol{\epsilon}_{jc}, \tag{27}$$

where $\boldsymbol{\eta}_{jc}$ are continuous common factors with class-specific variances $\Psi_c = \mathrm{cov}(\boldsymbol{\eta}_{jc})$ and the unique factors have class-specific covariance matrices $\boldsymbol{\Theta}_c = \mathrm{cov}(\boldsymbol{\epsilon}_{jc})$. Fokoué and Titterington assume that $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}$ and $\Psi_c = \mathbf{I}$. In the context of diagnostic test agreement, Qu *et al.* (1996) specify a unidimensional probit version of this model. They interpret the model as a latent class model with a 'random effect' (the common factor) to relax conditional independence.

Blåfield (1980), Jedidi *et al.* (1997), Dolan & van der Maas (1998), Arminger *et al.* (1999), McLachlan & Peel (2000), Wedel & Kamakura (2000), Muthén (2002) and others specify 'finite mixture structural equation models' by including a structural model

$$\boldsymbol{\eta}_{jc} = \mathbf{B}_c \boldsymbol{\eta}_{jc} + \boldsymbol{\Gamma}_c \mathbf{w}_{jc} + \boldsymbol{\zeta}_{jc}, \quad \Psi_c \equiv \mathrm{cov}(\boldsymbol{\zeta}_{jc})$$

for the factors in (27).

Mixture IRT. Rost (1990) specifies a 'mixed Rasch model' with class-specific difficulty parameters. By using conditional maximum-likelihood estimation for the item parameters of each class in the M step of an EM algorithm (which treats class membership as missing), he avoids making assumptions regarding the ability distribution in each class. By contrast, Mislevy &

Verhelst (1990) and Mislevy & Wilson (1996) specify normal ability distributions, the latter paper with class-specific means and variances.

Growth mixture models. Muthén *et al.* (2002) consider linear mixed models where all parameters are class specific. The models can be thought of as a special case of mixture structural equation models. Alternatively, they can be viewed as extensions of mixture regression models to allow for variability of the coefficients within latent classes instead of assuming that all individuals in the same class have the same coefficients.

5.3. Multiprocess and mixed response models

In multiprocess models, several distinct processes, such as survival and repeated measures, are modelled simultaneously to allow for dependence among them. Such models usually involve responses of mixed types, a challenge that has been addressed by, for instance Muthén (1984), Arminger & Küsters (1988, 1989), Moustaki (1996), Sammel *et al.* (1997), Bartholomew & Knott (1999), Rabe-Hesketh *et al.* (2004) and Skrondal & Rabe-Hesketh (2004). In the next sections, we discuss some common types of multiprocess models.

5.3.1. Covariate measurement error models

We first consider the problem of estimating a regression model for the relationship between an outcome y_j and covariates $\mathbf{x}_j$ and η_j,

$$g(\mathrm{E}(y_j \mid \mathbf{x}_j, \eta_j)) = \mathbf{x}_j'\boldsymbol{\beta} + \eta_j\lambda_y, \tag{28}$$

where the latent covariate η_j has been fallibly measured as v_{ij} at possibly multiple occasions i.

A model for this problem can be constructed by combining three submodels (e.g. Clayton, 1992): (i) an outcome model (called 'disease model' in epidemiology) such as that shown above, (ii) a measurement model,

$$g(\mathrm{E}(v_{ij} \mid \eta_j)) = \eta_j\lambda_i, \tag{29}$$

and (iii) an exposure model,

$$\eta_j = \boldsymbol{\gamma}'\mathbf{x}_j + \zeta_j, \quad \zeta_j \sim \mathrm{N}(0, \psi). \tag{30}$$

Here, we have used the same covariate vector $\mathbf{x}_j$ as in the outcome model, but some elements of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ may be zero. As usual, the measurements v_{ij} are conditionally independent of one another given η_j. An important assumption, known as non-differential measurement error, is that v_{ij} are also conditionally independent of the outcome y_j, given the latent covariate.

Note that the above covariate measurement error model represents a generalization of the conventional covariate measurement error model described in section 2.4 to include a structural model for exposure and a generalization of the conventional MIMIC model described in section 2.2 to mixed responses.

The basic structure of the model is perhaps best conveyed in a path diagram as shown in the upper left panel of Fig. 6, where there are two measurements v_{1j} and v_{2j} of the latent covariate η_j.

5.3.2. Models for partially missing covariates

In some cases, we have a validation sample with complete data as shown in the upper right panel of Fig. 6 where η_j is observed (denoted by placing η_j in a rectangle instead of a circle).

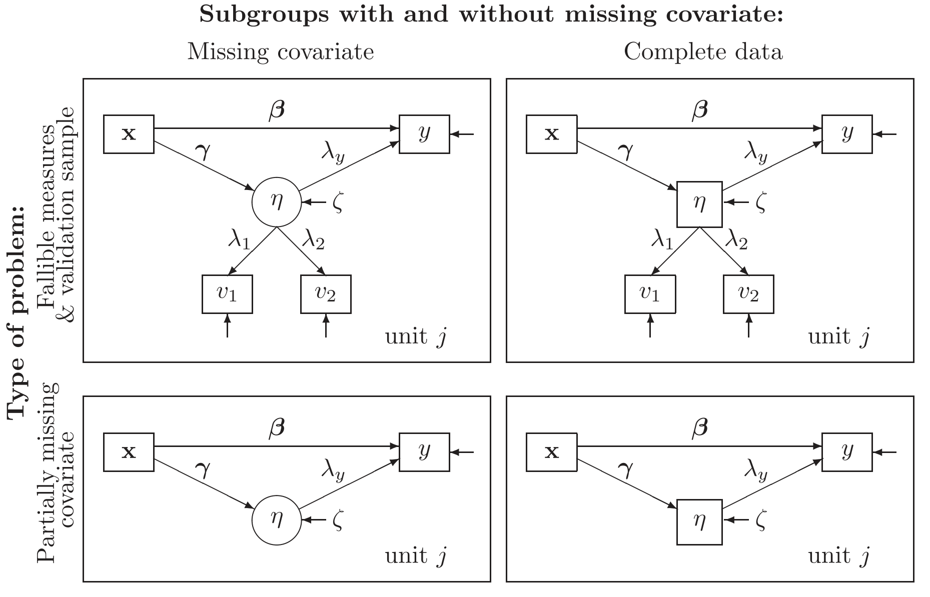

Fig. 6. Fallibly measured and/or partially missing covariates.

