Analysis of Multivariate Longitudinal Categorical Data Subject to Nonrandom Missingness: A Latent Variable Approach

Mai Sherif Hafez

A thesis submitted to the Department of Statistics of the London School of Economics for the degree of Doctor of Philosophy

London, September 2015

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of 33,615 words.

Statement of conjoint work:

I confirm that parts of Chapters 4 and 5 were jointly co-authored with my Supervisors, Professor Irini Moustaki and Dr. Jouni Kuha, and published in *Structural Equation Modeling: A Multidisciplinary Journal.*

Acknowledgements

First and foremost, I would like to express my sincere gratitude and appreciation to my supervisor, Professor Irini Moustaki for her continuous support, guidance and understanding and for making my Ph.D. a life-changing experience.

I am also thankful to Dr. Jouni Kuha, my advisor, for his valuable feedback and constructive recommendations during my studies.

My time at LSE was made enjoyable, thanks to the friendly and helpful staff and colleagues at the Department of Statistics of LSE. I particularly thank Ian Marshall for his genuine suppport.

I am both thankful and grateful to my loving parents, Mona and Sherif, and my supportive brother Kareem for their never-ending support and for always believing in me.

I am greatly indebted to my husband, Ismail, who made this experience possible by his undeniable sacrifice and support even at the hardest of times, and to baby Amina who made the completion of this thesis a much bigger challenge.

Last but not least, I gratefully acknowledge the Citadel Scholarship Foundation of Egypt for giving me this opportunity by their financial support.

Abstract

Longitudinal data are collected for studying changes across time. In social sciences, interest is often in theoretical constructs, such as attitudes, behaviour or abilities, which cannot be directly measured. In that case, multiple related manifest (observed) variables, for example survey questions or items in an ability test, are used as indicators for the constructs, which are themselves treated as latent (unobserved) variables. In this thesis, multivariate longitudinal data is considered where multiple observed variables, measured at each time point, are used as indicators for theoretical constructs (latent variables) of interest. The observed items and the latent variables are linked together via statistical latent variable models.

A common problem in longitudinal studies is missing data, where missingness can be classified into one of two forms. Dropout occurs when subjects exit the study prematurely, while intermittent missingness takes place when subjects miss one or more occasions but show up on a subsequent wave of the study. Ignoring the missingness mechanism can lead to biased estimates, especially when the missingness is nonrandom.

The approach proposed in this thesis uses latent variable models to capture the evolution of a latent phenomenon over time, while incorporating a missingness mechanism to account for possibly nonrandom forms of missingness. Two model specifications are presented, the first of which incorporates dropout only in the missingness mechanism, while the other accounts for both dropout and intermittent missingness allowing them to be informative by being modelled as functions of the latent variables and possibly observed covariates.

Models developed in this thesis consider ordinal and binary observed items, because such variables are often met in social surveys, while the underlying latent variables are assumed to be continuous.

The proposed models are illustrated by analysing people's perceptions on women's work using three questions from five waves of the British Household Panel Survey.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we study latent variable modelling of multivariate longitudinal data subject to different types of missingness. Dropout and intermittent missingness are two types of missing data that we incorporate within a latent variable modelling framework to account for missingness while capturing the evolution of the latent phenomenon of interest.

In social sciences, such as educational testing and psychometrics, interest is often in theoretical constructs, such as attitudes, behaviour or abilities, which cannot be directly measured. In that case, multiple related manifest (observed) variables, for example survey questions or items in an ability test, are used as indicators for the constructs, which are themselves treated as latent (unobserved) variables. The observed items and the latent variables are linked together by statistical latent variable models (see e.g. Skrondal and Rabe-Hesketh (2004) and Bartholomew et al. (2011) for overviews). Both manifest and latent variables can be either categorical or continuous resulting in different versions of latent variable models. In our research, we consider models with ordinal or binary observed items, because such variables are often met in social surveys, while we assume latent variables to be continuous.

Longitudinal data are collected for studying changes across time. Most of the existing research on longitudinal data focuses on repeated measures for one variable over time. Good starting points to the extensive literature on such univariate longitudinal data analysis are Diggle et al. (2013), who give a thorough overview of different methods, and Verbeke and Molenberghs (2000), who provide a comprehensive treatment of linear mixed models for continuous longitudinal data. However, when interest lies in how latent constructs change across time, the same multiple items are measured repeatedly at different time points, thus resulting in multivariate longitudinal data. Models for such data have been proposed by, for example, Fieuws and Verbeke (2004, 2006), Dunson (2003), and Cagnone et al. (2009), who model the associations of the latent and observed variables across time using random effects and/or latent variables.

Missing data is an unavoidable problem in almost every dataset, especially with longitudinal data. The most common type of missingness in longitudinal studies is dropout, where subjects exit the study prematurely. A crucial question for the analysis is whether or not those who drop out are systematically different from the ones who remain till the end of the study. Intermittent missingness where an individual misses an occasion and shows up on a subsequent wave, is also possible. In our research, these two types of missingness are incorporated within a latent variable model framework for multivariate longitudinal data.

The thesis is outlined as follows: Chapter 2 provides a literature review of latent variable models outlining the two main approaches for modelling categorical manifest variables; the underlying variable approach (UVA) and item response theory (IRT) approach. Chapter 3 gives a review on missing data in general with a focus on existing methods for modelling univariate longitudinal data subject to dropout. A latent variable model for analysing multivariate ordinal longitudinal data subject to dropout is developed in Chapter 4 under the underlying variable approach, with two possible model specifications. Chapter 5 provides an illustration for the developed models by applying them to a real dataset about people's attitudes towards women's work from five waves of the British Household Panel Survey (BHPS), along with a sensitivity analysis for different levels of dropout. In Chapter 6, one of the two model specifications presented in Chapter 4 is used to develop a similar model for multivariate binary longitudinal data under IRT. Chapter 7 extends the model developed for binary observed items within an IRT framework to accommodate intermittent missingness together with dropout. Two possible specifications are given for this model too. An application of this model is also presented using the British Household Panel Survey (BHPS) data. Chapter 8 gives a final conclusion with highlights on the contribution of the research. Future areas for research are introduced including the incorporation of item non-response within the same model framework.

1.1 Notation

Observed variables will be denoted by y, where $\mathbf{y}$ will denote a $(p \times 1)$ vector of p observed variables. Observed variables will be either ordinal or binary. Latent variables on the other hand will be denoted by z, where $\mathbf{z}$ will denote a $(q \times 1)$ vector of q latent variables. Latent variables are assumed to be normally distributed throughout. A vector of covariates will be denoted by $\mathbf{x}$.

The subscript i will be used to identify an observed variable y , while j will be used to identify a latent variable variable z. Subscript for an individual of the sample will be denoted by m.

Further notation will be introduced in relevant parts of the thesis.

Chapter 2

Literature Review on Latent Variable Models

In order to develop a latent variable model for multivariate longitudinal data subject to different forms of missingness, we first need to review the existing literature on several topics. In this chapter, a literature review of latent variable models for multivariate complete data is given, first in a cross-sectional context followed by the longitudinal case.

Latent variable modelling is an important tool in multivariate data analysis. One of the main reasons behind using such a technique is trying to measure constructs or concepts that cannot be directly measured, which are often met in social sciences (e.g. democracy, satisfaction, attitude,...). These are referred to as latent (unobserved) factors or variables, and can be measured via a number of manifest (observed) variables or items. The dimension reduction caused by summarising a set of related observed variables into one or few latent variables that can be used in further data analysis without losing much of the structure in the data is itself another main reason behind using latent variable models.

A latent variable model consists of two parts: a measurement part that links the observed variables to the latent variables; and a structural part that specifies relationships among latent variables and possibly covariates. It is assumed that associations among observed variables are explained by the latent variables. This is an assumption of conditional independence where the observed items are independent given the latent variables.

Both manifest and latent variables can be either metric or categorical. Metric variables can be either discrete or continuous while categorical variables can be ordered (ordinal) or unordered (nominal). When both manifest and latent variables are metric, factor analysis is implemented. On the other hand, latent class analysis is applied when both manifest and latent variables are categorical. When manifest variables are categorical while latent variables are metric, latent trait analysis is the appropriate technique to adopt. Oppositely, when manifest variables are metric and latent variables are categorical the suitable latent variable method is latent profile analysis. Bartholomew et al. (2011) present a unified approach for latent variable models for which each of the before-mentioned techniques can be viewed as a special case within the same general framework.

It is also possible for both manifest and latent variables to be of mixed types. When latent variables are of mixed type, hybrid models are used. These will not be discussed within the scope of this thesis. Moustaki and Knott (2000a) propose a generalised linear model framework which allows simultaneous analysis for different types of manifest variables from the exponential family including metric, binary and nominal items. Moustaki (1996) develops a method for analysing latent variable models with metric and binary manifest variables within the same approach. Moustaki and Steele (2005) discuss a latent variable model with a mixture of categorical and survival items.

When the observed variables are categorical, there are two approaches for estimating parameters of the latent variable model. The underlying variable approach developed within the structural equation modelling (SEM) framework regards categorical variables as manifestations of underlying unobserved continuous variables and thus the problem is converted into one with metric observed variables where factor analysis can be employed. The second approach is the response function approach, also known as item response theory (IRT) where distributional assumptions are directly made on categorical manifest variables. A function is defined to give the probability of obtaining a response in each category of the categorical variable given the respondent's position on the latent variable scale.

In this thesis, we develop two types of models. The first is for ordinal observed variables where the underlying variable approach is adopted. The second is for binary observed items in which the response function approach is employed. Latent variables are assumed to be continuous in both cases. We therefore present the underlying variable approach for ordinal variables, followed by the item response theory for binary items. Bartholomew et al. (2011) (pp. 79-81) prove the equivalence of the two approaches for binary data.

2.1 The Underlying Variable Approach

Structural equation models (SEM) can be viewed as an extension to factor analysis. Whereas factor analysis only focuses on relationships between observed and latent variables, structural equation models allow estimation and testing of relationships between observed and latent variables on one hand (measurement model), and relationships among latent variables on the other (structural model).

Structural equation modelling is developed to handle continuous observed variables. When observed variables are categorical, the underlying variable approach (UVA) is adopted. The underlying variable approach regards categorical variables as manifestations of underlying unobserved continuous variables and thus the problem is converted into one with metric observed variables. Early contributions to the development of this method can be found in Jöreskog (1990, 1994), Muthén (1984) and Arminger and Küsters (1988) among others. The underlying variable approach is supported by software such as LISREL (Jöreskog and Sörbom (1996)) and Mplus (Muthén and Muthén (2011)).

Jöreskog (1990) defines an ordinal variable as one that takes values out of a set of ordered categories, such as a five-category Likert scale. The categories are ordered ascendingly or descendingly but the distances between categories are neither specified nor equal (example: strongly agree, agree, don't know, disagree and strongly disagree). Even when the categories of an ordinal variable are assigned numeric values, these values should not be treated as values of a continuous variable. Means, variances and covariances should not be calculated for an ordinal variable, but rather counts of responses in each category. That is why different techniques are applied when ordinal variables are used within structural equation models.

The underlying variable approach assumes that each ordinal variable y is a manifestation of an underlying unobserved continuous variable y^* which is used in fitting the structural equation model. For an ordinal variable y with c categories, the relationship between these two variables is given in Jöreskog (2005) by

$$y = s \quad \Longleftrightarrow \quad \tau_{s-1} < y^* < \tau_s, \qquad s = 1, 2, \ldots, c, \tag{2.1}$$

where

$$-\infty = \tau_0 < \tau_1 < \tau_2 < \ldots < \tau_{c-1} < \tau_c = +\infty,$$

are parameters known as thresholds. There are $c-1$ estimable thresholds for an ordinal variable y with c categories. It is the continuous unobserved variable y^* that is used in structural equation models not the ordinal observed variable y. Since only ordinal information is available about the underlying continuous variable y^*, its mean and variance are not identified and it is therefore assumed to have a standard normal distribution with a density function $\phi(u)$ and a distribution function $\Phi(u)$. The choice of a standard normal distribution is explained in Jöreskog (2005) by the fact that any continuous distribution can be transformed by a monotonic transformation into a standard normal. The probability of y falling in category s can therefore be expressed by

$$\pi_s = \Pr[y = s] = \Pr[\tau_{s-1} < y^* < \tau_s] = \int_{\tau_{s-1}}^{\tau_s} \phi(u)du = \Phi(\tau_s) - \Phi(\tau_{s-1}),$$

and thus the threshold parameters are

$$\tau_s = \Phi^{-1}(\pi_1 + \pi_2 + \ldots + \pi_s), \qquad s = 1, \ldots, c-1,$$

where Φ^{-1} is the inverse of the standard normal distribution function. The quantity

$(\pi_1 + \pi_2 + \ldots + \pi_s)$ is the probability that a response falls in category s or lower.

2.1.1 Measurement Model

The measurement model is the classical linear factor model

$$y_i^* = \alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j + \varepsilon_i, \qquad i = 1, \ldots, p, \tag{2.2}$$

where α_i is the mean of the i^{th} item (here zero as the underlying variables are assumed to have a standard normal distribution), λ_{ij} is the loading of the latent variable z_j on the underlying continuous variable y_i^* and ε_i is a normally distributed random error; $\varepsilon_i \sim N(0, \omega_{ii}^2)$ that is uncorrelated with errors of other items. Or in matrix form

$$\mathbf{y}^* = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{z} + \boldsymbol{\epsilon},$$

where $\boldsymbol{\alpha}$ is a $(p \times 1)$ vector of zero means, $\boldsymbol{\Lambda}$ is a $(p \times q)$ matrix of loadings, and $\boldsymbol{\epsilon}$ is a $(p \times 1)$ vector of normally distributed random errors $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \boldsymbol{\Omega})$; such that $\boldsymbol{\Omega}$ is a $(p \times p)$ diagonal matrix of error variances.

Model (2.2) can also be referred to as a cumulative probit model or an ordered probit model (McElvey and Zavoina (1975)), an extension to the well-known probit model where the dependent variable is ordinal instead of binary. Alternatively, an ordered logit model, the counterpart of a logit model for modelling ordinal data, can be obtained by assuming a logistic distribution for the error term ε_i (McCullagh (1980)).

The linear model introduced above for modelling the underlying continuous variable y_i^* is an alternative representation of a generalised linear model for the ordinal variable y_i. This presentation will be introduced in Section 2.2 under item response theory, with binary items as a special case.

The latent factors are assumed to account for dependencies among the observed variables (in this case the underlying variables); such that conditional on the latent factors, the observed variables are independent. If both the underlying continuous variables and the latent variables are assumed to have standard normal distributions, the conditional distribution of $\mathbf{y}^*$ given $\mathbf{z}$ is

$$\mathbf{y}^* \mid \mathbf{z} \sim N_p(\mathbf{\Lambda z}, \Omega),$$

and the marginal distribution of $\mathbf{y}^*$ is thus a multivariate normal

$$\mathbf{y}^* \sim N_p(\mathbf{0}, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = \Lambda\Lambda' + \Omega$ is the theoretical covariance matrix of the underlying variables.

2.1.2 Structural Model

The structural part of the model which defines relationships among latent variables, possibly in addition to a set of observed covariates $\mathbf{x}$ is given by

$$z_j = \sum_{l=1}^{q} \phi_{jl} z_l + \sum_{h=1}^{r} \beta_{jh} x_h + \delta_j, \qquad j = 1, \ldots, q,$$

where ϕ_{jl} is a regression coefficient representing the dependence of a factor z_j on another latent factor z_l, β_{jh} is a regression coefficient representing the dependence of a factor z_j on an observed covariate x_h and δ_j is a normally distributed random error; $\delta_j \sim N(0, \upsilon_{jj}^2)$ that is uncorrelated with the latent factors z_l. Or in matrix notation

$$\mathbf{z} = \mathbf{\Phi z} + \boldsymbol{\beta}\mathbf{x} + \boldsymbol{\delta},$$

where $\mathbf{\Phi}$ is a $(q \times q)$ coefficient matrix representing relationships among latent variables, $\boldsymbol{\beta}$ is a $(q \times r)$ matrix of coefficients representing dependence of latent variables on covariates, and $\boldsymbol{\delta}$ is a $(q \times 1)$ vector of normally distributed random errors $\boldsymbol{\delta} \sim N_q(\mathbf{0}, \mathbf{\Upsilon})$; such that $\mathbf{\Upsilon}$ is a $(q \times q)$ covariance matrix of error terms that is possibly diagonal if the errors are not allowed to correlate.

2.1.3 Estimation

Estimation methods for the classical linear factor model, such as maximum likelihood (ML) and generalised least squares, provide parameter estimates that in some sense minimise the distance between the observed $\mathbf{S}$ and theoretical $\mathbf{\Sigma}$ covariance matrices of the items. However, an observed covariance matrix cannot be obtained for categorical variables. Therefore, the estimation procedure should start in this case by obtaining a covariance/correlation matrix that can be employed in the estimation process.

The following procedure described in Jöreskog (1990, 1994) is a three stage estimation method; also known as PRELIS/LISREL Approach (PLA) for parameters estimation of a structural equation model. A similar approach is also given in Muthén (1984). In the first step, thresholds are estimated from the univariate marginal distributions of the underlying variables. In the second step, polychoric correlations are estimated from the bivariate marginal distributions for given thresholds, thus a matrix of polychoric correlations which can be used in the estimation of the model parameters is obtained. The third step involves estimation of the measurement/structural model parameters.

STEP 1: The probabilities π_s are unknown population parameters and can be estimated by their corresponding sample quantities p_s, which represent the percentages of responses in category s. Therefore, the estimates of the thresholds become

$$\hat{\tau}_s = \Phi^{-1}(p_1 + p_2 + \ldots + p_s), \qquad s = 1, \ldots, c-1,$$

where $\hat{\tau}_s$ are the maximum likelihood estimators of τ_s based on the univariate marginal data.