The validation sample is usually small because measuring η_j directly is either expensive, intrusive or invasive. In this case, we can exploit the information in the larger sample for which only the fallible measures are available by considering a joint model with equality restrictions imposed on the parameters across samples. Note that replicate measures are not required in either sample in this case.

A similar idea can be used for partially observed covariates; see the two lower panels of Fig. 6. Here, we combine information from a sample with an observed covariate η_j (right panel) with a sample where the covariate is missing (left panel).

If the latent variable is categorical, the model is a latent class model. Such a model can be used for analysing data from randomized studies with non-compliance, where some individuals are not receiving the treatments they are randomized to receive. One way of handling the problem is to study the *effectiveness* of the treatment by comparing groups as randomized regardless of compliance (intention-to-treat). However, one is often interested in investigating *efficacy*, the effect of the treatment among those taking the treatment. To compare like with like it seems judicious to compare those taking the treatment in the active treatment group with those in control group who would have taken the treatment (latent compliance) given the opportunity, known as the *complier average causal effect* (CACE).

Complier average causal effects can be investigated by a latent class model with 'training data' available for the treatment group (e.g. Muthén, 2002; Skrondal & Rabe-Hesketh, 2004). Such a model, where η_j is dichotomous, is presented in the lower part of Fig. 6 where the left panel represents the model for the control group (where compliance η_j is latent) and the right panel the model for the treatment group (where compliance η_j is observed). We could of course have different covariates affecting compliance and the outcome.

5.3.3. Endogenous covariate models

In contrast to randomized studies, a major problem in estimating treatment or exposure effects from observational studies is that a covariate or 'treatment' may be endogenous, in the

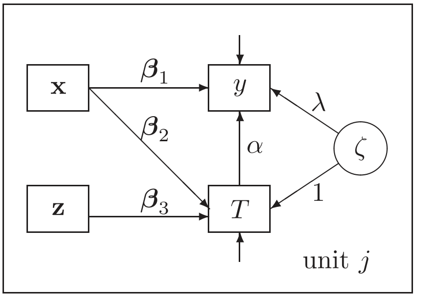

Fig. 7. Path diagram for endogenous treatment model.

sense that it is confounded with unobserved covariates affecting the outcome. For instance, subjects with a poor prognosis (unobserved heterogeneity) may self-select into the treatment perceived to be best. It is hardly surprising that conventional modelling in this case can produce severely biased estimates of the treatment effect. Hence, models attempting to correct for selection bias by jointly modelling the outcome and treatment processes, often called endogenous treatment models, have been suggested in econometrics (e.g. Heckman, 1978).

The outcome model for y_j is specified as

$$g(\mathrm{E}(y_j \mid \mathbf{x}_j, \eta_j)) = \alpha T_j + \mathbf{x}_j' \boldsymbol{\beta}_1 + \lambda \zeta_j,$$

where T_j is a dummy for treatment with corresponding parameter α and $\mathbf{x}_j$ a vector of covariates with parameter vector $\boldsymbol{\beta}_1$. $\zeta_j \sim \mathrm{N}(0, \psi)$ is a latent variable representing unobserved heterogeneity that is shared between the outcome and treatment processes and λ is a factor loading. An appropriate link and distribution are specified according to the type of response y_j.

The treatment model for T_j is specified as

$$g(\mathrm{E}(T_j \mid \mathbf{x}_j, \mathbf{w}_j, \eta_j)) = \mathbf{x}_j' \boldsymbol{\beta}_2 + \mathbf{z}_j' \boldsymbol{\beta}_3 + \zeta_j,$$

where $\boldsymbol{\beta}_2$ is the regression parameter vector for the covariates $\mathbf{x}_j$ that are also included in the outcome model and $\boldsymbol{\beta}_3$ is the regression parameter vector for covariates $\mathbf{z}_j$ which are not included in the outcome model. For the treatment model, a probit link and a Bernoulli distribution are typically specified.

Different types of parameter constraints are necessary for identification depending on the type of response y_i. For a dichotomous or continuous response y_i, we set $\psi = 1$. For a continuous response, we also use a scaled probit link for the treatment model with scale set equal to the residual standard deviation of y_i (see Skrondal & Rabe-Hesketh, 2004, p. 108).

A path diagram of a model with an endogenous covariate is shown in Fig. 7. Note that the endogenous treatment model reduces to the famous Heckman selection model (e.g. Heckman, 1979) if $\alpha = 0$ and the variable T_j plays the role of sample inclusion indicator.

5.3.4. Models for non-ignorable dropout

Intermittent missingness and dropout are fundamental problems with longitudinal data. If responses are missing for some units, these units can still contribute to parameter estimation as long as there is at least one observed response, leading to consistent parameter estimates if the data are missing at random (MAR) (e.g. Rubin, 1976; Little & Rubin, 2002) and the model is correctly specified. If the responses are not missing at random (NMAR), ignoring the missing data mechanism will, in general, lead to inconsistent parameter estimates. However, valid statistical inference can be achieved if a correctly specified joint model for the missingness and substantive processes is used.

Little (1993, 1995) distinguishes between two broad classes of joint models, 'selection models' and 'pattern mixture models'. In pattern mixture models, the distribution of the longitudinal responses is specified conditional on the missingness pattern, whereas missingness depends on covariates only. Here, we concentrate on selection models where dropout may depend either on the missing responses, called 'non-ignorable outcome-based dropout', or on random effects in the longitudinal model, called 'non-ignorable random coefficient-based dropout' by Little (1995).

As emphasized by Hogan & Laird (1997), the same models can be used when survival (a kind of dropout) is of primary interest and the longitudinal data represent time-varying covariates.

Outcome-based dropout. In econometrics, Hausman & Wise (1979) proposed a model for dropout or 'attrition' that was later reinvented in the statistical literature by Diggle & Kenward (1994). Here, the dropout at each time-point (given that it has not yet occurred) is modelled using a logistic (or probit) regression model with previous responses and the contemporaneous response as covariates. If dropout occurs, the contemporaneous response is not observed but represented by a latent variable y^*_{ij}. The marginal likelihood is the joint likelihood of the response and dropout status after integrating out the latent variable.

A simple version of the model for the response of interest y_{ij} at time i for unit j can be written as

$$y_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta} + \eta_j + \epsilon_{ij},$$

where η_j is a unit-specific random intercept. The dropout variable, $d_{ij} = 1$ if unit j drops out at time i and 0 otherwise, can be modelled as

$$\text{logit}\{\Pr(d_{ij} = 1 \mid \mathbf{y}_j)\} = \alpha_1 y_{ij} + \alpha_2 y_{i-1,j},$$

where y_{ij} is replaced by a latent variable y^*_{ij} when it is unobserved. Figure 8 shows path diagrams for a unit with complete data across three time-points and a unit dropping out at time 2. In the latter case, the response at time 2 is represented by a latent variable y^*_2.

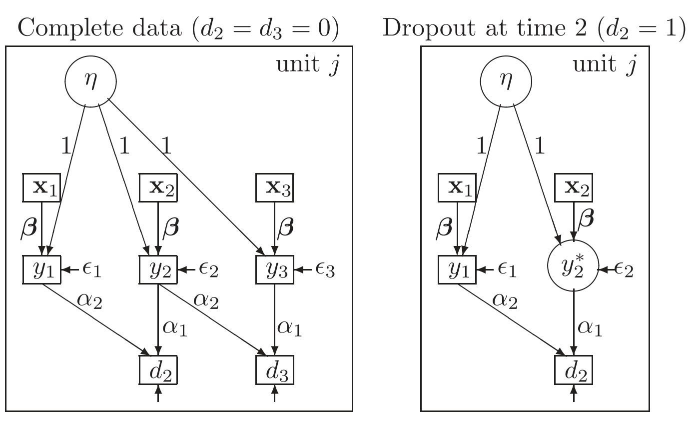

Fig. 8. Path diagram of Hausman–Wise–Diggle–Kenward dropout model.

This approach can be viewed literally as correcting for missing data that are NMAR. However, as the estimation of α_1 relies heavily on the distributional assumption for y_{ij}, it might be advisable to instead interpret the results as a sensitivity analysis of the MAR assumption.

Random coefficient-based dropout. Wu & Carroll (1988) consider a linear growth model for the longitudinal data with a subject-specific random intercept b_{0j} and slope of time b_{1j}. The term $\alpha_0 b_{0j} + \alpha_1 b_{1j}$ is included in the dropout model, where α_0 and α_1 are unknown parameters that could be viewed as factor loadings. This type of model has also been referred to as a 'shared parameter model' by Follmann & Wu (1995).

Another specification is to treat the subject-specific mean $b_{0j} + b_{1j}t_{ij}$ in the linear mixed model as a covariate in the dropout model, interpreting the subject-specific mean as the 'latent process' or true response with measurement error removed (e.g. Faucett & Thomas, 1996; Xu & Zeger, 2001).

Ten Have *et al.* (2000) specify a two-level unidimensional common factor model for a multivariate dichotomous response with a subject-specific random intercept which is also included in the dropout model.

5.4. Nonlinear latent variable models

In this section, we consider models that are nonlinear in the latent variables and possibly in the parameters.

5.4.1. Nonlinear mixed models

Nonlinear mixed models are used either because theory prescribes a particular functional form, as, for instance in compartment models for drug absorption, distribution and elimination (see below) or because the relationship between the response and explanatory variable(s) cannot be well approximated by a polynomial, due to asymptotes, inflection points or other features.

The models usually have the form

$$\begin{aligned} y_{ij} &= f(\mathbf{t}_{ij}, \boldsymbol{\eta}_j) + \epsilon_{ij}, \quad \epsilon_{ij} \sim \mathrm{N}(0, \sigma^2) \\ \boldsymbol{\eta}_j &= \mathbf{X}_j\boldsymbol{\beta} + \mathbf{Z}_j\mathbf{b}_j, \quad \mathbf{b}_j \sim \mathrm{N}(\mathbf{0}, \mathbf{G}), \end{aligned} \tag{31}$$

where $f(\cdot)$ is some nonlinear function, $\mathbf{t}_{ij}$ is a covariate vector, $\mathbf{X}_j$ and $\mathbf{Z}_j$ are covariate matrices, $\boldsymbol{\beta}$ are fixed effects and $\mathbf{b}_j$ are random effects.

Davidian & Giltinan (2003) discuss the following one-compartment model for the blood concentration of an anti-asthmatic agent t_{ij} time units after the administration of the drug,

$$y_{ij} = \frac{Dk_j}{V_j(k_j - C_j/V_j)}\{\exp(-k_j t_{ij}) - \exp(-C_i t_{ij}/V_j)\} + \epsilon_{ij},$$

where k_j is the fractional rate of absorption from the gut into the bloodstream, V_j is roughly the volume required to account for all drug in the body and C_j is the clearance rate, representing the volume of blood from which the drug is eliminated per time unit. The model for the vector of subject-specific random parameters $\boldsymbol{\eta}_j = (\ln k_j, \ln V_j, \ln C_j)'$ has the form of (31) with $\mathbf{Z}_j = \mathbf{I}_3$. We refer to Davidian & Giltinan (2003) for a useful recent overview of nonlinear mixed effects models.

5.4.2. Nonlinear factor models

Nonlinear factor models for continuous responses (e.g. Bartlett, 1953; McDonald, 1967; Yalcin & Amemiya, 2001) have been proposed to relax the standard linearity specification. The vector of responses $\mathbf{x}_j$ is modelled as

$$\mathbf{x}_j = \boldsymbol{\Lambda}_x \boldsymbol{\alpha}_j + \boldsymbol{\delta}_j, \quad \boldsymbol{\alpha}_j \equiv \mathbf{g}(\boldsymbol{\xi}_j),$$

where $\mathbf{g}(\boldsymbol{\xi}_j)$ is a known deterministic vector function and normality is often assumed for $\boldsymbol{\xi}_j$ and $\boldsymbol{\delta}_j$. Special cases include polynomial factor models where $\boldsymbol{\alpha}_j = (\xi_j, \xi_j^2, \ldots, \xi_j^p)'$, and models with (first-order) interactions where

$$\boldsymbol{\alpha}_j = (\xi_{1j}, \xi_{2j}, \ldots, \xi_{qj}, \xi_{1j}\xi_{2j}, \xi_{1j}\xi_{3j}, \ldots, \xi_{q-1,j}\xi_{qj})'.$$

For example, Etezadi-Amoli & McDonald (1983) consider a two-factor model for aphasic dysfunction where the first factor is interpreted as a general factor and the second as a verbal–non-verbal bipolar factor. The responses are a cubic function of the first factor, a linear function of the second factor and the interaction between the two factors.