STEP 2: Considering the bivariate distribution, suppose there are two ordinal variables y_1 and y_2 with c_1 and c_2 categories, respectively. The bivariate marginal distribution can be represented by a $c_1 \times c_2$ contingency table that cross tabulates the two variables, such that the $(s_1, s_2)^{th}$ cell contains the counts $n_{s_1 s_2}$ of cases in category s_1 for the first variable y_1 and in category s_2 for the second variable y_2. Since the underlying continuous variables y_1^* and y_2^* are both standard normal, their bivariate distribution is assumed to be standard bivariate normal with a correlation ρ_{12} (known as *polychoric correlation*). However this is an assumption to be tested as the normality of y_1^* and y_2^* does not guarantee their joint bivariate normality.

Let $\tau_1^{(1)}, \tau_2^{(1)}, \ldots, \tau_{c_1-1}^{(1)}$ be thresholds for the underlying variable y_1^* and $\tau_1^{(2)}, \tau_2^{(2)}, \ldots, \tau_{c_2-1}^{(2)}$ be the corresponding thresholds for y_2^*. Jöreskog (1994, 2005) outlines the estimation of the polychoric correlation for y_1^* and y_2^* by maximising the loglikelihood of the multinomial distribution,

$$\ln L = \sum_{s_1=1}^{c_1} \sum_{s_2=1}^{c_2} n_{s_1 s_2} \log \pi_{s_1 s_2}(\boldsymbol{\theta}),$$

where

$$\pi_{s_1 s_2}(\boldsymbol{\theta}) = \Pr\left[y_1 = s_1, y_2 = s_2\right] = \int_{\tau_{s_1-1}^{(1)}}^{\tau_{s_1}^{(1)}} \int_{\tau_{s_2-1}^{(2)}}^{\tau_{s_2}^{(2)}} \phi_2(u,v) du dv,$$

such that

$$\phi_2(u,v) = \frac{1}{2\pi\sqrt{(1-\rho_{12}^2)}} e^{-\frac{1}{2(1-\rho_{12}^2)}(u^2 - 2\rho_{12}uv + v^2)}$$

is the standard bivariate normal density with correlation ρ_{12}. There are $c_1 \times c_2$ probabilities $\pi_{s_1 s_2}(\boldsymbol{\theta})$ that are functions of the parameter vector

$$\boldsymbol{\theta} = (\tau_1^{(1)}, \tau_2^{(1)}, \ldots, \tau_{c_1-1}^{(1)}, \tau_1^{(2)}, \tau_2^{(2)}, \ldots, \tau_{c_2-1}^{(2)}, \rho_{12}).$$

Maximising $ln\, L$ is equivalent to minimising the bivariate fit function

$$F(\boldsymbol{\theta}) = \sum_{s_1=1}^{c_1} \sum_{s_2=1}^{c_2} p_{s_1 s_2} [ln\, p_{s_1 s_2} - ln\, \pi_{s_1 s_2}(\boldsymbol{\theta})] = \sum_{s_1 s_2} p_{s_1 s_2} ln\, [p_{s_1 s_2}/\pi_{s_1 s_2}(\boldsymbol{\theta})],$$

where $p_{s_1 s_2} = n_{s_1 s_2}/N$ are the sample proportions.

A full information maximum likelihood estimation approach assumes a multivariate normal distribution for all the underlying variables $y_1^*, y_2^*, \ldots, y_p^*$. Estimation involves minimising the $p-$dimensional fit function over all response patterns present in the data. This requires the evaluation of a $p-$dimensional integral for each response pattern in the sample, which is done numerically. This approach becomes computationally infeasible as the number of observed variables increases ($p > 4$). That is why the approach based on bivariate normality assumption

-outlined above- is usually adopted.

STEP 3: Whereas Muthén (1984) uses a generalised least squares method for estimating parameters of the structural part of the model in the third step, Jöreskog (1990, 1994) uses a weighted least squares method where the weight matrix is an estimate of the inverse of the asymptotic covariance matrix of the polychoric correlations to estimate those parameters.

2.2 Item Response Theory Approach

The second approach for estimating parameters of a latent variable model with categorical manifest variables is the response function/item response theory (IRT) approach, where distributional assumptions are directly made on categorical manifest variables. A function is defined to give the probability of obtaining a response in each category of the categorical variable given the respondent's position on the latent variable scale. Within the response function framework, Moustaki (1996) develops a method for analysing latent variable models with metric and binary manifest variables. Moustaki and Knott (2000a) propose a generalised linear model framework which allows simultaneous analysis for different types of manifest variables from the exponential family including metric, binary and nominal items. They define a generalised linear model as a model of three components:

1. The random component: each manifest variable y_i has a distribution from the exponential family with a canonical link function η_i taking the form

$$f_i(y_i, \eta_i, \varphi_i) = \exp\left\{\frac{y_i\eta_i - b_i(\eta_i)}{\varphi_i} + d_i(y_i, \varphi_i)\right\}, \qquad i = 1, ..., p, \tag{2.3}$$

where $b_i(\eta_i)$ and $d_i(y_i, \varphi_i)$ take different forms depending on the distribution of the manifest variable y_i, and φ_i is a scale parameter.

2. The systematic component: latent variables $z_1, ..., z_q$ produce a linear predictor η_i corresponding to each manifest variable y_i as follows

$$\eta_i = \alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j, \qquad i = 1, ..., p.$$

3. The link function: it provides the link between the systematic component η_i and the conditional mean of the random component $E(y_i \mid \mathbf{z})$ such that

$$\eta_i = \nu_i(\mu_i(\mathbf{z})) = \nu_i(E(y_i \mid \mathbf{z})),$$

where the link function ν_i can take diferent forms for different manifest variables.

Binary variables are very common in social sciences. Even when responses fall in more than two categories, in many cases these are collapsed into just two whether the original categorical variable is ordinal or nominal. In this section, a model is outlined for binary manifest variables where latent variables are assumed to be continuous. For latent trait models with polytomous data, see for example; Bartholomew et al. (2011) and Moustaki and Knott (2000a).

2.2.1 A Measurement Model for Binary Manifest Variables

Let $\mathbf{y} = (y_1, ..., y_p)'$ denote a vector of p binary manifest variables and $\mathbf{z} = (z_1, ..., z_q)'$ a vector of q latent variables, where q is practically much smaller than

p. Possible responses to each binary variable y_i, $i = 1, \dots, p$ are coded as 0 or 1. A sensible assumption would be that the manifest binary variable y_i has a Bernoulli distribution with expected value $\pi_i(\mathbf{z}) = \Pr(y_i = 1 \mid \mathbf{z})$, which is a member of the exponential family thus taking the form of equation (2.3) with $d_i(y_i, \varphi_i) = 0$ and $\varphi_i = 1$. An appropriate *link function* in the case of binary items is the logit function

$$\eta_i = \operatorname{logit} \pi_i(\mathbf{z}) = \alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j, \quad i = 1, \dots, p, \tag{2.4}$$

where α_i is a constant term , and λ_{ij} is the loading of the latent variable z_j on the i^{th} binary item y_i. The intercept α_i is known in educational testing, where this model originates, as the *difficulty parameter* because increasing its value increases the probability of a positive response $\pi_i(\mathbf{z}) = \Pr(y_i = 1 \mid \mathbf{z})$ for all respondents with different levels on the latent scale. The loading λ_{ij} is known as the *discrimination parameter* because the larger its value, the easier it becomes to discriminate between two respondents at a given distance on the latent scale. This can be viewed as a logistic latent trait model with response function

$$\pi_i(\mathbf{z}) = \frac{e^{\alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j}}{1 + e^{\alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j}}.$$

An alternative model for binary responses uses the inverse of the normal distribution function

$$\Phi^{-1} \pi_i(\mathbf{z}) = \alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j, \quad i = 1, \dots, p,$$

as a link function instead of the logit given in (2.4). This leads to a model equivalent to the underlying variable approach outlined in Section 2.1. This equivalence extends to the case of ordinal items y_i , but does not hold for unordered polytomous variables due to the fact that the categories are necessarily ordered in an underlying variable approach (Bartholomew et al. (2011)).

2.2.2 Structural Model

Latent variables are still assumed to be continuous. The structural part of the model is the same as that in a structural equation model outlined in Section 2.1.2.

2.2.3 Estimation

For a binary item y_i, the conditional distribution of y_i given $\mathbf{z}$ is taken to be the Bernoulli distribution,

$$g_i(y_i \mid \mathbf{z}) = \{\pi_i(\mathbf{z})\}^{y_i}\{1-\pi_i(\mathbf{z})\}^{1-y_i}, \quad y_i = 0, 1;\ i = 1, \ldots, p, \tag{2.5}$$

$$= \{1-\pi_i(\mathbf{z})\} \exp\{y_i(\alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j)\}.$$

Since only $\mathbf{y}$ can be observed, Bartholomew et al. (2011) define the joint distribution density function of $\mathbf{y}$ by

$$f(\mathbf{y}) = \int_{R_z} g(\mathbf{y} \mid \mathbf{z}) h(\mathbf{z}) d\mathbf{z},$$

where $h(\mathbf{z})$ is the prior distribution of $\mathbf{z}$, and $g(\mathbf{y} \mid \mathbf{z})$ is the conditional distribution of $\mathbf{y}$ given $\mathbf{z}$. Our assumption is that of conditional independence, meaning that if the set of latent variables $\mathbf{z}$ is *complete*, then $\mathbf{z}$ is sufficient to explain all dependencies among the y's. In other words, conditioning on $\mathbf{z}$, the y's are independent. Therefore, their joint distribution $g(\mathbf{y} \mid \mathbf{z})$ can be expressed as the multiplication of their marginal distributions, conditioning on $\mathbf{z}$, as follows

$$g(\mathbf{y} \mid \mathbf{z}) = \prod_{i=1}^{p} g_i(y_i \mid \mathbf{z}),$$

and thus $f(\mathbf{y})$ can be written as

$$f(\mathbf{y}) = \int \left[\prod_{i=1}^{p} g_i(y_i \mid \mathbf{z}) \right] h(\mathbf{z}) d\mathbf{z}. \tag{2.6}$$

Parameters of the model (α_i and λ_{ij}s) are estimated by maximum likelihood based on the joint distribution of the manifest variables given by equation (2.6). For a random sample of size n, the loglikelihood is written as

$$L = \sum_{m=1}^{n} \log f(\mathbf{y}_m)$$

$$= \sum_{m=1}^{n} \log \int \left[\prod_{i=1}^{p} g_i(y_i \mid \mathbf{z}) \right] h(\mathbf{z}) d\mathbf{z}.$$

$$= \sum_{m=1}^{n} \log \int \left[\prod_{i=1}^{p} \{1 - \pi_i(\mathbf{z})\} \exp\{y_i(\alpha_i + \sum_{j=1}^{q} \lambda_{ij} z_j)\} \right] h(\mathbf{z}) d\mathbf{z}. \tag{2.7}$$

The loglikelihood given in equation (2.7) is differentiated with respect to the model parameters α_i, λ_{ij}, where the resulting integral cannot be found analytically and is approximated numerically using techniques such as Gauss-Hermite quadrature. Equating the partial derivatives to zero, ML equations are obtained. For non-normal manifest variables, the ML equations are nonlinear and estimates for parameters are found by solving the equations using an iterative scheme, such as Newton-Raphson. The maximisation of the loglikelihood (2.7) can be done by an

EM algorithm explained in Moustaki and Knott (2000a). Alternatively, Bayesian estimation using Markov Chain Monte Carlo (MCMC) can be used (see Patz and Junker (999a, 1999b)).

2.3 Goodness-of-Fit

A latent variable model is accepted as a good fit when the latent variables account for most of the associations among the observed variables. Testing whether the model provides a good fit for the data involves comparing observed frequencies and estimates of expected frequencies under the model being tested. However, a global goodness-of-fit measure that compares frequencies for full response patterns cannot be obtained under a limited information likelihood estimation approach that only estimates pairwise probabilities assuming underlying bivariate normality. Alternatively, instead of looking at whole response patterns, one may consider two-way margins. The likelihood ratio test statistic is given by

$$X^2_{LR} = 2\sum_{s_1=1}^{c_1}\sum_{s_2=1}^{c_2} n_{s_1s_2} ln\,[p_{s_1s_2}/\hat{\pi}_{s_1s_2}] = 2N\sum_{s_1=1}^{c_1}\sum_{s_2=1}^{c_2} p_{s_1s_2} ln\,[p_{s_1s_2}/\hat{\pi}_{s_1s_2}] = 2NF(\hat{\boldsymbol{\theta}}),$$

where $\hat{\boldsymbol{\theta}}$ is the estimated parameter vector and $\hat{\pi}_{s_1s_2} = \pi_{s_1s_2}(\hat{\boldsymbol{\theta}})$. If the model holds, the above statistic has an approximate chi-square distribution with $c_1c_2 - c_1 - c_2$ degrees of freedom (Jöreskog (2005)). By adding up all univariate and bivariate X^2_{LR}s, an overall likelihood ratio statistic is obtained. The alternative goodness-of-fit statistic

$$X^2_{GF} = \sum_{s_1=1}^{c_1}\sum_{s_2=1}^{c_2}[(n_{s_1s_2} - N\hat{\pi}_{s_1s_2})^2/(N\hat{\pi}_{s_1s_2})] = N\sum_{s_1=1}^{c_1}\sum_{s_2=1}^{c_2}(p_{s_1s_2} - \hat{\pi}_{s_1s_2})^2/\hat{\pi}_{s_1s_2},$$

has the same asymptotic distribution as X^2_{LR} when the fit is good.

Likelihood ratio and goodness-of-fit tests can be greatly distorted in cases of sparseness in contingency tables leading to unreliable estimates, especially for binary variables (Jöreskog (2005)). The test statistics are sensitive to sample sizes too. Large sample sizes lead to large values thus rejecting models even if the difference between the sample and fitted covariance matrices is small. On the other hand, small sample sizes lead to small values thus failing to reject the model due to lack of evidence.

The Root Mean Squared Error of Approximation (RMSEA) is a more robust measure, first introduced by Steiger (1990), that is based on the non-central chi-square distribution and tests whether the model holds "approximately". Values of thr RMSEA greater than 0.1 are indications of poor fit. One advantage of the RMSEA is that it is usually reported with a confidence interval. The Comparative Fit Index (CFI) is another fit index, proposed by Bentler (1990), that compares the sample covariance matrix to a null model that assumes all latent variables are uncorrelated. Values of the CFI range between 0.0 and 1.0, with values closer to 1.0 indicating good fit. Hooper et al. (2008) provide a list of available fit indices for structural equation modelling in the literature, along with guidelines on their use.

An alternative way that does not provide a test statistic but rather focuses on measurement of fit is proposed by Jöreskog and Moustaki (2001). It measures the fit to the univariate and bivariate marginal distributions and allows pointing out the source of lack of fit by defining a LR-fit and a GF-fit for each category of the univariate and bivariate contingency tables. These represent the individual cell contributions to the LR and GF-fits and they do not have a chi-square distribution.

For category s of variable i , the LR and GF-fits are defined as

$$LR - fit_s^{(i)} = 2np_s^{(i)} \ln(p_s^{(i)}/\hat{\pi}_s^{(i)}),$$

$$GF - fit_s^{(i)} = n(p_s^{(i)} - \hat{\pi}_s^{(i)})^2/\hat{\pi}_s^{(i)}.$$

Summing these over s gives the univariate LR- and GF-Fits for variable i. Similarly, the bivariate LR and GF-fits for category s_1 of variable i and category s_2 of variable i' are defined as

$$LR - fit_{s_1 s_2}^{(ii')} = 2np_{s_1 s_2}^{(ii')} \ln(p_{s_1 s_2}^{(ii')}/\hat{\pi}_{s_1 s_2}^{(ii')}),$$

$$GF - fit_{s_1 s_2}^{(ii')} = n(p_{s_1 s_2}^{(ii')} - \hat{\pi}_{s_1 s_2}^{(ii')})^2/\hat{\pi}_{s_1 s_2}^{(ii')}.$$

Summing over s_1 and s_2 gives the bivariate LR and GF-fits for variables i and i'. Since each of these fit measures is based on a different contingency table with a different number of cells, they are divided by the number of cells to allow for comparison across variables and pairs of variables. The overall fit measure is the average of all pairwise fit measures. Jöreskog and Moustaki (2001) suggest considering a value larger than 4 to indicate a poor fit. Cells with large contributions to the LR or GF-statistics will be nominated as the source of bad fit. Bartholomew and Tzamourani (1999) propose alternative ways for assessing the goodness-of-fit of this model based on Monte Carlo methods and residual analysis.

Criterion for selecting among a set of nested models could be used instead of goodness-of-fit tests. Akaike Information Criterion (AIC) takes into account both the value of the likelihood at the maximum likelihood solution and the number of estimated parameters

$$\text{AIC} = -2[\max L] + 2m,$$

where m is the number of estimated parameters. AIC can be used to compare models with different numbers of latent factors, where the model with the smallest AIC is taken to be the best.

2.4 Latent Variable Models for Multivariate Longitudinal Data

When a single variable is measured repeatedly over time, the data is said to be *longitudinal*. Modelling univariate longitudinal data will be discussed briefly in Chapter 3. However, when interest lies in capturing the evolution of a latent construct over time, latent variable models are used. The latent variables are measured via a number of observed items at each time point. When dealing with such models, there are two types of relationships to account for; those between different items within the same time point and those between the same items at different time points. At a given time, one or more latent variables can be used to account for dependencies among items, as outlined earlier in this chapter for cross-sectional data. The structural part of the model in this case addresses the question: how should the latent variables be linked in order to capture the longitudinal nature of the data?