A novel idea would be to extend multilevel structural equation models to include products of latent variables varying at different levels. For instance, consider a latent variable $\eta_{jk}^{(2)}$ for unit j in cluster k and a latent variable $\eta_k^{(3)}$ for cluster k. A product term such as $\eta_k^{(3)}\eta_{jk}^{(2)}$ could then represent either (i) a cross-level interaction of latent covariates $\eta_k^{(3)}$ and $\eta_{jk}^{(2)}$; (ii) a random coefficient $\eta_k^{(3)}$ of a latent covariate $\eta_{jk}^{(2)}$; or (iii) a latent variable $\eta_{jk}^{(2)}$ with random, cluster-specific, standard deviation $\eta_k^{(3)}$.

5.4.3. Nonlinear structural equation models

Structural models that are nonlinear in the latent variables have also been proposed (e.g. Kenny & Judd, 1984). Arminger & Muthén (1998) and Lee & Zhu (2002), among others, discuss nonlinear versions of the LISREL model for continuous responses (see section 2.2). The structural model for latent response variables $\boldsymbol{\eta}_j$ in terms of the latent explanatory variables $\boldsymbol{\xi}_j$ is given by

$$\boldsymbol{\eta}_j = \mathbf{B}\boldsymbol{\eta}_j + \boldsymbol{\Gamma}\boldsymbol{\alpha}_j + \boldsymbol{\zeta}_j, \quad \boldsymbol{\alpha}_j \equiv \mathbf{g}(\boldsymbol{\xi}_j).$$

As for nonlinear factor models, $\mathbf{g}(\boldsymbol{\xi}_j)$ is a known deterministic vector function and normality is typically assumed for $\boldsymbol{\xi}_j$ and $\boldsymbol{\zeta}_j$. An example of a nonlinear relationship between latent variables is discussed by Wall & Amemiya (2000) where the effect of parenting skills on children's self-restraint is large for low parenting skills and eventually levels off. Wall & Amemiya (2007) provide a recent overview of nonlinear structural equation modelling.

6. Concluding remarks

Although our survey has been quite extensive, reflecting the major developments since Andersen (1982), we do not claim that it is exhaustive. For instance, we have omitted model types such as state-space models (e.g. Jones, 1993; Durbin & Koopman, 2001), Bayesian models (e.g. Fox & Glas, 2001; Lee & Song, 2003; Dunson & Herring, 2005) and spatial models (e.g. Knorr-Held & Best, 2001; Christensen & Amemiya, 2003; Liu *et al.*, 2005).

We have not discussed identification and equivalence which are important challenges in latent variable modelling (e.g. Dupacŏvá & Wold, 1982; Rabe-Hesketh & Skrondal, 2001; Skrondal & Rabe-Hesketh, 2004, Chapter 5). Roughly speaking, identification concerns whether the parameters of a specific model are unique in the sense that there is only one set of parameter values that can produce a given probability distribution, whereas equivalence concerns whether different models can produce identical probability distributions.

Regarding models with continuous latent variables, we have focused on the standard multivariate normal case. Approaches to relaxing this assumption include using multivariate t-distributions (e.g. Pinheiro *et al.*, 2001), finite mixtures of normal distributions (e.g. Uebersax, 1993; Magder & Zeger, 1996), truncated Hermite series expansions (e.g. Gallant &

Nychka, 1987; Davidian & Gallant, 1992), non-parametric maximum-likelihood estimation (e.g. Lindsay, 1995), other non-normal distributions (e.g. Lee & Nelder, 1996; Wedel & Kamakura, 2001; Lee *et al.*, 2006) or, in a Bayesian setting, semiparametric mixtures of Dirichlet processes (e.g. Müller & Roeder, 1997).

Finally, we have not discussed the estimation of latent variable models and therefore give a very brief outline here. When the marginal distribution of the responses is multivariate normal, maximum-likelihood estimation is relatively straightforward because the marginal likelihood can be written in closed form. In non-normal models, such at item–response models, there is generally no analytic expression for the likelihood (exceptions being models with conjugate latent variable distributions). In this case, the integrals are typically approximated in one of three ways: (i) Laplace approximation and penalized quasi likelihood (e.g. Breslow & Clayton, 1993); (ii) numerical integration (e.g. Bock & Lieberman, 1970; Rabe-Hesketh *et al.*, 2005); or (iii) Monte Carlo integration (e.g. Meng & Schilling, 1996; Train, 2003; Lee & Song, 2004). For models with multivariate normal latent responses, such as that shown in (17), Muthén (1984) and Muthén & Satorra (1996) developed a limited-information approach based on univariate and bivariate marginal distributions. Markov chain Monte Carlo has also been used (e.g. Zeger & Karim, 1991; Lee, 2007), typically with vague prior distributions. We refer to Skrondal & Rabe-Hesketh (2004, Chapter 6) for a comprehensive treatment of the estimation of latent variable models.

Acknowledgements

We wish to thank The Research Council of Norway for a grant supporting our collaboration.

References

Adcock, R. J. (1877a). Note on the method of least squares. *Analyst* **4**, 183–184.
Adcock, R. J. (1877b). A problem in least squares. *Analyst* **5**, 53–54.
Airy, G. B. (1861). *On the algebraical and numerical theory of errors of observations and the combination of observations*. MacMillan, Cambridge.
Aitkin, M., Anderson, D. & Hinde, J. (1981). Statistical modeling of data on teaching styles. *J. Roy. Statist. Soc. Ser. A* **144**, 419–461.
Allenby, G. M., Arora, N. & Ginter, J. L. (1998). On the heterogeneity of demand. *J. Mark. Res.* **35**, 384–389.
Allison, P. D. (1982). Discrete time methods for the analysis of event histories. In *Sociological methodology 1982* (ed. S. Leinhardt), 61–98. Jossey-Bass, San Francisco, CA.
Andersen, E. B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **32**, 283–301.
Andersen, E. B. (1972). The numerical solution of a set of conditional estimation equations. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **34**, 42–54.
Andersen, E. B. (1973). *Conditional inference and models for measuring*. Mentalhygiejnisk Forlag, Copenhagen.
Andersen, E. B. (1980a). Comparing latent populations. *Psychometrika* **45**, 121–134.
Andersen, E. B. (1980b). *Discrete statistical models with social science applications*. North-Holland, Amsterdam.
Andersen, E. B. (1982). Latent structure analysis: a survey. *Scand. J. Statist.* **9**, 1–12.
Andersen, E. B. & Madsen, M. (1977). Estimating the parameters of the latent population distribution. *Psychometrika* **42**, 357–374.
Anderson, T. W. & Rubin, H. (1956). Statistical inference in factor analysis. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* (ed. J. Neyman), 111–150. University of California Press, Berkeley, CA.
Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika* **43**, 561–573.
Ansari, A. & Jedidi, K. (2000). Bayesian factor analysis for multilevel binary observations. *Psychometrika* **65**, 475–496.