Within a SEM framework, Jöreskog (2005) discusses a confirmatory SEM for multivariate ordinal longitudinal data where measurement invariance is assumed by setting thresholds and loadings of the same items to be equal over time. Dependence among latent constructs over time is captured by regressing a latent variable on the same latent variable measured at a preceding time point. Measurement errors for the same items are correlated over time to account for their repitition.

Within an IRT approach, Dunson (2003) proposes a generalised latent trait model that accommodates different types of observed items and accounts for dependencies within time using time-specific factors and across time using item-specific random effects. A linear transition model is used to link the latent variables. Inference is carried out using a Bayesian approach. Cagnone et al. (2009) use a similar framework for modelling multivariate longitudinal ordinal responses where a first order autoregressive structure is used to link latent variables over time whereas item-specific random effects or a single common factor are used to account for associations of items across time. An EM algorithm is used for estimation.

Chapter 3

Missing Data: Review of Literature

Standard statistical techniques are designed to analyse *complete* datasets. They are developed under the assumption that the values for all variables recorded for all observations in the dataset are present. In practice, this is not usually the case. It is very often when dealing with a real dataset that some of the values are missing.

There are different types of missing data. Unit non-response is a severe type of missingness where data for an observation is completely missing, and thus no information can be inferred about this observation. Item non-response is another type of missing data where data for a respondent is collected for some variables but is missing for others. Intermittent missingness and dropout are two types of missingness specific to longitudinal data. Intermittent missingness occurs when a subject misses one or more occasions of a longitudinal study, but shows up on subsequent waves. Dropout is a more common type of missingness in longitudinal studies where subjects exit the study prematurely.

In this chapter, we present existing literature on missing data in general, followed by a review of how dropout is treated in longitudinal studies.

3.1 Missingness in Cross-sectional Data

Rubin (1976) and Little and Rubin (2002) classify missing data into:

1. Data Missing Completely At Random (MCAR) where missingness is independent of both observed and unobserved data.

2. Data Missing At Random (MAR) such that missingness depends on the observed data, but is independent of the unobserved.

3. Data Missing Not At Random (MNAR) where missingness depends on the unobserved data, and possibly the observed data as well.

When data is missing completely at random (MCAR), it is reasonable to think of the observed data as a random subset of the complete data. If data is missing at random (MAR), it can still be viewed as a random subset defined for different values of the observed data. In these cases, the missingness mechanism is said to be *ignorable*. For the case when data is missing not at random (MNAR), the missingness depends on the missing value itself and possibly on observed outcomes too, hence it is said that the missingness mechanism is *non-ignorable* or *informative*.

It is possible to test whether the MCAR assumption is met. For example, one could compare men and women to test whether they differ in the proportion of cases with missing data on income. Any such difference would be a violation of MCAR. However, it is impossible to test whether the data are MAR or MNAR. For obvious reasons, after controlling for observed variables, one cannot tell whether respondents with higher income are more likely than those with lower income to have missing data on income (Allison (2012)).

There are various ways in the literature for dealing with missing values, the simplest of which is complete-case analysis; also known as *listwise deletion* in which incompletely recorded units are discarded and a case is included in the analysis only if it is fully observed on all variables. Simplicity is the main advantage of this method. However, it can involve a great loss of information since values of a certain variable are discarded when they belong to cases that are missing for other variables. It can thus lead to serious biases and is not efficient except when the data is MCAR. Available-case analysis is a possible alternative for complete-case analysis that includes all cases where the variable of interest is recorded, thus making use of all available information when making inference on a single variable. The main limitation of this method is that the sample base is not the same from one variable to another. This variability can cause practical problems and does not allow for comparability across variables if the missingness mechanism is not MCAR. A natural extension to accommodate meausures of covariation is pairwise deletion, in which a case remains in the analysis if the pair of variables being referenced have complete data for that case.

Weighting procedures for missing data use weights for observed units in an attempt to adjust for bias. This is a relatively simple device for reducing bias from complete-case analysis by yielding the same weight for all variables measured for each case. On the other hand, this simplicity entails a cost, in that weighting generally involves an increase in variance and is thus inefficient (Little and Rubin (2002)).

Imputation is a widely used method in which the missing values are filled in with one of several options and analysis is carried out with the imputed data as if the dataset was completely observed (Little and Rubin (2002)). Single imputation can be applied to impute one value for each missing item. Options for imputation include unconditional or conditional mean imputation, where means from observed values of a variable or conditional means given data observed on other variables are substituted respectively; imputation by regression, where the missing variables for a unit are estimated by predicted values from the regression on the known variables for that unit; and hot deck imputation, where recorded units in the sample are used to substitute missing values. An important drawback of single imputation methods is that they do not account for imputation uncertainty and thus standard variance formulas applied to the imputed data systematically underestimate the variance of estimates, even if the model used to generate the imputations is correct.

Multiple Imputation (MI) has the added bonus of largely correcting this disadvantage by imputing each missing value by more than one value, allowing for appropriate assessment of imputation uncertainty and increasing the efficiency of estimates compared to those obtained from single imputation. MI was first introduced by Rubin (1978) and keynote references include Rubin (1987) and Rubin (1996). When MI is implemented, each missing value is replaced by a vector of D imputed values. Replacing each missing value by the first component in its vector of imputations creates the first completed data set, replacing each missing value by the second component in its vector creates the second completed data set, and so on. Standard complete-data methods are then used to analyse each of the D imputed data sets. The D sets of imputations can be viewed as repeated random draws from the predictive distribution of the missing values under a particular model for non-response. In that case, the D complete-data inferences can be combined to form one inference that properly reflects uncertainty due to missingness under that model, where the standard error adjusted for imputation accounts both for variation within and between imputed sets. When imputations are from two or more models for non-response, the combined inferences under the models can be contrasted across models to display the sensitivity of inference to models for non-response. MI provides statistically valid estimates, unilke simple ad hoc methods for handling missing data such as complete-case analysis, available-case analysis or mean imputation which only return valid estimates under the assumption that data is MCAR (Rubin (1996)).

A different approach for handling missing data is to rely on model-based procedures in which a model is defined for missingness. Inference is then made on the likelihood defined under that model. A brief review of likelihood-based estimation for data with missingness is given below.

3.1.1 Likelihood-Based Estimation for Data With Missing Values

Little and Rubin (2002) present a likelihood-based estimation method for data with missing values. Let y denote the complete data with no missing values such that y can be written as $y = (y_{obs}, y_{mis})$ where y_{obs} denotes the observed values and y_{mis} denotes the missing values. The probability density function of y with a scalar or vector parameter θ; $f(y \mid \theta)$, can be written as a joint distribution of y_{obs} and y_{mis} as $f(y_{obs}, y_{mis} \mid \theta)$. The marginal distribution of y_{obs} is obtained by integrating out y_{mis} from the joint probability density function

$$f(y_{obs} \mid \theta) = \int f(y_{obs}, y_{mis} \mid \theta) dy_{mis}. \tag{3.1}$$

The missingness mechanism can be incorporated in the model by introducing an indicator random variable for missingness. For the i^{th} variable of the m^{th} observation, an indicator random variable r_{mi} is defined as

$$r_{mi} = \begin{cases} 1, & y_{mi} \ \textit{observed}, \\ 0, & y_{mi} \ \textit{missing}. \end{cases}$$

The joint distribution of y and r, a $(p \times 1)$ vector of missingness indicators, can be written as

$$f(y, r \mid \theta, \psi) = f(y \mid \theta) f(r \mid y, \psi),$$

where $f(r \mid y, \psi)$ is the distribution of the missingness mechanism and ψ is an unknown parameter related to the missingness process. What is actually observed is values for y_{obs} and r. To obtain the distribution of (y_{obs}, r), we integrate out y_{mis} from the joint distribution of $y = (y_{obs}, y_{mis})$ and r:

$$f(y_{obs}, r \mid \theta, \psi) = \int f(y_{obs}, y_{mis} \mid \theta) f(r \mid y_{obs,} y_{mis}, \psi) dy_{mis}. \tag{3.2}$$

When the missingness mechanism does not depend on the missing values y_{mis}; that is to say data is MAR, $f(r \mid y_{obs}, y_{mis}, \psi) = f(r \mid y_{obs}, \psi)$, and equation (3.2) can be written as

$$f(y_{obs}, r \mid \theta, \psi) = f(r \mid y_{obs,}, \psi) \int f(y_{obs}, y_{mis} \mid \theta) dy_{mis}$$

$$= f(r \mid y_{obs}, \psi) f(y_{obs} \mid \theta).$$

In some cases, the parameters θ and ψ are said to be *"distinct"* in the sense that their joint parameter space is the product of the parameter space of θ and that of ψ. In this case, the likelihood-based inferences for θ from the likelihood $L(\theta, \psi \mid y_{obs}, r) = f(y_{obs}, r \mid \theta, \psi)$ is the same as that from the simpler $L(\theta \mid y_{obs}) = f(y_{obs} \mid \theta)$. That is, if the data is MAR (missingness does not depend on the missing values), and the parameters θ and ψ are distinct, the missingness mechanism is ignorable when inferences are made about θ. Inferences can then be based on equation (3.1) rather than (3.2). However, it is not always easy to justify the assumption of random missingness. See Little and Rubin (2002) and Verbeke and Molenberghs (2000) for details.

The procedure for maximum likelihood estimation of parameters of an incomplete dataset will be the same as that of a complete dataset with the difference that it is based on the observed part of the data. The likelihood function is derived, and the ML estimates are obtained by solving the likelihood equations. The following are three available approaches to ML estimation with missing data.

3.1.1.1 Factoring the Likelihood

Assuming the missing-data mechanism is ignorable, the loglikelihood $l(\theta \mid y_{obs}) = \log f(y_{obs} \mid \theta)$ based on the incomplete data may be rewritten as

$$l(\phi \mid y_{obs}) = l_1(\phi_1 \mid y_{obs}) + l_2(\phi_2 \mid y_{obs}) + ... + l_J(\phi_J \mid y_{obs}),$$

where ϕ is a one-to-one monotone function of θ and $\phi_1, \phi_2, ..., \phi_J$ are distinct parameters (Little and Rubin (2002)). If this decomposition can be found, $l(\phi \mid y_{obs})$ can be maximised by maximising $l_j(\phi_j \mid y_{obs})$ separately. This is done by factoring the likelihood into functions that depend on distinct ϕ_j's. However, this factorisation does not always exist. Moreover, it can even exist but with non-distinct parameters ϕ_j's; and thus maximising the factors separately does not maximise the likelihood.

3.1.1.2 Direct ML

Assuming the data is MAR (missingness mechanism is ignorable), ML estimates can be found by solving the score function

$$S(\theta \mid y_{obs}) = \frac{\partial l(\theta \mid y_{obs})}{\partial \theta} = 0.$$

If a closed-form solution can be found for the above equation, ML estimates are obtained directly. If no closed-form solution can be attained, iterative methods are used to obtain the ML estimates. Some of these iterative methods, such as the Newton-Raphson algorithm, require calculating second derivatives of θ which can become rather complicated.

3.1.1.3 The EM Algorithm

The Expectation-Maximisation (EM) algorithm, first introduced by (Dempster et al. (1977)), is an iterative algorithm for ML estimation. It consists of two steps: the E-step which finds the conditional expectation of the loglikelihood $l(\theta \mid y)$ - or functions of y_{mis} appearing in the complete data loglikelihood, given the observed data and the current estimate of the parameter θ, say $\theta^{(t)}$

$$E\left(l(\theta \mid y) \mid y_{obs}, \theta = \theta^{(t)}\right) = \int l(\theta \mid y) f(y_{mis} \mid y_{obs}, \theta = \theta^{(t)}) dy_{mis},$$

and the M-step which maximises this expected loglikelihood thus determining the new parameter $\theta^{(t+1)}$. The new estimated parameter is substituted back in the E-step and the above steps are repeated until convergence is attained. The expectations calculated in each step involve estimating functions of the missing data y_{mis} which are then used to re-estimate the parameters. The new parameters are used to re-estimate functions of the missing values, and so on. Here data are assumed to be MAR, but in case the missingness is not at random, a factor representing the missing-data mechanism has to be included in the model. See examples for clarification in Little and Rubin (2002).

Bayesian Estimation

The likelihood function also plays a central role in Bayesian inference. In the Bayesian approach, parameters θ are treated as random variables rather than fixed quantities, and uncertainty about them is quantified using probability distributions. A parameter θ is assigned a prior distribution $h(\theta)$, and inference about θ after observing the data y is based on its posterior distribution $h(\theta \mid y)$, which is given by Bayes' theorem as:

$$h(\theta \mid y) = \frac{h(\theta) L(y \mid \theta)}{f(y)}.$$

Point estimates of θ can be obtained as measures of the center of the posterior distribution. Little and Rubin (2002) outline the strong parallels between Bayesian and likelihood inference in case of large samples, and point out that Bayesian estimates correspond to ML estimates of θ when the prior distribution is uniform. Furthermore, they highlight that in principle there is no difference between ML or Bayes inference for incomplete data and ML or Bayes inference for complete data. The likelihood for the parameters based on the incomplete data is derived, ML estimates are found by solving the likelihood equation,while in case of Bayesian inference the posterior distribution is obtained by incorporating a prior distribution and performing the necessary integrations. More on Bayesian estimation is given in Chapter 6.

To sum up, there are two main competing approaches with very similar statistical properties for handling missing data, namely MI and ML. Under the assumption that data are MAR, both approaches return valid estimates that are consistent and asymptotically efficient. One main advantage of MI is to provide ultimate users of the data, with varying models and degrees of statistical and computing capabilities, with data sets that can be analysed with standard complete-data techniques without the need for special knowledge and techniques for handling missing data. This is particularly useful when database constructors and ultimate users are distinct entities, so that database constructors may focus on handling the missing data problem while ultimate users focus on their substantive scientific analysis for which missing data are merely a nuisance factor (Rubin (1996)). On the other hand, Allison (2012) highlights the strength points for ML-based procedures over MI, the most important of which being that there is always a potential conflict between the imputation model and the analysis model when MI is implemented while no such conflict exists in ML because everything is done under one model. Possible sources of incompatability that can cause serious bias in estimates when

MI is used include cases such as: when the analysis model contains variables that were not included in the imputation model, and when the analysis model contains interactions and non-linearities, but the imputation model is strictly linear, and so on... Another point is that ML is more efficient than MI, where in case of MI full efficiency requires an infinite number of data sets to be analysed, which is not possible. Moreover, for a given set of data, ML always produces the same result, while MI gives a different result every time it is implemented. This particular drawback of MI can be overcome by increasing the number of imputed data sets. On comparing five methods (mean imputation, regression imputation, MI, EM and ML) for dealing with missing data in SEM with respect to the percent of bias in estimating parameters, Olinsky et al. (2003) find ML to be superior in the estimation of most different types of parameters, followed by MI which is found superior in estimating standard errors but suffers with increasing percentage of missing data.

3.1.2 Models for Data Missing Not At Random (MNAR)

So far, data has been assumed to be MAR. However, when the missingness is non-ignorable, the missingness mechanism should be incorporated in the model as in equation (3.2). There are two cases when missing data is non-ignorable:

1. *The missing data mechanism is non-ignorable but known.* The conditional distribution of r given $y = (y_{obs}, y_{mis})$ depends on y_{mis}, but does not depend on unknown parameters ψ.

   An example with a known non-ignorable mechanism is given in Little and Rubin (2002) for incomplete data that are created by censoring at some known censoring point c, so that only values less than c are recorded. The missingness mechanism is given by

$$f(r \mid y, \psi) = \prod_{m=1}^{n} f(r_m \mid y_m, \psi),$$

where only n' respondents are observed; $r_m = 1$, $m = 1, ..., n'$ and $r_m = 0$, $m = n' + 1, ..., n$. Hence,

$$f(r_m \mid y_m, \psi) = \begin{cases} 1, & r_m = 1 \text{ and } y_m < c, \quad \text{or } r_m = 0 \text{ and } y_m \geq c, \\ 0, & otherwise, \end{cases}$$

This leads to a likelihood function involving an exponential distribution that depends on the known censoring point c, but not on unknown parameters ψ defining the missingness process:

$$\begin{aligned} f(y_{obs}, r \mid \theta) &= \prod_{m=1}^{n'} f(y_m, r_m \mid \theta) \prod_{m=n'+1}^{n} f(r_m \mid \theta) \\ &= \prod_{m=1}^{n'} f(y_m \mid \theta) \Pr(y_m < c \mid y_m) \prod_{m=n'+1}^{n} \Pr(y_m \geq c \mid \theta) \\ &= \theta^{-n'} \exp\left(-\sum_{m=1}^{n'} \frac{y_m}{\theta}\right) \exp\left(-\frac{(n-n')c}{\theta}\right), \end{aligned}$$

since $\Pr(y_m < c \mid y_m) = 1$ for respondents and $\Pr(y_m \geq c \mid \theta) = \exp\left(-\frac{c}{\theta}\right)$ for non-respondents, using properties of the exponential distribution.

2. *The missing data mechanism is non-ignorable and unknown.* The conditional distribution of y depends both on y_{mis} and unknown parameters ψ. In

practice, most non-ignorable missing data mechanisms will also be unknown as non-reponse will usually be related in some unknown way to the missing values even after adjusting for covariates known for both respondents and non-respondents.

Although the more general MNAR assumption explicitly incorporates the missingness mechanism, inferences produced are based on untestable assumptions about the distribution of the unobserved data given the observed. Methods involving non-ignorable missing data should always be viewed as part of a sensitivity analysis in which the consequences of different modelling assumptions are explored (Little and Rubin (2002)).

3.2 Missingness in Longitudinal Data

In this section, we first look at how longitudinal data is modelled, has the data been fully observed. Then, we review models for longitudinal data subject to dropout, the most common form of missing data in longitudinal studies.

3.2.1 Modelling Complete Univariate Longitudinal Data

When a variable is measured repeatedly for each subject in a study to monitor its evolution over time, the data is said to be *longitudinal*. Analysing such type of data, correlations among measurements of the same subject over time should be accounted for. Each individual in the study is affected by average trends that affect the whole population. These are called *population-specific* or *fixed effects.* One is also affected by *subject-specific* or *random effects* that are unique or specific to this subject in particular.

For an individual m, longitudinal data is modelled using general linear mixed-effects models (see for example Verbeke and Molenberghs (2000), Diggle et al. (2013) and Molenberghs and Verbeke (2001)) of the form

$$\mathbf{y}_m = \mathbf{X}_m\boldsymbol{\beta} + \mathbf{W}_m\mathbf{b}_m + \boldsymbol{\varepsilon}_m, \quad m = 1, \ldots, n; \tag{3.3}$$

where $\mathbf{y}_m = (y_{m1}, y_{m2}, ..., y_{mT_m})'$ is a vector of all repeated measurements for the mth subject at T_m occasions, $\mathbf{X}_m$ is a $(T_m \times r_1)$ matrix of known covariates, $\boldsymbol{\beta}$ is an $(r_1 \times 1)$ vector of corresponding regression parameters (fixed effects), $\mathbf{W}_m$ is a $(T_m \times r_2)$ matrix of subject-specific covariates that usually includes time as a covariate modelling how y_m evolves over time, $\mathbf{b}_m$ is an $(r_2 \times 1)$ vector of subject-specific parameters (random effects) describing how the evolution of the mth subject deviates from the average evolution in the population, and $\boldsymbol{\varepsilon}_m$ is a $(T_m \times 1)$ vector of residuals for the mth subject. This model assumes that the vector of repeated measurements for each subject follows a linear regression with some population-specific parameters $\boldsymbol{\beta}$ and some subject-specific parameters $\mathbf{b}_m$.

It is also assumed that $\mathbf{b}_m \sim N(0, \mathbf{B})$, $\boldsymbol{\varepsilon}_m \sim N(0, \sigma^2\mathbf{I}_{T_m})$, and that $\mathbf{b}_m$ and $\boldsymbol{\varepsilon}_m$ are independent. Conditioning on the random effect $\mathbf{b}_m$, $\mathbf{y}_m$ follows a normal distribution with mean $\mathbf{X}_m\boldsymbol{\beta} + \mathbf{W}_m\mathbf{b}_m$ and covariance matrix $\Sigma_m = \sigma^2\mathbf{I}_{T_m}$. The marginal density function of $\mathbf{y}_m$, which is given by

$$g(\mathbf{y}_m \mid \mathbf{X}_m) = \int g(\mathbf{y}_m \mid \mathbf{X}_m, \mathbf{b}_m)h(\mathbf{b}_m)d\mathbf{b}_m,$$

will have a normal distribution with mean vector $\mathbf{X}_m\boldsymbol{\beta}$ and covariance matrix $\mathbf{V}_m = \mathbf{W}_m\mathbf{B}\mathbf{W}_m' + \Sigma_m$ where $\Sigma_m = \sigma^2\mathbf{I}_{t_m}$ (Verbeke and Molenberghs (2000)).

Molenberghs et al. (2004) consider adding a term $\mathbf{u}_m$ to account for serial correlation in equation (3.3):

$$\mathbf{y}_m = \mathbf{X}_m\boldsymbol{\beta} + \mathbf{W}_m\mathbf{b}_m + \mathbf{u}_m + \boldsymbol{\varepsilon}_m,$$

where $\mathbf{u}_m$ is assumed a normal distribution; $N(0, \sigma_u^2\mathbf{H}_m)$. The serial covariance matrix $\mathbf{H}_m$ only depends on m through the number of observations T_m and the time points at which measurements are taken. The structure of the matrix $\mathbf{H}_m$ is determined through the autocorrelation function $\rho(t_m - t'_m)$. This function decreases such that $\rho(0) = 1$ and $\rho(t) \to \infty$ as $t \to \infty$. In this case the covariance matrix of the marginal distribution of $\mathbf{y}_m$ becomes $\mathbf{V}_m = \mathbf{W}_m\mathbf{B}\mathbf{W}'_m + \Sigma_m$, where $\Sigma_m = \sigma^2\mathbf{I}_{T_m} + \sigma_u^2\mathbf{H}_m$ combines the measurement errors and the serial components.

Longitudinal data can be viewed as two-level data, where occasions are nested within subjects. Rabe-Hesketh and Skrondal (2012) present a comprehensive review of models for univariate longitudinal data within this context, including fixed-effects models where unobserved between-subject heterogeneity is represented by fixed subject-specific effects, random-effects models where unobserved between-subject heterogeneity is represented by subject-specific effects that vary randomly, and dynamic models where the response at a given occasion depends on previous or lagged responses. In practice, different disciplines adopt different modelling strategies for longitudinal data. A summary of some of the main modelling techniques for longitudinal data presented by Rabe-Hesketh and Skrondal (2012) is given below.

A fixed-effects model may include a subject-specific fixed intercept or fixed coefficient for some of the covariates, or both. An example of a fixed-effects model would be

$$y_{mt} = \beta_0 + \alpha_{1m} + \beta_1 x_{1m} + \beta_2 x_{2m} + (\beta_3 + \alpha_{2m}) x_{3m} + \varepsilon_{mt},$$

where α_{1m} and α_{2m} are fixed subject-specific intercept and slope parameters, respectively, x_1 and x_2 are covariates having the same effect for all subjects, and x_3 is a covariate having a subject-specific effect. Fixed-effects models are used to estimate average within-subject relationships between time-varying covariates and the response variable, where every subject acts as its own control..

Similarly, a random-effects model may include a subject-specific random intercept resulting in a random-intercept model, or random coefficient for some of the covariates resulting in a random-coefficient model, or both. An example of a random-effects model would be

$$y_{mt} = \beta_0 + \zeta_{1m} + \beta_1 x_{1m} + \beta_2 x_{2m} + (\beta_3 + \zeta_{2m}) x_{3m} + \varepsilon_{mt},$$

where ζ_{1m} and ζ_{2m} are random subject-specific intercept and slope parameters, respectively, x_1 and x_2 covariates having the same effect for all subjects, and x_3 is a covariate having a subject-specific effect. Random-effects models explain individual differences by allowing subject-specific relationships to vary randomly around average trends of the population. A special case of random effects-models are growth curve models, in which time is always given a random coefficient to model individual growth trajectories. Random-effects models are widely used in areas of psychology and education, where both the nature and reasons for variability are of interest.

Dynamic models, also known as lagged-response models, autoregressive-response models and Markov models, model the response variable as a function of the same variable at previous occasions. One of the most widely used dynamic models is the first-order autoregressive model [AR(1)], where the response $y_{m,t}$ is regressed on the preceding response $y_{m,t-1}$. An example of an [AR(1)] model is

$$y_{mt} = \beta_0 + \phi y_{m,t-1} + \beta_1 x_{1m} + \beta_2 x_{2m} + \beta_3 x_{3m} + \varepsilon_{mt}. \tag{3.4}$$

Conditional on covariates, residuals ε_{mt} are assumed to be uncorrelated. This model assumes that all the within-subject dependence is due to the lagged response. It is noted that a fixed autoregressive parameter ϕ is used here. It is appropriate to use such a model when occasions are equally spaced in time. Otherwise, it would seem unrealistic or unjustifiable to assume that the lagged response has the same effect on the current response regardless of the time interval between them. The above model can be extended to have a time-dependent autoregressive parameter ϕ_t instead of ϕ, to accommodate unequal spacing in time. Another extension can combine a lagged-response model with a random (or fixed) intercept as follows

$$y_{mt} = \beta_0 + \zeta_{1m} + \phi y_{m,t-1} + \beta_1 x_{1m} + \beta_2 x_{2m} + \beta_3 x_{3m} + \varepsilon_{mt}, \tag{3.5}$$

in order to distinguish between two explanations of within-subject dependence over time; namely unobserved heterogeneity where individual differences affect both past and future responses (represented by the random intercepts), and state dependence where previous responses somehow determine /affect future responses (represented by the lagged responses). Dynamic models are very popular among economists.

3.2.2 Longitudinal Data Subject to Dropout

The problem of missing data is very common in longitudinal studies. A distinction should be made here between *intermittent* missing values where a subject has missing data at some of the waves of the study, and *dropout* or *attrition* where missing values are only followed by missing values (i.e. subjects exit the study prematurely and never come back).

Most of the literature on longitudinal data is restricted to dropout as it is the most common type of missingness to appear in longitudinal studies. Again the simplest way to deal with dropout is to discard incomplete cases. However, this can be very misleading especially in cases when there is a systematic difference between subjects who stay in the study and those who dropout, that is to say dropout is not at random. Another way to overcome the problem of dropout is using imputation. Methods for imputing longitudinal data include: replacing a missing observation by the mean of non-missing subjects with the same covariates, carrying the last available measurement of the subject onwards, and regressing the missing value on available past data. However, as before-mentioned in Section 3.1, these naïve approaches are only valid under the assumption that data are MCAR, and are thus not recommended as they systemetically lead to underestimation of the variance (between-individual variance in case of mean imputation and within-individual variance in case of last observation carried forward). Neither complete-case analysis nor imputation is completely natural here. The most natural thing is to use all available data for each subject for as long as they last, without further adjustment, thus justifying our choice of likelihood-based approaches to handling dropout. This approach is fully efficient when dropout is at random.

When modelling longitudinal data subject to dropout, the joint density function of both the measurement and dropout processes is considered. Suppose dropout occurs at occasion t of a longitudinal study. If dropout neither depends on observed history data nor on the currently unobserved value, then missingness is considered to be completely at random. However, if the probability of dropout depends on previously observed values, but not on the currently missing value, then dropout is considered to be at random. If dropout is at random (or completely at random), and the parameters of the dropout process are distinct from those of the measurement process (an assumption we make throughout), the dropout is said to be ignorable and a valid analysis can be based on a likelihood that ignores the dropout mechanism. However, it is not always easy to justify the assumption of random dropout. If dropping out depends on the unobserved value at time of dropout, data is MNAR as in that case the missingness depends on the missing value itself which implies a systematic difference between respondents who remain in the study and those who drop out. For example, in a medical study, those dropping out may be those with a deteriorating health condition. Hence the dropout mechanism is non-ignorable and should be incorporated in the analysis of the data, as ignoring it may lead to biased estimates of the parameters of interest.

There are three general model-based approaches for modelling univariate longitudinal data subject to dropout, that are outlined below. As a start, let us define a time of dropout variable k_m that denotes the occasion at which subject m drops out - in case of an incomplete sequence - and that is equal $T_m + 1$ in case of a complete sequence. Since in case of dropout, r_m is of the form $(1, ..., 1, 0, ..., 0)$, k_m is then given by

$$k_m = 1 + \sum_{t=1}^{T_m} r_{mt}.$$

3.2.2.1 Selection Models

Selection models factorise the joint density of the full data $f(\mathbf{y}_m, k_m \mid \mathbf{X}_m, \mathbf{W}_m, \theta, \psi)$ into the product of the marginal density of the measurement process and the conditional density of the missingness mechanism given the measurement as follows

$$f(\mathbf{y}_m, k_m \mid \mathbf{X}_m, \mathbf{W}_m, \theta, \psi) = f(\mathbf{y}_m \mid \mathbf{X}_m, \mathbf{W}_m, \theta) f(k_m \mid \mathbf{y}_m, \mathbf{X}_m, \psi).$$

It is possible to have additional covariates in the missingness model but this is suppressed from notation (Molenberghs and Verbeke (2001)). In their key paper on selection models for non-ignorable dropout, Diggle and Kenward (1994) combine a multivariate Gaussian linear model for the measurement process with a logistic dropout model. A general model for informative dropout in longitudinal data for which completely random and random dropouts are special cases is proposed, and an associated methodology for likelihood-based inference is developed. A linear mixed model of the form (3.3) is assumed to model the measurement process. Assuming that the first measurement y_{m1} is obtained for every subject in the study, the model for the dropout process is based on a logistic regression for the probability of dropout at occasion t, given the subject was still in the study up to occasion t. Let $g(\mathbf{h}_{mt}, y_{mt})$ denote this probability, where the history vector $\mathbf{h}_{mt}$ contains all responses and covariates observed up to but not including occasion t.

The dropout process is thus modelled by a logistic linear model of the form

$$\text{logit}\left[g_t(\mathbf{h}_{mt}, y_{mt})\right] = \text{logit}\left[\Pr\left(k_m = t \mid k_m \geqslant t,\, \mathbf{y}_m\right)\right] = \tilde{\mathbf{h}}_{mt}\boldsymbol{\Psi} + \psi_1 y_{mt}$$

$$= \psi_0 + \psi_1 y_{mt} + \sum_{l=2}^{t} \psi_l y_{m,t+1-l}, \tag{3.6}$$

where $\tilde{\mathbf{h}}_{mt}$ is a suitable subset of $\mathbf{h}_{mt}$. If $\psi_1 = 0$, the dropout process is random, since the dropout will depend only on history and not on y_{mt}. In this case, the measurement and dropout models can be fitted separately. If $\psi_1 \neq 0$, the dropout is nonrandom, since the dropout will depend on y_{mt}, and the measurement and dropout models cannot be fitted separately.

Suppressing the index m for a subject, let $\mathbf{y}^* = (y_1^*, y_2^*, ..., y_T^*)^T$ denote the complete vector of measurements at T time points. Let $\mathbf{y} = (y_1, y_2, ..., y_T)^T$ denote the vector of observed measurements with missing values recorded as 0. Therefore, it can be said that

$$y_t = \begin{cases} y_t^*, & t = 1, ..., k-1, \\ 0, & t \geqslant k, \end{cases}$$

where $2 \leqslant k \leqslant T$ identifies the dropout time. Let $f^*(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\phi})$ denote the joint probability density function (pdf) of $\mathbf{y}^*$ which follows a multivariate normal distribution: $\mathbf{y} \sim MVN\{\mathbf{X}\boldsymbol{\beta}, V(\mathbf{t}, \boldsymbol{\phi})\}$, where $V(\mathbf{t}, \boldsymbol{\phi})$ is a block diagonal matrix with non-zero $(T_m \times T_m)$ blocks depending on some parameters $\boldsymbol{\phi}$. We will combine $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$ in one parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\phi})$ that defines the measurement process.

Also, let $\mathbf{h}_t = (y_1, ..., y_{t-1})$ denote an observed sequence of measurements observed up to time $t-1$, and y_t^* the value that would be observed at time t if the unit did not drop out.

Let $f_t^*(y_t \mid \mathbf{h}_t^*; \boldsymbol{\theta})$ denote the conditional univariate normal pdf of y_t^* given $(y_1, ..., y_{t-1}) = \mathbf{h}_t^*$ and $f_t(y_t \mid \mathbf{h}_t; \boldsymbol{\theta})$ the conditional pdf of y_t given $(y_1, ..., y_{t-1}) = \mathbf{h}_t$. It follows that

$$\Pr(y_t = 0 \mid \mathbf{h}_t, y_{t-1} = 0) = 1, \tag{3.7}$$

$$\Pr(y_t = 0 \mid \mathbf{h}_t, y_{t-1} \neq 0) = \int p_t(\mathbf{h}_t, y_t; \boldsymbol{\psi}) f_t^*(y_t \mid \mathbf{h}_t; \boldsymbol{\theta}) dy_t \tag{3.8}$$

and, for $y_t \neq 0$,

$$f_t(y_t \mid \mathbf{h}_t; \boldsymbol{\theta}, \boldsymbol{\psi}) = \{1 - p_t(\mathbf{h}_t; y_t; \boldsymbol{\psi})\} f_t^*(y_t \mid \mathbf{h}_t; \boldsymbol{\theta}). \tag{3.9}$$

The above equations determine the joint distribution of $\mathbf{y}$. For a complete sequence $\mathbf{y} = (y_1, ..., y_T)$, and supressing the dependence on the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$,

$$f(\mathbf{y}) = f_1^*(y_1) \prod_{t=2}^{T} f_t(y_t \mid \mathbf{h}_t)$$

$$= f^*(\mathbf{y}) \prod_{t=2}^{T} \{1 - p_t(\mathbf{h}_t, y_t)\}, \tag{3.10}$$

while for an incomplete sequence $\mathbf{y} = (y_1, ..., y_{k-1}, 0, ..., 0)$ with dropout at time k

$$f(\mathbf{y}) = f_1^*(y_1)\left\{\prod_{t=2}^{k-1} f_t(y_t \mid \mathbf{h}_t)\right\}\Pr(y_k = 0 \mid \mathbf{h}_k)$$

$$= f_{k-1}^*(\mathbf{y}^{k-1})\left\{\prod_{t=2}^{k-1}[1 - p_t(h_t, y_t)]\right\}\Pr(y_k = 0 \mid \mathbf{h}_k), \tag{3.11}$$

where $f_{k-1}^*(\mathbf{y}^{k-1})$ denotes the joint pdf of the first $k-1$ non-missing elements.