Arminger, G. & Küsters, U. (1988). Latent trait models with indicators of mixed measurement levels. In *Latent trait and latent class models* (eds R. Langeheine & J. Rost), 51–73. Plenum Press, New York.
Arminger, G. & Küsters, U. (1989). Construction principles for latent trait models. In *Sociological methodology 1989* (ed. C. C. Clogg), 369–393. Blackwell, Oxford.
Arminger, G. & Muthén, B. O. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis–Hastings algorithm. *Psychometrika* **63**, 271–300.
Arminger, G., Stein, P. & Wittenberg, J. (1999). Mixtures of conditional mean- and covariance-structure models. *Psychometrika* **64**, 475–494.
Bartholomew, D. J. (1980). Factor analysis for categorical data (with discussion). *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **42**, 293–321.
Bartholomew, D. J. (1987). *Latent variable models and factor analysis*. Oxford University Press, Oxford.
Bartholomew, D. J. & Knott, M. (1999). *Latent variable models and factor analysis*. Arnold, London.
Bartlett, M. S. (1953). Factor analysis in psychology as a statistician sees it. In *Uppsala Symposium on Psychological Factor Analysis*, 23–34. Almqvist and Wiksell, Stockholm.
Bentler, P. M. & Weeks, D. G. (1980). Linear structural equations with latent variables. *Psychometrika* **45**, 289–308.
Berkson, J. (1950). Are there two regressions? *J. Amer. Statist. Assoc.* **45**, 164–180.
Birnbaum, A. (1968). Test scores, sufficient statistics, and the information structures of tests. In *Statistical theories of mental test scores* (eds F. M. Lord & M. R. Novick), 425–435. Addison-Wesley, Reading, MA.
Blåfield, E. (1980). *Clustering of observation from finite mixtures with structural information*. PhD thesis, Jyväskyla University, Finland.
Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37**, 29–51.
Bock, R. D. & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika* **31**, 507–534.
Bock, R. D. & Lieberman, M. (1970). Fitting a response model for *n* dichotomously scored items. *Psychometrika* **33**, 179–197.
Böckenholt, U. (2001). Mixed-effects analyses of rank-ordered data. *Psychometrika* **66**, 45–62.
Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.* **88**, 9–25.
Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models, applications and data analysis methods*. Sage, Newbury Park, CA.
Carroll, R. J., Roeder, K. & Wasserman, L. (1999). Flexible parametric measurement error models. *Biometrics* **86**, 44–54.
Carroll, R. J., Ruppert, D., Stefanski, L. A. & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective,* 2nd edn. Chapman & Hall/CRC, Boca Raton, FL.
Christensen, W. F. & Amemiya, Y. (2003). Modeling and prediction for multivariate spatial factor analysis. *J. Statist. Plann. Inference* **115**, 543–564.
Christofferson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika* **40**, 5–32.
Clayton, D. G. (1988). The analysis of event history data: a review of progress and outstanding problems. *Statist. Med.* **7**, 819–841.
Clayton, D. G. (1992). Models for the analysis of cohort and case–control studies with inaccurately measured exposures. In *Statistical models for longitudinal studies on health* (eds J. H. Dwyer, M. Feinlieb, P. Lippert & H. Hoffmeister), 301–331. Oxford University Press, New York.
Cochran, W. G. (1968). Errors of measurement in statistics. *Technometrics* **4**, 637–666.
Costa, P. T. & McCrae, R. R. (1985). *The NEO personality inventory manual*. Psychological Assessment Resources, Odessa, FL.
Coull, B. A. & Agresti, A. (1999). The use of mixed logit models to reflect heterogeneity in capture–recapture studies. *Biometrics* **55**, 294–301.
Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. Wiley, London.
Davidian, M. & Gallant, A. R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quindine. *J. Pharmacokinetics Biopharmaceutics* **20**, 529–556.
Davidian, M. & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: an overview and update. *J. Agric. Biol. Environ. Statist.* **8**, 387–419.
Dayton, C. M. & MacReady, G. B. (1988). Concomitant variable latent class models. *J. Amer. Statist. Assoc.* **83**, 173–178.

De Boeck, P. & Wilson, M. (eds) (2004). *Explanatory item response models: a generalized linear and non-linear approach*. Springer, New York.

Diggle, P. J. & Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis (with discussion). *J. Roy. Statist. Soc. Ser. C* **43**, 49–93.

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

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

Dupacŏvá, J. & Wold, H. (1982). On some identification problems in ML modeling of systems with indirect observation. In *Systems under indirect observation – causality structure prediction – Part II* (eds K. G. Jöreskog & H. Wold), 293–315. North-Holland, Amsterdam.

Durbin, J. (1954). Errors in variables. *Int. Statist. Rev.* **22**, 2332.

Durbin, J. & Koopman, S. J. (2001). *Time series analysis by state space methods*. Oxford University Press, Oxford.

Eisenhart, C. (1947). The assumptions underling the analysis of variance. *Biometrics* **3**, 1–21.

Etezadi-Amoli, J. & McDonald, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika* **48**, 315–342.

Faucett, C. L. & Thomas, D. C. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statist. Med.* **15**, 1663–1685.

Fischer, G. H. (1973). Linear logistic test model as an instrument in educational research. *Acta Psychol.* **37**, 359–374.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinb.* **52**, 399–433.

Fokoué, E. & Titterington, D. M. (2003). Mixtures of factor analysers. Bayesian estimation and inference by stochastic simulation. *Mach. Learn.* **50**, 73–94.

Follmann, D. A. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika* **53**, 553–562.

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

Formann, A. K. (1992). Linear logistic latent class analysis for polytomous data. *J. Amer. Statist. Assoc.* **87**, 476–486.

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

Friedman, M. (1957). *A theory of the consumption function*. Princeton University Press, Princeton, NJ.

Gallant, A. R. & Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica* **55**, 363–390.

Gibson, W. A. (1959). Three multivariate models: factor analysis, latent structure analysis and latent profile analysis. *Psychometrika* **24**, 229–252.

Gilmour, A. R., Anderson, R. D. & Rae, A. L. (1985). The analysis of binomial data by a generalized linear mixed model. *Biometrika* **72**, 593–599.

Goldberger, A. S. (1971). Econometrics and psychometrics: a survey of communalities. *Psychometrika* **36**, 83–107.

Goldstein, H. (1987). Multilevel covariance component models. *Biometrika* **74**, 430–431.

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

Goldstein, H. & Browne, W. J. (2005). Multilevel factor analysis models for continuous and discrete data. In *Contemporary psychometrics: a festschrift for Roderick P. McDonald* (eds A. Maydeu-Olivares & J. J. McArdle), 453–475. Erlbaum, Mahwah, NJ.

Goodman, L. A. (1974a). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I – a modified latent structure approach. *Amer. J. Sociol.* **79**, 1179–1259.

Goodman, L. A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika* **61**, 215–231.

Green, B. F. (1952). Latent structure analysis and its relation to factor analysis. *J. Amer. Statist. Assoc.* **47**, 71–76.

Gulliksen, H. (1950). *Theory of mental tests*. Wiley, New York.

Haavelmo, T. (1944). The probability approach in econometrics. *Econometrica* **12** (Suppl.), 1–118.

Hagenaars, J. A. P. (1993). *Loglinear models with latent variables*. Sage, Newbury Park, CA.

Hagenaars, J. A. P. (2002). Directed loglinear modeling with latent variables: causal models for categorial data with nonsystematic and systematic measurements errors. In *Applied latent class* *analysis* (eds J. A. P. Hagenaars & A. L. McCutcheon), 234–286. Cambridge University Press, Cambridge.

Harville, D. A. (1976). Extensions of the Gauss–Markov theorem to include random effects. *Ann. Statist.* **2**, 384–395.

Harville, D. A. (1977). Maximum likelihood approaches to variance components estimation and related problems. *J. Amer. Statist. Assoc.* **72**, 320–340.

Hauser, R. M. & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. In *Sociological methodology 1971* (ed. H. L. Costner), 81–117. Jossey-Bass, San Francisco, CA.

Hausman, J. A. & Wise, D. A. (1979). Attrition bias in experimental and panel data: the Gary income maintenance experiment. *Econometrica* **47**, 455–473.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica* **46**, 931–959.

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

Heckman, J. J. & Willis, R. J. (1976). Estimation of a stochastic model of reproduction: an econometric approach. In *Household production and consumption* (ed. N. Terleckyj), 99–138. National Bureau of Economic Research, New York.

Hogan, J. W. & Laird, N. M. (1997). Model-based approaches to analyzing incomplete repeated measures and failure time data. *Statist. Med.* **16**, 259–271.

Hougaard, P. (2000). *Analysis of multivariate survival data*. Springer, New York.

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

Jones, R. H. (1993). *Longitudinal data with serial correlation: a state-space approach*. Chapman & Hall, London.

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika* **32**, 443–482.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika* **34**, 183–202.

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

Jöreskog, K. G. (1971b). Statistical analysis of sets of congeneric tests. *Psychometrika* **36**, 109–133.

Jöreskog, K. G. (1973a). Analysis of covariance structures. In *Multivariate analysis*, Vol. III (ed. P. R. Krishnaiah), 263–285. Academic Press, New York.

Jöreskog, K. G. (1973b). A general method for estimating a linear structural equation system. In *Structural equation models in the social sciences* (eds A. S. Goldberger & O. D. Duncan), 85–112. Seminar Press, New York.

Jöreskog, K. G. (1977). Structural equation models in the social sciences: specification, estimation and testing. In *Applications of statistics* (ed. P. R. Krishnaiah), 265–287. North-Holland, Amsterdam.

Kenny, D. A. & Judd, C. M. (1984). Estimating the non-linear and interactive effects of latent variables. *Psychol. Bull.* **96**, 201–210.

Knorr-Held, L. & Best, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *J. Roy. Statist. Soc. Ser. A* **164**, 73–85.

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

Lawley, D. N. (1939). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edinb.* **60**, 64–82.

Lawley, D. N. (1943). On problems connected with item selection and test construction. *Proc. Roy. Soc. Edinb.* **61**, 273–287.

Lawley, D. N. & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. Butterworths, London.

Lazarsfeld, P. F. (1950). The logical and mathematical foundation of latent structure analysis. In *Studies in social psychology in World War II, vol. IV, measurement and prediction* (eds S. A. Stouffer, L. Guttmann, E. A. Suchman, P. F. Lazarsfeld, S. A. Star & J. A. Clausen), 362–412. Princeton University Press, Princeton, NJ.

Lazarsfeld, P. F. (1954). A conceptual introduction to latent structure analysis. In *Mathematical thinking in the social sciences* (ed. P. F. Lazarsfeld), 349–387. Free Press, Glencoe, IL.

Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mill, Boston, MA.

Lee, S.-Y. (2007). *Structural equation modelling: a Bayesian approach*. Wiley, New York.

Lee, Y. & Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **58**, 619–678.

Lee, S.-Y. & Song, X.-Y. (2003). Bayesian analysis of structural equation models with dichotomous variables. *Statist. Med.* **22**, 3073–3088.

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

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

Lee, Y., Nelder, J. A. & Pawitan, Y. (2006). *Generalized linear models with random effects: unified analysis via H-likelihood*. Chapman & Hall/CRC, Boca Raton, FL.

de Leeuw, J. & Verhelst, N. D. (1986). Maximum likelihood estimation in generalized Rasch models. *J. Educ. Statist.* **11**, 183–196.

Lenk, P. J. & DeSarbo, W. S. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* **65**, 93–119.