The loglikelihood function for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ based on the observed data $\{\mathbf{y}_m : m = 1, ..., n\}$ is given by Diggle and Kenward (1994) as

$$l(\boldsymbol{\theta}, \boldsymbol{\psi}) = l_1(\boldsymbol{\theta}) + l_2(\boldsymbol{\psi}) + l_3(\boldsymbol{\theta}, \boldsymbol{\psi}), \tag{3.12}$$

where

$$l_1(\boldsymbol{\theta}) = \sum_{m=1}^{n} \log f_m^*(\mathbf{y}_m),$$

$$l_2(\boldsymbol{\psi}) = \sum_{m=1}^{n}\sum_{t=2}^{k-1} \log\{1 - p_t(\mathbf{h}_{mt}, y_{mt})\},$$

and

$$l_3(\boldsymbol{\theta}, \boldsymbol{\psi}) = \sum_{m:k_m \le T} \log\Pr(k = k_m \mid \mathbf{y}_m).$$

In case of random dropout, l_3 reduces to depend only on $\boldsymbol{\psi}$, and therefore likelihoods for $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ can be maximised separately. It is recommended anyways to maximise l_1 and $(l_2 + l_3)$ separately assuming random dropout as a means of obtaining initial values for the full maximisation of $l(\boldsymbol{\theta}, \boldsymbol{\psi})$. If random dropout holds, only l_1 is required to make valid inference about the marginal measurement process. However, the full likelihood is still needed for inferences about the conditional process $Y \mid non-dropout$.

The model outlined above, introduced by Diggle and Kenward (1994), for non-random dropout is used as basis for many articles to follow. Molenberghs et al. (1997) use the same framework to model dropout probabilities when the variable of interest is ordinal. This requires a different approach to modelling the response and to the resulting likelihood calculations. Jansen et al. (2006) also study non-Gaussian outcomes such as binary, categorical or count data. They consider both generalised linear mixed models, for which the parameters can be estimated using maximum likelihood, and marginal models estimated through generalised estimating equations, which is a non-likelihood method and hence requires modification to be valid under MAR. Molenberghs and Verbeke (2001) provide a review on linear mixed models for continuous longitudinal data focusing on the problem of missing data within both selection models and *pattern-mixture models*.

Sensitivity Analysis Within Selection Models

When fitting a nonrandom dropout model, both the impact of the assumed distributional form and the impact one or a few influential subjects may have on the model parameters, should be considered.

In their work, Verbeke et al. (2001) adopt the same model proposed by Diggle and Kenward (1994) with the following perturbed version of the dropout model

$$\text{logit}\left[g_t(\mathbf{h}_{mt}, y_{mt})\right] = \text{logit}\left[\Pr\left(k_m = t \mid k_m \geq t,\, \mathbf{y}_m\right)\right] = \tilde{\mathbf{h}}_{mt}\boldsymbol{\Psi} + \psi_{1m} y_{mt}.$$

The difference in this model is that the ψ_{1m}s are not viewed as parameters but rather as local, individual-specific perturbations around a null model. The null model is taken to be the MAR model corresponding to setting $\psi_1 = 0$. Verbeke et al. (2001) state that " When small perturbations in a specific ψ_{1m} lead to relatively large differences in the model parameters, it suggests that the subject is likely to drive the conclusions...Such an observation is important also for our approach because then the impact (e.g., from influential subjects) on dropout model parameters extends to all functions that include these dropout parameters...Therefore, influence on measurement model parameters can arise not only from incomplete observations but also from complete ones". A similar model is studied in Steen et al. (2001) for incomplete longitudinal multivariate ordinal data.

3.2.2.2 Pattern-Mixture Models

Pattern-mixture models, introduced by Little (1993), represent an alternative to selection models. They factorise the joint density in the opposite way, that is as product of the marginal density of the dropout mechanism, and the conditional density of the measurement process given the dropout. In other words, the measurement process is defined over different dropout patterns. The density of the full data $(\mathbf{y}_m, k_m)$ in the pattern-mixture model context can be written in the form

$$f(\mathbf{y}_m, k_m \mid \mathbf{X}_m, \mathbf{W}_m, \theta, \psi) = f(\mathbf{y}_m \mid k_m, \mathbf{X}_m, \mathbf{W}_m, \theta) f(k_m \mid \mathbf{X}_m, \psi),$$

where the first factor is the density of the measurement process conditional on dropout, and the second factor is the marginal density of the missingness mechanism. For example, assuming that the response will have a multivariate normal distribution, with the possibility of dropout at any time point after the first one, it follows that the response will have a different distribution at each time of dropout k_m:

$$\mathbf{y}_m \mid k_m \sim N(\mu(k_m), V(k_m)),$$

where $k = 2, ..., T$. Let $\pi_k = f(k_m \mid \psi)$ denote the marginal density of the missingness mechanism. The marginal distribution of the response is then a mixture of normals with mean

$$\mu = \sum_{k=1}^{T} \pi_k \mu(k).$$

In general, pattern-mixture models depend on restrictions defined in terms of conditional distributions of the response given the dropout $\mathbf{y}_m \mid k_m$.

3.2.2.3 Shared-parameter Models

The third general approach for modelling dropout are *shared-parameter models*, in which both the measurement process and dropout are influenced by a latent variable or random effect (e.g. Wu and Carroll (1988); Wu and Bailey (1989); Henderson et al. (2000)). A shared parameter model is thus a selection model which is also conditional on a latent variable. This specification allows the dropout to be non-ignorable given the observed data only, but ignorable given also the latent variables. Wu and Carroll (1988) proposed a model where a random effect is shared between a mixed effects linear model for a normal repeated measure, and a discrete-time survival model for the missingness mechanism, thus allowing for informative dropout. Their model is extended by Wu and Bailey (1989) who proposed conditioning on the time to censoring and using censoring time as a covariate in the random effects model. Ten Have et al. (1998) also proposed a shared-parameter model with a logistic link for a longitudinal binary outcome subject to informative dropout. Roy (2003) introduced a shared-parameter model in which the dependence between the measurement process $\mathbf{y}_m$ and time of dropout k_m is due to a shared latent variable η that is assumed to be discrete, so that the marginal distribution of the measurement is a mixture over the dropout classes of the latent variable

$$f(\mathbf{y}_m, k_m) = \sum_{\eta} f(\mathbf{y}_m \mid \eta) f(\eta \mid k_m) f(k_m), \tag{3.13}$$

rather than the dropout times themselves. This latent dropout class model is used for univariate longitudinal data, and is estimated by maximum likelihood.

Dantan et al. (2008) compare pattern-mixture models and latent class models in dealing with informative dropout.

The nature of the problem studied in this thesis makes shared-parameter models an appealing option to adopt, where instead of a shared parameter, a latent variable is employed to affect both the observed variables and the missingness mechanism. The use of a latent variable is unavoidable in our case since interest lies in unobserved phenomena which are measured by observed items through latent variable models. Allowing the missingness mechanism to be affected by the latent variables, makes the setup of the model fall within a shared-parameter model. However, a selection model which is also conditional on a latent variable is explored in Chapter 7 where latent variables at time t are allowed to affect non-response at the same time t. More on shared-parameter models will be introduced later, in relevant sections of the thesis.

Chapter 4

A SEM for Multivariate Ordinal Longitudinal Data Subject to Dropout[1]

The models developed in this chapter are latent variable models in which a continuous latent variable is used at each time point to explain the associations among multiple ordinal observed response items. The models are fitted in a SEM framework where the underlying variable approach is adopted. Item-specific random effects are included to account for repetition of items over time. For modelling dropout, we introduce dropout indicators which are modelled with a hazard function. Different structures among the latent variables and the dropout mechanism are explored in two different model specifications which allow attitudes and co-

[1] A paper based on parts of Chapters 4 and 5 has been published in *Structural Equation Modeling* journal under the title : Hafez, M. S., Moustaki, I. and Kuha, J. (2015). Analysis of Multivariate Longitudinal Data Subject to Nonrandom Dropout. *Structural Equation Modeling: A Multidisciplinary Journal,* 22(2), 193-201.

variates to affect both the latent variables and dropout indicators.

A latent variable model is first specified for the complete case multivariate data, disregarding dropout. This model is formed of two parts: the measurement part in which the observed variables are explained by a latent variable at each time point, and the structural part which defines relationships among the latent variables over time. Having specified this model for the complete data, we then define models for the dropout mechanism with a hazard function. Finally, the link between attitudes and dropout is specified.

4.1 Modelling The Observed Indicators: The Measurement Model

We will consider ordinal items as they are among the most common type of items used for measuring attitudes in social surveys. Suppressing the index m for a subject (e.g. survey respondent) for convenience, let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{pt})$ be $(p \times 1)$ vectors of observed ordinal variables for a single subject at times $t = 1, 2, ..., T$. Within a SEM framework, it is assumed that each y_{it} is a manifestation of an underlying unobserved continuous variable y^*_{it} as outlined in Section 2.1.1. Let c_{it} denote the number of categories for y_{it}, the ith variable $(i = 1, 2, ..., p)$, at time t. For an ordinal variable y_{it} with c_{it} categories, its relationship with y^*_{it} is as defined in equation (2.1)

$$y_{it} = s \Leftrightarrow \tau^{(i)}_{s-1} < y^*_{it} \leq \tau^{(i)}_{s}, \quad s = 1, \cdots, c_{it}, \tag{4.1}$$

where $\tau_0^{(i)} = -\infty$, $\tau_1^{(i)} < \tau_2^{(i)} < \ldots < \tau_{c_{it-1}}^{(i)}$, and $\tau_{c_{it}}^{(i)} = \infty$ are thresholds, to be estimated.

The items $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{pt})$ at each time t are regarded as measures of one or more continuous attitudinal time-dependent latent variables z_{a_t}, that are assumed to be normally distributed as presented by equation (2.2). For simplicity, the model below is presented assuming that the items are unidimensional (i.e. one latent variable is sufficient to explain dependencies among items at a given time point), but it can be extended to accommodate more latent variables. To account for repitition of the same items at each time point, two possible options are available. The first option is to allow for correlated errors in the measurement model for z_{a_t}:

$$y_{it}^* = \lambda_i z_{a_t} + \varepsilon_{it}; \qquad i = 1, ..., p\,;\; t = 1, ..., T,$$

where λ_i is the loading of the latent variable z_{a_t} on y_{it}^* and ε_{it} is a normally distributed random error that is correlated with errors $\varepsilon_{it'}$ of the same item across time.

Equivalently, we introduce an item-specific random effect u_i instead of correlating errors

$$y_{it}^* = \lambda_i z_{a_t} + u_i + \varepsilon_{it}; \qquad i = 1, ..., p\,;\; t = 1, ..., T, \tag{4.2}$$

where λ_i is the loading of the latent variable z_{a_t} on y_{it}^*, and ε_{it} is a normally distributed random error that is uncorrelated with other errors. In this model, associations among different items at the same time ($y_{it}^*, y_{i't}^*$ for $i \neq i'$) are explained by the dependence on the common latent variable z_{a_t}, while associations between the values of the same item measured at different time points ($y_{it}^*, y_{it'}^*$ for $t \neq t'$)

are explained both by the covariance between corresponding attitudinal latent variables $(z_{a_t}, z_{a_{t'}})$ and the item-specific random effect u_i. It is assumed that the random effects u_i are independently normally distributed as $u_i \sim N(0, \sigma^2_{u_i})$ for $i = 1, ..., p$, and that ε_{it} are independent and normally distributed as $\varepsilon_{it} \sim N(0, \omega^2_{it})$ for $i = 1, ..., p$ and $t = 1, ..., T$, where $\omega^2_{it} = 1 - (\lambda^2_i \operatorname{var}(z_{a_t}) + \sigma^2_{u_i})$ since each y^*_{it} is assumed to have a standard normal distribution. The error terms ε_{it} and random effects u_i are assumed to be uncorrelated.

The aim of a longitudinal study is to monitor changes that occur between occasions, and to attribute these changes to background characteristics. Jöreskog (2005) explain that in order to estimate changes in means and variances of latent variables, they should be on the same scale over time. In case of continuous items, this is simply done by choosing the same reference variable at different occasions and assuming the mean of the latent variable is zero at the first occasion, thus monitoring how it changes on subsequent waves. However, this is not sufficient in case of ordinal items as they do not have metric scales. The underlying variables are used instead. These can be put on the same scale by assuming equal thresholds for the underlying variables of the same items over time.

We have imposed the assumption of *invariance of measurement* across time for each item $i = 1, \ldots, p$, by constraining the thresholds $\tau^{(i)}_s$ in equation (4.1) (for each $s = 1, \ldots, c_{it}$) and the loading λ_i in equation (4.2) for each $i = 1, \ldots, p$ to be the same at all time points $t = 1, \ldots, T$. In order to set the scale for the time-dependent attitude latent variables, the loading λ_1 on the first underlying variable y^*_{1t} is set to 1 at all occasions. The advantages of these constraints are both technical and conceptual. On the technical side, it yields a more parsimonious model and avoids some possible identification problems that may arise with increasing the number of time points (Bijleveld et al. (1998)). The conceptual advantage is clearer interpretation of the model results. If the loadings and thresholds are not constrained to be time-invariant, we cannot guarantee that the latent variable has the same scale or interpretation at each time point.

4.2 Modelling The Latent Variables: The Structural Model

The structural part of the model addresses the question: how should the attitudinal latent variables be linked in order to capture the longitudinal nature of the data? Throughout, we will assume that the possible measurement occasions $t = 1, \ldots, T$ are the same for every subject, and evenly spaced in time. We then specify that the $(T \times 1)$ vector of attitude latent variables $\mathbf{z}_a = (z_{a_1}, ..., z_{a_T})'$ follows a multivariate normal distribution $\mathbf{z}_a \sim MVN_{(T)}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ where $\boldsymbol{\mu}$ is a vector of means and $\boldsymbol{\Gamma}$ a covariance matrix with diagonal elements σ_t^2 representing the variances of the latent variables, and off-diagonal elements $\sigma_{tt'}$ their covariances such that $\sigma_{tt'}$ is the covariance between z_{a_t} and $z_{a_{t'}}$. The values of these parameters may be unconstrained, or depend further on the model specification, as defined below. For example, it is logical to expect that attitudes are more strongly correlated when they are measured at closer time points, in which case $\sigma_{tt'}$ should be higher when t and t' are close to each other. For identification, the mean of z_{a_1} is set to 0.

A specification for the structural part which takes the time ordering explicitly into account is the first-order autoregressive [AR(1)] structure, presented in equation (3.4) for modelling univariate longitudinal data. A subject-specific random effect analogous to that included in equation (3.5) is not considered here as it would act as yet another latent variable, resulting in a more complex model with multiple levels of latent variables. The model thus assumes that all the within-subject dependence is due to the lagged attitude. This specification has been used by Dunson (2003) and Cagnone et al. (2009) to model the dependence of latent variables over time within a similar context, where $z_{a_1} \sim N(0, \sigma_1^2)$ and

$$z_{a_t} = a_t + \phi z_{a_{t-1}} + \delta_t, \quad t = 2, ..., T, \tag{4.3}$$

where a_t is an intercept, ϕ a time-constant regression coefficient representing the dependence of the attitude at time t on that at the previous occasion $t-1$ justified by the equally spaced time intervals, and $\delta_t \sim N(0, \upsilon_t^2)$ is a random error which is uncorrelated with $z_{a_1}, ..., z_{a_{t-1}}$. This formulation explicitly captures the time ordering in the data, by presenting the model as a sequence of conditional distributions rather than a joint distribution with an unstructured correlation matrix $\mathbf{\Gamma}$. It expresses the dynamic nature of the latent attitude variable and accounts for the serial correlation in it in a form where the latent variable at time point 3, say, is only related to that measured at time 1 via the latent variable at time 2. Another alternative specification would be a random effects model in which a random intercept and possibly a random slope, where time is a covariate, affect the time-dependent latent variables as in a standard growth mixture model for observed repeated measures; for example, see Muthén and Masyn (2005) and Muthén et al. (2011). However, this type of models is not considered here.

More generally, we may also be interested in studying the associations between the attitudinal latent variables and observed covariates (explanatory variables), such as demographic and socioeconomic characteristics of survey respondents. Let $\mathbf{x}_t$ denote a vector of such covariates, noting that some components of $\mathbf{x}_t$ (e.g. sex and race) may be constant over time while others (e.g. marital status and health condition) may be time-varying. In this case, the AR(1) structure in equation (4.3) can be extended to include covariates, as

$$z_{a_t} = a_t + \phi z_{a_{t-1}} + \boldsymbol{\beta}_t' \mathbf{x}_t + \delta_t, \quad t = 2, ..., T, \tag{4.4}$$

where $\boldsymbol{\beta}_t$ is a vector of regression coefficients for $\mathbf{x}_t$.

4.3 Modelling The Dropout

Dropout is a form of missing data in which a respondent in a longitudinal study fails to respond at a given occasion and never comes back to the study. Dropout is typically the most common form of missingness in longitudinal studies. In this chapter, we will focus solely on dropout, and assume that there is no intermittent missingness in the data. We also assume that, at each time point, variables for a respondent are either fully observed or totally missing, i.e. that there is no item non-response.

Our approach to handling dropout in multivariate longitudinal data draws on ideas of shared-parameter models for univariate longitudinal data as in the model presented by Roy (2003) where both the measurement process and dropout mechanism are modelled conditional on a discrete latent variable (see equation (3.13)), and on previous work on modelling non-ignorable item non-response in multivariate cross-sectional data. Early examples of the latter are Knott et al. (1990) and

O'Muircheartaigh and Moustaki (1999), who present a latent variable approach that allows missing values to be included in the analysis and information about latent attitudes to be inferred from non-response. They propose two latent dimensions, one to summarise the attitude and the other to summarise response propensity. For each observed variable, an indicator variable for responding is created, taking the value 1 if the individual responds and 0 if he or she does not respond. The probability of responding depends both on an individual's position on the attitudinal latent variable and the response propensity latent variable. The attitude items are explained only by an individual's position on the attitudinal latent variable given that this individual has responded. Holman and Glas (2005) use reformulations of the models of O'Muircheartaigh and Moustaki (1999) to assess the extent to which the missing data can be ignored. Within the same framework, Moustaki and Knott (2000b) present a latent variable model for binary and nominal observed items which includes covariate effects on attitudinal and response propensity items. In our study, we extend this approach to the longitudinal case taking dropout into account.