Lindsay, B. G. (1995). *Mixture models: theory, geometry and applications. NSF-CBMS Regional Conference Series in Probability and Statistics*, Vol. 5. Institute of Mathematical Statistics, Hayward, CA.

Lindsay, B. G., Clogg, C. C. & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *J. Amer. Statist. Assoc.* **86**, 96–107.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.* **88**, 125–134.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated measures studies. *J. Amer. Statist. Assoc.* **90**, 1112–1121.

Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data*, 2nd edn. Wiley, New York.

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

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

Lord, F. M. (1952). *A theory of test scores.*. Psychometric Monograph 7. Psychometric Society, Bowling Green, OH.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum, Hillsdale, NJ.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley, Reading, MA.

Magder, S. M. & Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *J. Amer. Statist. Assoc.* **11**, 86–94.

Magidson, J. & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. In *Sociological methodology 2001* (eds M. Sobel & M. Becker), 223–264. Blackwell, Boston, MA.

Maier, K. S. (2001). A Rasch hierarchical measurement model. *J. Educ. Behav. Statist.* **26**, 307–330.

Mason, W. M., Wong, G. Y. & Entwisle, B. (1983). Contextual analysis through the multilevel linear model. In *Sociological methodology 1983–1984* (ed. S. Leinhart), 72–103. Jossey-Bass, San Francisco, CA.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika* **47**, 149–174.

McArdle, J. J. (1988). Dynamic but structural equation modeling with repeated measures data. In *Handbook of multivariate experimental psychology*, Vol. II (eds J. R. Nesselroade & R. B. Cattell), 561–614. Plenum Press, New York.

McArdle, J. J. & McDonald, R. P. (1984). Some algebraic properties of the reticular action model. *Br. J. Math. Statist. Psychol.* **37**, 234–251.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **42**, 109–142.

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

McCulloch, C. E., Lin, H., Slate, E. H. & Turnbull, B. W. (2002). Discovering subpopulation structure with latent class mixed models. *Statist. Med.* **21**, 417–429.

McDonald, R. P. (1967). *Nonlinear factor analysis. Psychometric Monograph 15*. Psychometric Society, Bowling Green, OH.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Erlbaum, Hillsdale, NJ.

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

Mehta, P. D. & Neale, M. C. (2005). People are variables too: multilevel structural equations modeling. *Psychol. Methods* **10**, 259–284.

Mellenbergh, G. J. (1994). Generalized linear item response theory. *Psychol. Bull.* **115**, 300–307.

Meng, X.-L. & Schilling, S. G. (1996). Fitting full-information item factor models and an empirical investigation of bridge sampling. *J. Amer. Statist. Assoc.* **91**, 1254–1267.
Meredith, W. & Tisak, J. (1990). Latent curve analysis. *Psychometrika* **55**, 107–122.
Mislevy, R. J. (1983). Item response models for grouped data. *J. Educ. Statist.* **8**, 271–288.
Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Appl. Psychol. Meas.* **11**, 81–91.
Mislevy, R. J. & Verhelst, N. (1990). Modeling item responses when different subjects employ different strategies. *Psychometrika* **55**, 195–215.
Mislevy, R. J. & Wilson, M. (1996). Marginal maximum likelihood estimation for a psychometric model of discontinuous development. *Psychometrika* **61**, 41–71.
Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *Br. J. Math. Statist. Psychol.* **49**, 313–334.
Müller, P. & Roeder, K. (1997). A Bayesian semiparametric model for case–control studies with errors in variables. *Biometrika* **84**, 523–537.
Muthén, B. O. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika* **43**, 551–560.
Muthén, B. O. (1979). A structural probit model with latent variables. *J. Amer. Statist. Assoc.* **74**, 807–811.
Muthén, B. O. (1983). Latent variable structural modelling with categorical data. *J. Econom.* **22**, 43–65.
Muthén, B. O. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. *Psychometrika* **49**, 115–132.
Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika* **54**, 557–585.
Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociol. Methods Res.* **22**, 376–398.
Muthén, B. O. (2002). Beyond SEM: general latent variable modeling. *Behaviormetrika* **29**, 81–117.
Muthén, B. O. & Satorra, A. (1996). Technical aspects of Muthén's LISCOMP approach to estimation of latent variable relations with a comprehensive measurement model. *Psychometrika* **60**, 489–503.
Muthén, B. O., Brown, C. H., Masyn, K., Jo, B., Khoo, S. T., Yang, C. C., Wang, C. P., Kellam, S. G., Carlin, J. B. & Liao, J. (2002). General growth mixture modeling for randomized preventive interventions. *Biostatistics* **3**, 459–475.
Nagin, D. S. & Land, K. C. (1993). Age, criminal careers, and population heterogeneity: specification and estimation of a nonparametric mixed Poisson model. *Criminology* **31**, 327–362.
Neale, M. C. & Cardon, L. R. (1992). *Methodology for genetic studies of twins and families*. Kluwer, London.
Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica* **16**, 1–32.
Pearson, K. (1894). Contributions to the mathematical theory of evolution. *Philos. Trans. Roy. Soc. A* **185**, 71–110.
Peirce, C. S. (1884). The numerical measure of success of predictions. *Science* **4**, 453–454.
Pinheiro, J. C., Liu, C. & Wu, Y. (2001). Efficient algorithms for robust estimation in linear mixed-effects models using the multivariate *t*-distribution. *J. Comput. Graph. Statist.* **10**, 249–276.
Poon, W.-Y. & Lee, S.-Y. (1992). Maximum likelihood and generalized least squares analyses of two-level structural equation models. *Statist. Probab. Lett.* **14**, 25–30.
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.
Quandt, R. E. (1972). A new approach to estimating switching regressions. *J. Amer. Statist. Assoc.* **67**, 306–310.
Rabe-Hesketh, S. & Skrondal, A. (2001). Parameterization of multivariate random effects models for categorical data. *Biometrics* **57**, 1256–1264.
Rabe-Hesketh, S. & Skrondal, A. (2007). Multilevel and latent variable modeling with composite links and exploded likelihoods. *Psychometrika* **72**, 123–140.
Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika* **69**, 167–190.
Rabe-Hesketh, S., Skrondal, A. & Pickles, A. (2005). Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J. Econom.* **128**, 301–323.
Rabe-Hesketh, S., Skrondal, A. & Zheng, X. (2007). Multilevel structural equation modeling. In *Handbook of latent variable and related models* (ed. S. Y. Lee), 209–227. Elsevier, Amsterdam.

Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika* **20**, 93–111.
Rao, C. R. (1958). Some statistical methods for comparison of growth curves. *Biometrics* **14**, 1–17.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danmarks Pædagogiske Institut, Copenhagen.
Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (ed. J. Neymann), 321–333. University of California Press, Berkeley, CA.
Raudenbush, S. W. & Sampson, R. (1999). Ecometrics: toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. In *Sociological methodology 1999* (ed. P. V. Marsden), 1–41. Blackwell, Oxford.
Reiersøl, O. (1945). Confluence analysis by means of instrumental sets of variables. *Arkiv for Mathematik, Astronomi och Fysik* **32A**, 1–119.
Richardson, S., Leblond, L., Jaussent, I. & Green, P. (2002). Mixture models in measurement error problems, with reference to epidemiological studies. *J. Roy. Statist. Soc. Ser. A* **165**, 549–556.
Rijmen, F. & De Boeck, P. (2002). The random weights linear logistic test model. *Appl. Psychol. Meas.* **26**, 271–285.
Rindskopf, D. & Rindskopf, W. (1986). The value of latent class analysis in medical diagnosis. *Statist. Med.* **5**, 21–27.
Rost, J. (1990). Rasch models in latent classes: an integration of two approaches to item analysis. *Appl. Psychol. Meas.* **14**, 271–282.
Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
Rue, H. & Held, L. (2005). *Gaussian Markov random fields*. Chapman & Hall/CRC, Boca Raton, FL.
Samejima, F. (1969). *Estimation of latent trait ability using a response pattern of graded scores.*. Psychometric Monograph 17. Psychometric Society, Bowling Green, OH.
Sammel, M. D., Ryan, L. M. & Legler, J. M. (1997). Latent variable models for mixed discrete and continuous outcomes. *J. Roy. Statist. Soc. Ser. B Statist. Methodol.* **59**, 667–678.
Skrondal, A. & Rabe-Hesketh, S. (2003). Multilevel logistic regression for polytomous data and rankings. *Psychometrika* **68**, 267–287.
Skrondal, A. & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: multilevel, longitudinal, and structural equation models*. Chapman & Hall/CRC, Boca Raton, FL.
Spearman, C. (1904). General intelligence, objectively determined and measured. *Amer. J. Psychol.* **15**, 201–293.
Steele, F. & Goldstein, H. (2006). A multilevel factor model for mixed binary and ordinal indicators of women's status. *Sociol. Methods Res.* **35**, 137–153.
Sutton, A. J., Abrams, K. R., Jones, D. R., Sheldon, T. A. & Song, F. (2000). *Methods for meta-analysis in medical research*. Wiley, New York.
Swamy, P. A. V. B. (1970). Efficient inference in a random coefficient regression model. *Econometrica* **38**, 311–323.
Swamy, P. A. V. B. (1971). *Statistical inference in random coefficient regression models*. Springer, New York.
Takane, Y. & de Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika* **52**, 393–408.
Ten Have, T. R., Miller, M. E., Reboussin, B. A. & James, M. K. (2000). Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple cause drop-out from the longitudinal study of aging. *Biometrics* **56**, 279–287.
Thompson, R. & Baker, R. J. (1981). Composite link functions in generalized linear models. *J. Roy. Statist. Soc. Ser. C* **30**, 125–131.
Thomson, G. H. (1938). *The factorial analysis of human ability*. University of London Press, London.
Thurstone, L. L. (1931). Multiple factor analysis. *Psychol. Rev.* **38**, 406–427.
Thurstone, L. L. (1935). *The vectors of mind*. University of Chicago Press, Chicago, IL.
Thurstone, L. L. (1947). *Multiple-factor analysis: a development and expansion of the vectors of mind.* University of Chicago Press, Chicago, IL.
Tjur, T. (1982). A connection between Rasch's item analysis model and a multiplicative Poisson model. *Scand. J. Statist.* **9**, 23–30.
Train, K. E. (2003). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge.
Uebersax, J. S. (1993). Statistical modeling of expert ratings on medical treatment appropriateness. *J. Amer. Statist. Assoc.* **88**, 421–427.

Uebersax, J. S. & Grove, W. M. (1993). A latent trait finite mixture model for the analysis of rating agreement. *Biometrics* **49**, 823–835.

Verbeke, G. & Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *J. Amer. Statist. Assoc.* **91**, 217–221.

Verbeke, G. & Molenberghs, G. (2000). *Linear mixed models for longitudinal data*. Springer, New York.

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

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

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statist. Methods Med. Res.* **16**, in press.

Wall, M. M. & Amemiya, Y. (2000). Estimation for polynomial structural equation models. *J. Amer. Statist. Assoc.* **95**, 929–940.

Wall, M. M. & Amemiya, Y. (2007). Nonlinear structural equation modeling as a statistical method. In *Handbook of latent variable and related models* (ed. S. Y. Lee), 321–343. Elsevier, Amsterdam.

Wedel, M. & Kamakura, W. A. (2000). *Market segmentation: conceptual and methodological foundations*, 2nd edn. Kluwer, Dordrecht.

Wedel, M. & Kamakura, W. A. (2001). Factor analysis with (mixed) observed and latent variables. *Psychometrika* **66**, 515–530.

Wright, S. (1918). On the nature of size factors. *Genetics* **3**, 367–374.

Wu, M. C. & Carroll, R. J. (1988). Estimation and comparison of change in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

Xu, J. & Zeger, S. L. (2001). Joint analysis of longitudinal data comprising repeated measures and times to events. *J. Roy. Statist. Soc. Ser. C* **50**, 375–387.

Yalcin, I. & Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statist. Sci.* **16**, 275–294.

Yung, Y.-F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika* **62**, 297–330.

Zeger, S. L. & Karim, M. R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *J. Amer. Statist. Assoc.* **86**, 79–86.

Zellner, A. (1970). Estimation of regression relationships containing unobservable variables. *Int. Econom. Rev.* **11**, 441–454.

Zwinderman, A. H. (1991). A generalized Rasch model for manifest predictors. *Psychometrika* **56**, 589–600.

Received December 2006, in final form August 2007

Anders Skrondal, Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK.
E-mail: a.skrondal@lse.ac.uk