The dropout model outlined below has the form of a discrete-time hazard model, a widely used representation of event histories in social sciences. See for example Allison (1982) and Muthén and Masyn (2005). Let us define the probability that a respondent drops out at time t, given that they have remained in the study up to and including time $t-1$, by the hazard function $h_t = \Pr(k = t \mid k \geqslant t)$, $t = 2, ..., T$, where k is a discrete random variable that indicates the time of dropout. We also define a set of dropout indicators d_t, $t = 1, \ldots, T$, such that $d_t = 0$ when $\mathbf{y}_t$ is observed and $d_t = 1$ if a respondent drops out at time t (Muthén and Masyn (2005)). After the time of dropout, d_t itself is regarded as missing and can be set to an arbitrary value such as -1. We treat the observations at the first occasion as complete data, so that $d_1 = 0$ for all respondents, and define $\mathbf{d} = (d_2, ..., d_T)$. For an example with three waves ($T = 3$), an individual will have $\mathbf{d} = (0, 0)$ if they show up on all three occasions, $\mathbf{d} = (0, 1)$ if they drop out on the third occasion, and $\mathbf{d} = (1, -1)$ if they drop out on the second occasion. The data structure presented here is of "wide" form, which is necessary for estimating the model within a SEM framework. In "long" form, when a person experiences the event, they are removed from the risk of experiencing it at future waves thus no missing values are recorded after a dropout, and subjects will have dropout vectors $\mathbf{d}$ of different lengths according to their time of dropout. For example, a subject who drops out at the second occasion would have $\mathbf{d} = (1)$. With this notation, the hazard function can also be expressed as

$$h_t = \Pr(k = t \mid k \geqslant t) = \Pr(d_t = 1), \quad t = 2, ..., T.$$

In the more general case of intermittent missingness we could define binary missingness indicators such that the indicator d_t at time t has the value 0 if $\mathbf{y}_t$ is observed and 1 if it is missing. In that case, the missingness indicators may be assumed to measure a single latent variable z_{d_t} which summarises an individual's *response propensity*. Such a propensity may also be thought to exist in our case, where only dropout is considered. However, since the dropout indicators are created from a single variable (time of dropout), this latent propensity cannot be separately identified. Nevertheless, we will still employ such z_{d_t} as a convenient computational and presentational device, but with a formulation where they have a conditional variance of 0, given the attitude latent variables z_{a_t} and (possibly) covariates $\mathbf{x}_t$ (Muthén and Masyn (2005)). This means that z_{d_t} will be deterministic functions of z_{a_t} and $\mathbf{x}_t$, which will then affect the dropout indicators via z_{d_t}.

In the same way as for the observed items $\mathbf{y}_t$ in equation (4.1), we assume a set of continuous variables $\mathbf{d}^* = (d_2^*, \ldots, d_T^*)$ to underlie the set of dropout indicators $\mathbf{d} = (d_2, \ldots, d_T)$. Each of the d_t^* is assumed to have a standard normal distribution and to be modelled as

$$d_t^* = \lambda_{d_t} z_{d_t} + \varepsilon_{d_t}, \quad t = 2, ..., T, \tag{4.5}$$

where λ_{d_t} is the loading of z_{d_t} on the dropout variable at time t, and $\varepsilon_{d_t} \sim N(0, \omega_{d_t}^2)$ is a random error, with $\omega_{d_t}^2 = 1 - \lambda_{d_t}^2 \mathrm{var}(z_{d_t})$. Since the missingness indicators are all binary, only one threshold τ_{dt} is estimated for each variable d_t^*.

We will consider two special cases of this model. In the first, we take $z_{d_t} = z_{a_{t-1}}$ for $t = 2, \ldots, T$. Model (4.5) then becomes

$$d_t^* = \lambda_{d_t} z_{a_{t-1}} + \varepsilon_{d_t}, \quad t = 2, ..., T. \tag{4.6}$$

In this formulation, the probability of dropping out at a given time point depends only on the value of the latent attitude variable at the immediately preceding time point. The dropout indicators are thus in effect treated just like further "measures" of the attitude. Because the loadings λ_{d_t} can vary with t, the effect of attitude on dropout may depend on time.

In our second dropout model we define $z_{d_t} = z_d$ instead as a time-constant quantity which depends on the attitude only through its value z_{a_1} at the first time point. In this formulation we also allow for the possibility that the response propensity depends also on covariates $\mathbf{x}_1$ measured at the first time point. We thus define $z_d = \gamma z_{a1} + \boldsymbol{\beta}_d'\mathbf{x}_1$, where γ is a regression coefficient representing the dependence of the dropout "latent variable" z_d on the attitude latent variable z_{a_1} at the first time point, and $\boldsymbol{\beta}_d$ is a vector of the regression coefficients of covariates $\mathbf{x}_1$ similarly. Furthermore, in equation (4.5) we take $\lambda_{d_t} = 1$ for all t, to obtain

$$d_t^* = \gamma z_{a1} + \boldsymbol{\beta}_d'\mathbf{x}_1 + \varepsilon_{d_t}, \quad t = 2, ..., T. \tag{4.7}$$

Here the time-constant dropout variable z_d is regressed solely on z_{a_1} in order to avoid a multicollinearity problem that is very likely to occur if z_d was regressed on other attitude latent variables as well, due to the high correlation expected between the latent variable across different time points. Attitude at the first time is particularly chosen because it is the only occasion with complete data, and because it avoids a specification where dropout at time t would depend on attitude at future time points. Following the same argument, dropout is also regressed only on covariates measured at the first time.

When dropout is non-ignorable, a model for it needs to be incorporated in the estimation in order to obtain valid estimates for the parameters of interest in the structural and measurement models. For multivariate longitudinal data, unlike in many other situations, this can in fact be done without further unverifiable assumptions. In other words, combining the elements described above it is possible to fit models which combine multivariate longitudinal models for the latent attitude variables of interest with models for possibly non-ignorable dropout. In the next section, we discuss such joint models in more detail.

4.4 Joint Models for Attitudes, Measurements and Dropout

Having set the general layout of the model, we now look into two particular specifications of it. In both of them, the measurement model of the observed items y_{it} is defined by equations (4.1) and (4.2), and the corresponding assumptions. Differences lie in the definitions of the structural and dropout parts of the model, and the relationship between them. In both specifications, a latent variable is shared in the modelling of both the measurement process and the dropout mechanism, similar to the idea of the shared-parameter model presented in Roy (2003). A pattern-mixture model would assume that subjects with the same dropout time share a common response distribution, thus a response distribution is a mixture over response patterns. This assumption is considered too strong and may be unrealistic especially for studies with a large number of follow-up times where subjects drop out for a variety of reasons (Roy (2003)). A pattern-mixture model is thus not considered here. However, a shared-parameter that also classifies under selection models is considered later on in Chapter 7, where both dropout and intermittent missingness are incorporated. A special case of that model would only consider dropout, with dropout at time t depending on unobserved attitudes measured at time t, similar to selection models for the univariate longitudinal case presented in equation (3.6), with latent variables used instead of actual observed items.

The first model specification allows for the simple choice of a free mean structure and correlation matrix for the attitudinal latent variables $\mathbf{z}_a = (z_{a_1}, ..., z_{a_T})'$ at different time points. In other words, we assume a multivariate normal distribution $\mathbf{z}_a \sim MVN_{(T)}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$ with $\boldsymbol{\mu}$ and $\boldsymbol{\Gamma}$ unconstrained. For incorporating dropout, we assume model (4.6) where the attitudinal latent variable $z_{a_{t-1}}$ at each previous time point is allowed to directly affect the dropout at the next one. The employment of such a lagged effect can be useful for example when latent variables are measuring health condition and time points are not too far apart, so that a deteriorating health condition at time $t-1$ may affect the probability of dropping out at the next occasion t due to bad health. Non-zero dropout loadings λ_{dt} for $t = 2, \ldots, T$, will reflect dependence of dropout on unobserved attitudes. Figure 4.1 gives an illustration of this model by a path diagram for an example with three time points.

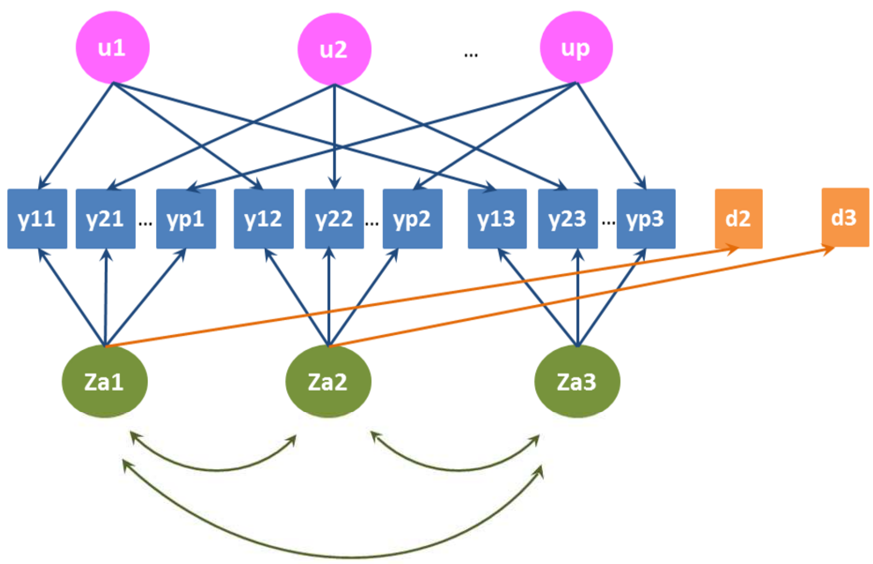


Figure 4.1: Path diagram for the first model specification (Model 1).

The second model specification assumes a first-order autoregressive structure among the latent variables $\mathbf{z}_a$, as presented in equation (4.3), instead of freely correlating them. With the attitude at the first time point also assumed to be normally distributed as $z_{a_1} \sim N(0, \sigma_1^2)$, this model too implies that $\mathbf{z}_a$ follows a multivariate normal distribution, but now with the covariance matrix $\mathbf{\Gamma}$ being a constrained function of the parameters ϕ, σ_1^2, and error variances $\upsilon_2^2, \ldots, \upsilon_T^2$, and the mean vector $\boldsymbol{\mu}$ unconstrained and depending on the parameters $a_2, \ldots, a_T$ and ϕ. For this model specification we also examine the extension of the structural model by including in it covariates $\mathbf{x}_t$ with coefficients $\boldsymbol{\beta}_t$, as shown in equation (4.4).

For the dropout model in the second model specification, we assume a model where the underlying dropout variables are modelled as a function of the dropout "latent variable" z_d which in turn is determined by the attitude latent variable z_{a_1} and covariates $\mathbf{x}_1$ at the first time point, thus resulting in the dropout model (4.7). Such a model would be reasonable to adopt in a case where the latent variable (attitude) being measured does not change much over time (example: political attitudes). In that case attitude measured at first occasion is used as it is based on complete data. Figure 4.2 gives an illustration of the joint model for the second model specification, for an example with three time points.

In the second specification, parameters of the dropout model include regression coefficients γ and $\boldsymbol{\beta}_d$, with non-zero γ indicating the dependence of dropout on attitude measured at first wave. These parameters are to be estimated, along with the parameters of the measurement model (including variances of the random effects u_i) and the structural model.

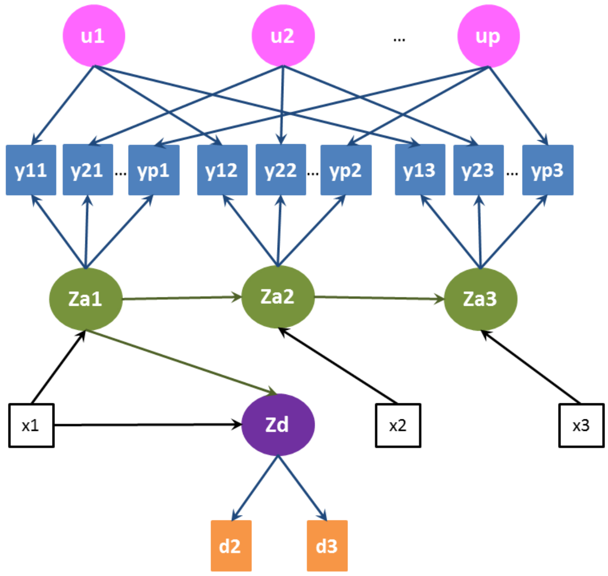


Figure 4.2: Path diagram for the second model specification (Model 2), with covariates.

4.5 Estimation

Within a SEM framework, parameter estimation is done in three steps as outlined in Section 2.1.3, where thresholds are estimated in the first step from the univariate marginal distributions. Since measurement invariance is assumed for each item across time, the thresholds $\tau_s^{(i)}$ (for each $s = 1, \ldots, c_{it}$) are estimated from the univariate marginal data of the same item at all time points $y_{i1}, \ldots, y_{iT}$. Thus the estimated thresholds for an item y_{it}, $t = 1, \ldots, T$ are

$$\hat{\tau}_s^{(i)} = \Phi^{-1}(p_1^{(i)} + p_2^{(i)} + \ldots + p_s^{(i)}), \qquad s = 1, \ldots, c_i - 1,$$

where p_s represents the percentage of responses in category s for items $y_{i1}, \ldots, y_{iT}$. Thresholds τ_{dt} for dropout indicators are obtained similarly.

The polychoric correlations are estimated in the second step from the bivariate distributions for given thresholds, by maximising the loglikelihood for each pair of items. These include correlations between items at the same time point, and across time points. Once the estimated polychoric correlation matrix is obtained, it is treated as an observed correlation/covariance matrix S. The theoretical covariance matrix is given by $\mathbf{\Sigma} = \mathbf{\Lambda\Gamma\Lambda'} + \mathbf{\Omega_u}$, where $\mathbf{\Gamma}$ is the covariance matrix of attitudinal latent variables and $\mathbf{\Omega_u}$ is a diagonal covariance matrix of item-specific random effects. Elements of $\mathbf{\Sigma}$ include covariances between different items within the same time point

$$\mathrm{Cov}(y^*_{it}, y^*_{i't}) = \lambda_i \lambda_{i'} \mathrm{Var}(z_{a_t}),$$

covariances between the same items at different time points

$$\mathrm{Cov}(y^*_{it}, y^*_{it'}) = \lambda_i^2 \mathrm{Cov}(z_{a_t}, z_{a_{t'}}) + \sigma^2_{u_i},$$

and covariances between different items at different time points

$$\mathrm{Cov}(y^*_{it}, y^*_{i't'}) = \lambda_i \lambda_{i'} \mathrm{Cov}(z_{a_t}, z_{a_{t'}}).$$

Variances and covariances of attitudinal latent variables will have different forms according to which model specification is being fitted. The covariance matrix $\mathbf{\Sigma}$ also includes covariances between dropout indicators, which are given by

$$\mathrm{Cov}(d^*_t, d^*_{t'}) = \lambda_{dt} \lambda_{dt'} \mathrm{Cov}(z_{a_{t-1}}, z_{a_{t'-1}}),$$

in case the first dropout model is used, while they are of the form

$$\text{Cov}(d_t^*, d_{t'}^*) = \gamma^2 \text{Var}(z_{a_1}),$$

for the second specification. For the first model specification, covariance between an item y_{it} and a dropout indicator d_t at the same time point t is

$$\text{Cov}(y_{it}^*, d_t^*) = \lambda_i \lambda_{dt} \text{Cov}(z_{a_t}, z_{a_{t-1}}),$$

while covariance between an item y_{it} and a dropout indicator at a different time point $d_{t'}$ is

$$\text{Cov}(y_{it}^*, d_{t'}^*) = \lambda_i \lambda_{dt'} \text{Cov}(z_{a_t}, z_{a_{t'-1}}).$$

For the second model specification, covariance between an item y_{it} and a dropout indicator d_t at the same time point t or at a different time point $d_{t'}$ is the same, and is given by

$$\text{Cov}(y_{it}^*, d_{t\prime}^*) = \lambda_i \gamma \text{Cov}(z_{a_t}, z_{a_1}).$$

The latent variable model can then be fitted to the estimated polychoric correlation matrix using unweighted least squares (ULS), diagonally weighted least squares (DWLS), or weighted least squares (WLS) by minimising some function of the difference $S - \Sigma$. In WLS, the weight matrix is an estimate of the inverse of the asymptotic covariance matrix of polychoric correlations, while DWLS involves only the diagonal elements of that weight matrix. Recent studies confirm (Forero et al. (2009); Yang-Wallentin et al. (2010)) that the WLS estimator converges very slowly to its asymptotic properties and therefore does not perform well in small sample sizes. DWLS and ULS are preferable to WLS and they seem to perform similarly well in finite samples. However, in order to compute correct standard errors and goodness-of-fit tests, the full weight matrix is needed. In our application, DWLS is used for estimation and WLS for obtaining the standard errors and test statistics.

A full information maximum likelihood estimation approach would assume a multivariate normal distribution for all the underlying variables $y_{1t}^*, y_{2t}^*, \ldots, y_{pt}^*$ at different time points $t = 1, \ldots, T$. Estimation involves maximising the $(p \times T)$-dimensional loglikelihood over all response patterns present in the data. This requires the evaluation of a $(p \times T)$-dimensional integral for each response pattern in the sample, which is computationally infeasible as the number of observed variables increases (Jöreskog and Moustaki (2001)). That is why the above approach, based on bivariate normality assumption, is adopted.

Chapter 5

Application and Sensitivity Analysis

5.1 Attitudes Towards Women's Work: BHPS

In this chapter, we apply the models proposed in Chapter 4 to study the evolution of people's attitudes towards women's work, using data from the British Household Panel Survey (BHPS). The BHPS (Taylor et al. (2010)) is a multi-purpose study that started in 1991 aiming to further understand the social and economic change at the individual and household level in Britain and the UK. It followed the same sample of individuals, drawn from different parts of the United Kingdom, over a period of 18 years, thus providing a rich research resource for a wide range of social science disciplines. A section of the BHPS includes questions on opinions about women's work and family life. On carrying out an exploratory factor analysis over six items, the following three were found to measure a single factor:

- A woman and her family would all be happier if she goes out to work [Family]
- Both the husband and wife should contribute to the household income [Con-

tribution]

- Having a full-time job is the best way for a woman to be an independent person [Independent]

We treat these three items as measures of a respondent's attitude towards women's work. Previous studies have used similar items to measure such an attitude (See for example Donnelly et al. (2015)). For each item, the response options are *Strongly agree*, *Agree*, *Neither agree nor disagree*, *Disagree*, and *Strongly disagree*. The wording of the items implies that the attitudinal latent variable will be defined such that the higher an individual scores on the latent variable, the more conservative are his or her views towards women's work. Five waves of the survey (1993, 95, 97, 99, 2001) are considered here. Dropout occurs in all waves but the first one. A frequency distribution of the three items measured at the first wave (1993) is given in Table 5.1. It is noted that most responses are concentrated in the three middle categories (*Agree*, *Neither agree nor disagree*, and *Disagree*), with fewer responses in the two extreme categories (*Strongly agree* and *Strongly disagree*).

Table 5.1: Frequency distribution for items (Family, Contribution and Independent) measured at first wave

	Strongly agree	Agree	Neither	Disagree	Strongly Disgaree
Family	98	929	3059	1560	173
Contribution	833	2225	1979	727	55
Independent	437	1998	1920	1321	143

The sample size of individuals who gave complete answers in the first wave considered here (year 1993) is 5819. In the second wave, with 10% dropout the sample size decreases to 5227, and in the third wave, a further 6% dropout reduces it to 4901. Dropout continues at each wave until the sample size becomes 4296 at the last wave considered here (year 2001), constituting approximately 74% of the original sample size. The analysis aims to explore how much each of the three items contributes to measuring this attitude and how the attitude evolves over the nine-year period, accounting for dropout by incorporating the dropout mechanism in the model.

In the next section, results from the two different model specifications outlined in Chapter 4 are compared. Covariates are introduced and their effects studied under the second model specification.

5.2 Data Analysis

5.2.1 Fitting Two Model Specifications

The models being studied are the ones introduced in Chapter 4, with items y_{it}, $i = 1, 2, 3$, and the dropout indicators d_t used to give information on one attitudinal latent variable z_{a_t} at waves $t = 1, \ldots, 5$ (with $d_1 = 0$ for all, as the observations are regarded as complete at the first wave). The latent variable captures attitudes towards women's work, with higher values of it indicating more conservative attitudes. Data analysis is implemented in Mplus (Muthén and Muthén (2011)). Mplus codes for the two model specifications are given in Appendix A and B, respectively.

We first carried out two preliminary analyses, which allowed us to conclude that two assumptions introduced in Chapter 4 are satisfied in these data. First, we considered the assumption of measurement invariance, which states that thresholds

$\tau_s^{(i)}$ of an underlying variable y_{it}^* given in (4.1), and loadings λ_i of the measurement model (4.2) are the same at all time points t. For the model to be identified, Millsap and Yun-Tein (2004) prove that at least one threshold $\tau_s^{(i)}$ for each underlying variable y_{it}^* and two thresholds for each reference underlying variable are required to be the same at all time points t. However, Jöreskog (2005) explain that in order to estimate changes in means and variances of latent variables over time, they should be put on the same scale by assuming all thresholds for the underlying variables of the same items to be equal over time. A chi-square difference test[1] was thus carried out testing whether loadings too could be set equal over time, and this constraint was not rejected (p-value = 0.05) against a model which allowed for equal thresholds but free loadings across time. The reported p-value is just on the border of rejection of the restricted model at 5% level of significance, but we choose to continue with a measurement invariant model that provides both parsimony and clearer interpretation of results. Next, for the second model specification we examined the assumption that the dropout latent variable z_d in equation (4.5) has its loadings λ_{d_t} set to 1 at all of $t = 2, \ldots, 5$, which for this model specification also implies that the attitudinal latent variable measured at first wave z_{a_1}, will have the same effect γ in equation (4.7) on dropout indicators at all time points. The model with this constraint was also not rejected (p-value = 0.5941) against the unrestricted model where those loadings were allowed to vary freely across time points.

Table 5.2 gives parameter estimates for the two model specifications when

[1] To obtain a correct chi-square difference test for two nested models using WLS, a null model is compared to a less restrictive alternative model in which the null model is nested. The less restrictive model is first estimated followed by the restricted model, and the chi-square difference test is computed using derivatives from the analyses of both models (Muthén and Muthén (2011)).

covariates are not yet considered. The attitude towards women's work loads very similarly on all three items, suggesting that the items contribute almost equally to measuring the attitude. The estimated thresholds for the dropout model are given in the second part of Table 5.2.

Table 5.2: Parameter estimates for Models 1 and 2, for modelling attitudes towards women's work from five waves of the British Household Panel Survey

		Model 1			Model 2		
	Measurement model						
		Est.	Sig.	S.E.	Est.	Sig.	S.E.
Family	λ_1	1			1		
Contribution	λ_2	1.115	***	(0.023)	1.115	***	(0.023)
Independent	λ_3	1.151	***	(0.025)	1.149	***	(0.025)
	Dropout model						
d_2^*	τ_{d_2}	1.272	***	(0.022)	1.272	***	(0.022)
d_3^*	τ_{d_3}	1.534	***	(0.027)	1.535	***	(0.027)
d_4^*	τ_{d_4}	1.533	***	(0.028)	1.536	***	(0.028)
d_5^*	τ_{d_5}	1.506	***	(0.029)	1.512	***	(0.029)
z_{a_1} on d_2^*	λ_{d_2}	-0.014	*	(0.008)			
z_{a_2} on d_3^*	λ_{d_3}	-0.019	*	(0.011)			
z_{a_3} on d_4^*	λ_{d_4}	-0.044	**	(0.018)			
z_{a_4} on d_5^*	λ_{d_5}	-0.056	*	(0.029)			
Dropout parameter	γ				-0.036	***	(0.009)
	Random effects						
Variances							
u_1	$\sigma^2_{u_1}$	0.195	***	(0.007)	0.187	***	(0.007)
u_2	$\sigma^2_{u_2}$	0.229	***	(0.008)	0.220	***	(0.008)
u_3	$\sigma^2_{u_3}$	0.192	***	(0.008)	0.183	***	(0.008)
	Structural model						
Variance of z_{a_1}	σ^2_1	0.318	***	(0.011)	0.301	***	(0.010)
Autoregressive parameter	ϕ				0.874	***	(0.007)

Note 1: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.
Note 2: Significance of variances is not useful, as it is reported by Mplus based on a t-test, which is not suitable for variances.

In the first model specification, the variance of the attitudinal latent variable at wave 1 is estimated by 0.32. The variance does not change much across waves, indicating that the variability of attitudes remains almost the same over time. The estimated covariance matrix of $\mathbf{z}_a$ for the first model specification is given by

$$\hat{\mathbf{\Gamma}} = \begin{bmatrix} 0.32 & 0.25 & 0.22 & 0.21 & 0.19 \\ & 0.34 & 0.26 & 0.24 & 0.22 \\ & & 0.34 & 0.26 & 0.24 \\ & & & 0.34 & 0.26 \\ & & & & 0.33 \end{bmatrix}.$$

The estimated covariances among the attitudinal latent variables are positive and significant, indicating a strong positive correlation of a person's attitude towards women's work across waves. As one would expect, the further apart the waves, the weaker is the covariance between the attitudes. Furthermore, a loading is estimated for each time-dependent attitudinal latent variable on the corresponding dropout indicator at the next wave. From Table 5.2, these loadings are negative and significant at 10% level of significance, indicating that the more conservative an individual's attitude is towards women's work, the less likely they are to drop out of the study at the next wave, given the model specification and assumptions.

The last part of Table 5.2 gives results for the structural part of the second model specification. The estimated autoregressive parameter $\hat{\phi} = 0.874$, with estimated standard error of 0.007, again shows a highly significant and strong positive correlation of a person's attitude towards women's work over time. In other words, liberal/conservative views at a given wave are associated with liberal/conservative views at the preceding wave. The estimated dropout parameter $\hat{\gamma} = -0.036$, with estimated standard error of 0.009, shows a significant dependence of dropout on attitude at the first wave. The negative coefficient shows that the more conservative an individual's initial attitude is towards women's work, the less likely they are to drop out of the study. This conclusion too agrees with the one obtained from the first model specification. However, the BHPS is not a study of just women's work but it also includes many other items (not analysed here). We would not conclude that attitudes towards women's work are driving the dropout, but they are somehow associated, according to the given model specifications and assumptions, which makes the incorporation of a dropout mechanism essential to better model the data. Other factors or traits, such as demographic characteristics, socio-economic status,...etc are likely to affect both attitudes and dropout. Hence, we include covariates in Section 5.2.3, in an attempt to give better explanation of both attitudes and dropout.

The estimated means of the time-dependent attitudinal latent variable are, in order, 0.0, 0.057, 0.085, 0.101, and 0.103. This gradual increase in the mean indicates that as time goes by and people get older their views about women's work become more conservative. Another explanation is that since the more conservative people are less likely to drop out, the ones who remain in the study as time passes will tend to hold more conservative views.

5.2.2 Goodness-of-Fit

The sample size considered here is large. In this situation, the X^2 goodness of fit statistic is not very helpful, as it will tend to suggest significant lack of fit even given very small discrepancies between the fitted and observed covariance matrices

(Bijleveld et al. (1998)). We therefore evaluate the two models by their Root Mean Square Error of Approximation (RMSEA) and Comparative Fit Index (CFI). The first model specification has an RMSEA of 0.017 (probability RMSEA$\leqslant$ 0.05 is 1.00) and CFI 0.994, while the second has an RMSEA of 0.021 (probability RMSEA$\leqslant$ 0.05 is 1.00) and CFI 0.991. The two model specifications seem to fit the data almost equally well, giving us the choice of which one to adopt. In this case, the second specification seems to be the more attractive option since it is more parsimonious and involves directed relationships rather than free correlations among the latent variables.

5.2.3 Second Model Specification, with Covariates

Next, three time-invariant covariates (sex as a dummy variable for women, age at first wave and initial educational attainment) and one time-varying covariate (occupational status) are introduced to the second model specification and allowed to affect both the attitude towards women's work at each wave and the dropout mechanism. The sample is classified into 2660 (46%) male respondents, and 3159 (54%) females. The average age of respondents at the first wave is 45 years. Education is included as a binary variable that takes the value 1 if an individual has acquired at least a certificate of secondary education and 0 if no academic qualification is acquired. This is measured at the first wave and treated as time-invariant, as it tends to vary only slighty over time and is thus highly correlated across different waves. There are 2201 (38%) respondents with no academic qualification, while 3618 (62%) have acquired at least a certificate of secondary education, at the first wave. Occupational status is defined as a binary time-varying covariate which takes the value 1 if an individual is employed or out of the labour market including those who are retired, students or looking after family/home, and 0 if the individual is unemployed. At the first wave, 4724 (81%) respondents are employed while 1095 (19%) are classified as unemployed. The effect of covariates on the corresponding attitudes is constrained to be the same from wave 2 onwards. For the first wave, the effect of covariates on the attitude is allowed to be different, as this latent value is modelled solely as a function of covariates but not of previous attitudes.

Table 5.3 shows estimated regression coefficients of covariates on attitudes along with their significance. Sex, initial age and education seem to have a significant effect on attitudes towards women's work at the first wave. The negative coefficient of sex indicates that, as expected, women seem to have more liberal attitudes towards women's work. Both age and education have significant positive coefficients on attitude at the first wave. This indicates that older people and people with at least medium or high education at the beginning of the study have more conservative views about women's work. This is in addition to the before-mentioned conclusion that as people get older (i.e. in the subsequent waves) their views tend to get still more conservative. Although occupational status does not seem to have a significant effect on attitude at the first wave, it does have a significant effect from wave 2 onwards, indicating that those who are employed, retired or students have more liberal attitudes towards women's work than the unemployed. Sex ceases to have a significant effect from wave 2 onwards. This is probably due to the fact that its effect is already carried through the attitude from previous waves.

Table 5.4 shows estimated regression coefficients of covariates measured at first

Table 5.3: Parameter estimates for regression of attitudinal latent variables on covariates (sex, age, education and occupational status) for Model 2

	Effect on z_{a_1}		Effect on $z_{a_2}, \ldots, z_{a_5}$	
	Est.	Sig.	Est.	Sig.
Sex (woman)	-0.049	**	-0.001	
Age at first wave	0.001	*	-0.001	***
Education	0.197	***	0.026	***
Occupational status	-0.004		-0.104	***

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

wave on the dropout "latent variable" z_d, along with their significance. All the time-invariant covariates are significant. Sex has a negative coefficient, indicating that women are less likely to drop out. Age has a positive effect, meaning that older people are more likely to drop out, while the negative coefficient of education indicates that those with medium or high education are less likely to drop out of the study. In summary, older, less educated and male respondents have a higher propensity to drop out. It is worth mentioning that having accounted for those covariates, the dropout coefficient γ of the attitude at the first wave is still significant, indicating nonrandom dropout. However, it is now positive (0.027), opposite to the coefficient in the model without covariates. Thus it now indicates that controlling for these covariates, the more conservative an individual is at the first wave, the more likely he or she is to drop out. The likeliest explanation of this reversal is controlling for education, for which higher education is associated with more conservative attitudes but also with lower probability of dropout.

Table 5.4: Parameter estimates for regression of the dropout latent variable on covariates (sex, age, education and occupational status) for Model 2

	Effect on z_d	
	Est.	S.E.
Sex (woman)	-0.110	***
Age at first wave	0.011	***
Education	-0.080	**
Occupational status	0.033	

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

5.3 Sensitivity Analysis

Methods involving non-ignorable missing data should always be viewed as part of a sensitivity analysis in which the consequences of different modelling assumptions are explored (Little and Rubin (2002)). This section is dedicated to studying how sensitive the results of the models presented in Chapter 4 are to different levels of dropout under a model that includes the dropout mechanism, another that ignores it, and a third that uses listwise deletion and estimates parameters from the fully observed subjects only.

The same three items about attitudes towards women's work from the BHPS introduced in Section 5.1 are used here. Out of the 4296 cases that were fully observed on all five waves (1993, 95, 97, 99, 2001), a random sample of 1000 cases is selected for the purpose of this sensitivity analysis. The same covariates that were previously used in the analysis (sex, age at first wave and initial educational attainment as time-invariant covariates), and (occupational status) as a time-varying covariate are used here.

The model fitted to the data is the second model specification outlined in Section 4.4 and illustrated by Figure 4.2, where the measurement model is presented by equations (4.1)–(4.2) along with the corresponding assumptions and the structural model is (4.4) allowing for covariates to affect attitudes.

The model is fitted at four levels of dropout: all 1000 cases fully observed at all waves, 10% , 25% and 50% dropout. Dropout is artificially created on waves 97, 99 or 2001 such that, in general, the more conservative individuals are more likely to dropout, and the more they are conservative the earlier they drop out. This is done by running a factor analysis on the three items observed at the first wave considered here (wave 1993), and a factor score is computed accordingly. The higher an individual's score, the more conservative are his or her views towards women's work. Estimated factor scores from the sample give a mean of 0.03 and standard deviation 1.03, with a minimum of -2.8 and a maximum of 3.3. Dropout is created for a random sample -depending on the selected percentage of dropout- of those having a positive score. If the predicted factor score for a respondent is between 0 and 0.5, he/she may drop out on the last wave (2001). If his/her score is between 0.5 and 1.5, he/she may drop out in 1999; and if the score is 1.5 or higher the respondent may drop out on the third wave (1997). In case of 50% dropout, all those who have a positive score drop out at some point. The choice of these cutoff points is quite subjective. Covariates are fully observed in all cases for all scenarios.

5.3.1 Listwise Deletion

Table 5.5 summarises results for a model that uses a dataset with listwise deletion for all cases that drop out at some point of the study, and carries out the analysis on complete cases only. The first column shows results from analysis of the fully observed data, without any dropouts. The sample size is different at each level of dropout, according to the number of deleted cases. In case of 50% dropout, all respondents with a positive score at first wave are deleted as they drop out at some point of the study, leaving 521 cases for analysis. The model is a good fit even with 50% dropout, as it only considers the complete data. Results of the measurement model and variances of random effects are robust to a great extent. The autoregressive parameter ϕ, also changes only slightly as the level of dropout increases, still capturing the dependence of attitudes on previous waves. A natural result of this model is underestimation of the variance of the attitudinal latent variable at first wave, especially at the extreme case of 50% dropout. As the more conservative respondents are eliminated from the dataset, attitudes of the remaining respondents exhibit less variability.

Problems with detecting the same significant covariates effects as for the complete cases arise even with 10% dropout, where Employment 93 is declared significant and Sex is not, opposite to the complete cases results. It gets even worse as the level of dropout increases. At 50% dropout, only one of the initially significant covariates (Employment t) is detected as a significant covariate on attitude at waves 2 onwards.

Table 5.5: A sensitivity analysis for parameter estimates at four levels of dropout (0%, 10%, 25% and 50%) when dropouts are treated by listwise deletion

	Complete cases		10% dropout		25% dropout		50% dropout	
Sample size	1000		900		750		521	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
	Goodness of fit							
RMSEA	0.021	***	0.021	***	0.021	***	0.030	***
CFI	0.990	-	0.990	-	0.991	-	0.971	-
	Measurement model							
Loadings								
Family λ_1	1		1		1		1	
Contribution λ_2	1.052	***	1.060	***	1.078	***	1.081	***
Independent λ_3	1.135	***	1.143	***	1.141	***	1.081	***
	Random effects							
Variances								
$\sigma^2_{u_1}$	0.188	***	0.191	***	0.184	***	0.226	***
$\sigma^2_{u_2}$	0.231	***	0.247	***	0.234	***	0.248	***
$\sigma^2_{u_3}$	0.154	***	0.148	***	0.161	***	0.229	***
	Structural model							
Variance of z_{a1}; σ^2_1	0.320	***	0.313	***	0.307	***	0.110	***
Autoregressive parameter; ϕ	0.883	***	0.878	***	0.862	***	0.788	***
	Covariates effects							
	Covariates effects on attitude at first wave z_{a_1}							
Education 93	0.236	***	0.229	***	0.203	***	0.113	
Employment 93	0.090		0.138	*	0.161	*	0.168	
	Covariates effects on attitude at subsequent waves $z_{a_2}, ..., z_{a_5}$							
Sex	-0.020	*	-0.017		-0.017		-0.017	
Education 93	0.026	*	0.029	*	0.026		0.030	
Employment t	-0.125	***	-0.132	***	-0.165	***	-0.164	***

Note 1: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.
Note 2: Significance of variances is not useful, as it is reported by Mplus based on a t-test, which is not suitable for variances.

Table 5.6 shows means of time-dependent attitudinal latent variables calculated from the estimated parameters obtained from analysing the data where dropouts are deleted. As the level of dropout increases, the calculated means fail to capture the same trend of decreasing over time as does the complete case. This may be due to the fact that attitudes of those who remain in the study (the more liberal) behave differently.

Table 5.6: Estimated means for attitudinal latent variables under four levels of dropout (0%, 10%, 25% and 50%) when dropouts are treated by listwise deletion

	Means of attitudinal latent variables			
	Complete cases	10% dropout	25% dropout	50% dropout
z_{a_1}	0.260	0.314	0.328	0.212
z_{a_2}	0.321	0.322	0.489	0.447
z_{a_3}	0.184	0.178	0.214	0.402
z_{a_4}	0.172	0.216	0.299	0.601
z_{a_5}	0.099	0.109	0.155	0.458

5.3.2 Ignoring The Dropout Mechanism

When cases with missing data are not deleted, WLS in MPlus uses pairwise information for estimation. That is, first and second stage WLS estimates are obtained by univariate and bivariate listwise deletion, respectively, using ML thus making use of all available pairwise data. Weighted least squares is then used to estimate the weight matrix and to fit the model. This means that different sample sizes will be used for different pairs of items depending on data available for each pair of items.

Table 5.7 summarises parameter estimates and their significance if none of the observations is deleted, yet no dropout model is specified at different levels of dropout (0%, 10%, 25% and 50%). Both measurement and structural parts of the model are quite robust, even with 50% dropout, with an exception of the autoregressive parameter ϕ, which drops to 0.699 in case of 50% dropout indicating weaker dependence of attitude on the previous wave. Two of the significant covariates effects (Sex and Education) on attitudes measured at second wave onwards are not detected at 25% or 50% dropout levels.

Table 5.7: A sensitivity analysis for parameter estimates at four levels of dropout (0%, 10%, 25% and 50%) when the dropout mechanism is ignored

	Complete cases		10% dropout		25% dropout		50% dropout	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
	Goodness of fit							
RMSEA	0.021	***	0.021	***	0.018	***	0.017	***
CFI	0.990	-	0.990	-	0.991	-	0.991	-
	Measurement model							
Loadings								
Family λ_1	1		1		1		1	
Contribution λ_2	1.052	***	1.053	***	1.050	***	1.033	***
Independent λ_3	1.135	***	1.136	***	1.119	***	1.092	***
	Random effects							
Variances								
$\sigma^2_{u_1}$	0.188	***	0.189	***	0.185	***	0.203	***
$\sigma^2_{u_2}$	0.231	***	0.245	***	0.236	***	0.252	***
$\sigma^2_{u_3}$	0.154	***	0.155	***	0.163	***	0.187	***
	Structural model							
Variance of z_{a1}; σ^2_1	0.320	***	0.321	***	0.330	***	0.375	***
Autoregressive parameter; ϕ	0.883	***	0.871	***	0.844	***	0.699	***
	Covariates effects							
	Covariates effects on attitude at first wave z_{a_1}							
Education 93	0.236	***	0.235	***	0.237	***	0.267	***
Employment 93	0.090		0.110		0.113		0.114	
	Covariates effects on attitude at subsequent waves $z_{a_2}, ..., z_{a_5}$							
Sex	-0.020	*	-0.016		-0.016		-0.027	
Education 93	0.026	*	0.029	*	0.030		0.027	
Employment t	-0.125	***	-0.125	***	-0.115	***	-0.163	***

Note 1: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.
Note 2: Significance of variances is not useful, as it is reported by Mplus based on a t-test, which is not suitable for variances.

Table 5.8 gives means of time-dependent attitudinal latent variables calculated from the estimated model, which reflects a closer trend at 50% dropout to the complete case than the one captured in case of listwise deletion.

Table 5.8: Estimated means for attitudinal latent variables under four levels of dropout (0%, 10%, 25% and 50%) when the dropout mechanism is ignored

	Means of attitudinal latent variables			
	Complete cases	10% dropout	25% dropout	50% dropout
z_{a_1}	0.260	0.272	0.271	0.303
z_{a_2}	0.321	0.338	0.313	0.278
z_{a_3}	0.184	0.173	0.129	0.005
z_{a_4}	0.172	0.162	0.146	0.058
z_{a_5}	0.099	0.043	0.051	-0.200

5.3.3 Incorporating The Dropout Mechanism

Table 5.9 summarises parameter estimates and their significance for the 4 scenarios: all 1000 cases fully observed, 10% , 25% and 50% dropout; for a model that accounts for dropout using the dropout function (4.7). It is noted from the RMSEA and CFI that the fit of the model gets worse as the percentage of dropout increases, indicating lack of fit at 50% and 25% dropout. However, the estimated parameters of the measurement and structural models are quite robust up to 25% level of dropout. Parameters of the structural model break down at 50% dropout. The estimated autoregressive parameter ϕ decreases as the level of dropout increases. It drops from 0.88 in a complete case analysis to 0.45 in case of 50% dropout. The dropout indicators are, in an indirect way, measures of the attitude at first wave. This may explain why dependence among attitudes becomes weaker, and also why the estimated variance σ_1^2 of the first attitudinal latent variable gets bigger, as the level of dropout increases. The dropout parameter γ varies at different levels of dropout, but remains highly significant in all cases capturing the artificially created nonrandom dropout.

Table 5.9: A sensitivity analysis for parameter estimates at four levels of dropout (0%, 10%, 25% and 50%) when the dropout mechanism is incorporated

	Complete cases		10% dropout		25% dropout		50% dropout	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
	Goodness of fit							
RMSEA	0.021	***	0.036	***	0.060		0.110	
CFI	0.990	-	0.966	-	0.898	-	0.677	-
	Measurement model							
Loadings								
Family λ_1	1		1		1		1	
Contribution λ_2	1.052	***	1.051	***	1.052	***	1.108	***
Independent λ_3	1.135	***	1.132	***	1.116	***	1.161	***
	Random effects							
Variances								
$\sigma^2_{u_1}$	0.188	***	0.191	***	0.188	***	0.284	***
$\sigma^2_{u_2}$	0.231	***	0.248	***	0.238	***	0.319	***
$\sigma^2_{u_3}$	0.154	***	0.160	***	0.168	***	0.265	***
	Structural model							
Variance of z_{a1}; σ^2_1	0.320	***	0.332	***	0.338	***	0.500	***
Autoregressive parameter; ϕ	0.883	***	0.854	***	0.830	***	0.450	***
Dropout parameter z_d on z_{a1}; γ			0.207	***	0.052	***	0.355	***

Note 1: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.
Note 2: Significance of variances is not useful, as it is reported by Mplus based on a t-test, which is not suitable for variances.

Table 5.10 gives estimated parameters for covariates effects at the four levels of dropout studied here. Even at the severe case of 50% dropout, the significant covariates affecting attitude at first wave and at subsequent waves are still detected by the model, with only one exception (Sex) which is never detected under any treatment for dropout (listwise deletion, ignoring dropout or incorporating dropout mechanism) even at 10%. Education is the only significant covariate on the dropout mechanism at 25% and 50% levels of dropout.

Table 5.10: A sensitivity analysis for covariates effects at four levels of dropout (0%, 10%, 25% and 50%) when the dropout mechanism is incorporated

	Complete cases		10% dropout		25% dropout		50% dropout	
	Est.	Sig.	Est.	Sig.	Est.	Sig.	Est.	Sig.
	Covariates effects on attitude at first wave z_{a_1}							
Education 93	0.236	***	0.236	***	0.238	***	0.269	***
Employment 93	0.090		0.112		0.114		0.117	
	Covariates effects on attitude at subsequent waves $z_{a_2}, ..., z_{a_5}$							
Sex	-0.020	*	-0.017		-0.017		-0.040	
Education 93	0.026	*	0.032	**	0.034	*	0.075	**
Employment t	-0.125	***	-0.128	***	-0.157	***	-0.178	***
	Covariates effects on dropout latent variable z_d							
Education 93			0.123		0.162	***	0.191	***

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

The calculated means of time-dependent attitudinal latent variables are given in Table 5.11 under the four scenarios of dropout. There is a gradual decrease in means over time captured at all levels of dropout, despite the difference in estimated values. The estimated means remain close to the complete case when the dropout is at 10% and 25%. However, the values estimated in case of 50% dropout are somehow further. This can be again due to dropout indicators being measures of the attitude at first wave thus affecting its mean, and that as conservatives disappear from the study the dependence on previous attitudes becomes weaker.

Table 5.11: Estimated means for attitudinal latent variables under four levels of dropout (0%, 10%, 25% and 50%) when the dropout mechanism is incorporated

	Means of attitudinal latent variables			
	Complete cases	10% dropout	25% dropout	50% dropout
z_{a_1}	0.260	0.274	0.273	0.308
z_{a_2}	0.321	0.337	0.311	0.211
z_{a_3}	0.184	0.171	0.126	-0.027
z_{a_4}	0.172	0.160	0.144	0.073
z_{a_5}	0.099	0.043	0.052	-0.120

In summary, it is noted that estimation of the measurement model parameters is quite robust under different treatments for dropouts. Whereas listwise deletion and ignoring dropout may give closer estimates to the complete cases analysis in case of 50% dropout, especially on the structural part of the model, incorporating the dropout mechanism has an advantage on the fixed part of the model. While both a model that uses listwise deletion and another that ignores the dropout fail to identify some of the significant covariates effects on attitudes at 25% and 50% dropout level, incorporating the dropout mechanism still captures their significance even at this high level of dropout despite its poor fit. Up to 25% dropout, which is the same percent of dropout that we had in the real data analysis, incorporating the dropout mechanism does equally well in terms of estimating model parameters as listwise deletion or ignoring dropout, yet it outperforms those two in detecting significant covariates.

Chapter 6

An IRT Model for Multivariate Binary Longitudinal Data Subject to Dropout

This chapter presents a latent variable model for multivariate binary longitudinal data subject to dropout. The model is similar in its setup to the second model specification presented in Chapter 4. However, the items considered here are all binary, and the model is fitted within an IRT approach where distributional assumptions are directly made on the items intead of assuming underlying continuous variables as in a SEM framework. We choose to fit this model for binary items to start with, as they are the simplest to model. Once established, the model can be extended to accommodate ordinal items as in Samejima (1969)'s Graded Response model or a Partial Credit model (Masters (1982, 1988)). Further extensions may include nominal, count, metric or mixed items.

The measurement model is first outlined, in which the observed binary variables are explained by a continuous latent variable at each time point, followed by the structural part which defines relationships among the latent variables over time, allowing for covariates effects. The dropout mechanism is then introduced and linked to the latent attitudes, and covariates, allowing the dropout to be nonrandom.

6.1 Modelling The Observed Indicators: The Measurement Model

The model developed here is for binary observed variables. Suppressing the index m for a subject for convenience, let $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{pt})$ be $(p \times 1)$ vectors of observed binary variables for a single subject at times $t = 1, 2, ..., T$. The items $\mathbf{y}_t = (y_{1t}, y_{2t}, ..., y_{pt})$ at each time t are regarded as measures of a continuous attitudinal time-dependent latent variable z_{a_t}, which is assumed to be normally distributed. For simplicity, the model below is presented assuming that the items are unidimensional, but it can be extended to accommodate more latent variables. It is assumed that item-specific random effects u_i, introduced to account for repitition of items over time, are independently normally distributed; $u_i \sim N(0, \sigma^2_{u_i})$ for $i = 1, ..., p$.

Each manifest binary variable y_{it} is assumed to have a Bernoulli distribution with a conditional probability of positive response denoted by $\pi_{it} = \Pr(y_{it} = 1 \mid z_{a_t}, u_i)$. This probability is modelled via a logit link as a function of the underlying time-dependent latent variable z_{a_t} and the item-specific random effect u_i,

$$\operatorname{logit} \pi_{it}(z_{a_t}, u_i) = \alpha_i + \lambda_i z_{a_t} + u_i, \quad i = 1, \ldots, p, \; t = 1, ...T, \tag{6.1}$$

where α_i is a constant term, and λ_i is the loading of the latent variable z_{a_t} on the i^{th} binary item y_{it}. The probability of a positive response can thus be expressed as

$$\pi_{it}(z_{a_t}, u_i) = \frac{e^{\alpha_i + \lambda_i z_{a_t} + u_i}}{1 + e^{\alpha_i + \lambda_i z_{a_t} + u_i}}. \tag{6.2}$$

As in the model defined in Chapter 4, we have imposed the assumption of *invariance of measurement* across time for each item $i = 1, \ldots, p$, by constraining the constant term α_i and the loading λ_i for each $i = 1, \ldots, p$ to be the same at all time points $t = 1, \ldots, T$; in order to guarantee that the latent variable has the same interpretation at each time point. In order to set the scale for the time-dependent attitude latent variables, and for their variances $\sigma_1^2, ..., \sigma_T^2$ to be estimable, the loading λ_1 on the first observed variable y_{1t} is set to 1, at all time points t.

6.2 Modelling The Latent Variables: The Structural Model

The attitudinal latent variables are assumed to be linked via a first-order autoregressive [AR(1)] structure as outlined in Chapter 4, allowing for covariates effects

$$z_{a_t} = a_t + \phi z_{a_{t-1}} + \boldsymbol{\beta}_t' \mathbf{x}_t + \delta_t, \quad t = 2, ..., T, \tag{6.3}$$

with the same set of assumptions and constraints, while the attitudinal latent variable at the first wave z_{a_1} is modelled solely as a function of covariates

$$z_{a_1} = \boldsymbol{\beta}_1' \mathbf{x}_1 + \delta_1. \tag{6.4}$$

6.3 The Dropout Mechanism

As in Chapter 4, dropout is the only form of missingness considered here, assuming there is no intermittent missingness or item non-response. Dropout indicators d_t, $t = 1, \ldots, T$, are as defined previously such that $d_t = 0$ when $\mathbf{y}_t$ is observed and $d_t = 1$ if a respondent drops out at time t. After the time of dropout, d_t itself is regarded as missing and can be set to an arbitrary value such as -1. Observations at the first occasion are assumed to be complete, so that $d_1 = 0$ for all respondents, hence the dropout vector is given by $\mathbf{d} = (d_2, ..., d_T)$.

Adopting the second model specification presented in Chapter 4, dropout indicators are modelled as functions of attitude measured at first wave z_{a_1}, in addition to covariates $\mathbf{x}_1$; also measured at first wave. However, a dropout latent variable is not considered here and dropout indicators are modelled directly as functions of attitude and covariates. Each binary dropout indicator d_t is assumed to follow a Bernoulli distribution, with a conditional probability (hazard) of dropping out at time t denoted by $h_t = \Pr(d_t = 1 \mid z_{a_1}, \mathbf{x_1})$. This probability is modelled via a logit link as a function of the first attitudinal latent variable z_{a_1}, and covariates $\mathbf{x_1}$

$$\operatorname{logit} h_t(z_{a_1}, \mathbf{x_1}) = \alpha_{dt} + \gamma z_{a1} + \boldsymbol{\beta}_d' \mathbf{x}_1, \quad t = 2, ..., T, \tag{6.5}$$

where α_{dt} is a constant, γ is a regression coefficient representing the dependence of the dropout indicator d_t on the attitude latent variable z_{a_1} at the first time point, and $\boldsymbol{\beta}_d$ is a vector of regression coefficients of covariates $\mathbf{x}_1$. The effect γ of attitude at first wave on dropping out is constrained to be the same at all waves, and is thus made responsible for determining whether or not the dropout is random. If the coefficient γ turns out to be significant, this reflects the dependence of probability of dropout at any wave on attitude at first wave, which means that the dropout is nonrandom.

Probability of a dropout indicator being missing ($d_t = -1$) is fully determined by the value of the preceding indicator d_{t-1}. A dropout indicator d_t is recorded as missing if dropout has occured at any previous wave, $\Pr(d_t = -1 \mid d_{t-1} = 1\, or\, -1) = 1$. On the other hand, if a respondent has been observed at the directly preceding wave $t-1$, the dropout indicator d_t cannot be missing; $\Pr(d_t = -1 \mid d_{t-1} = 0) = 0$. It can either be observed $d_t = 0$ or a dropout at time t, $d_t = 1$.

The latent variable model outlined above by equations (6.1, 6.3, 6.4 and 6.5) is illustrated by a path diagram for an example with four time points in Figure 6.1.

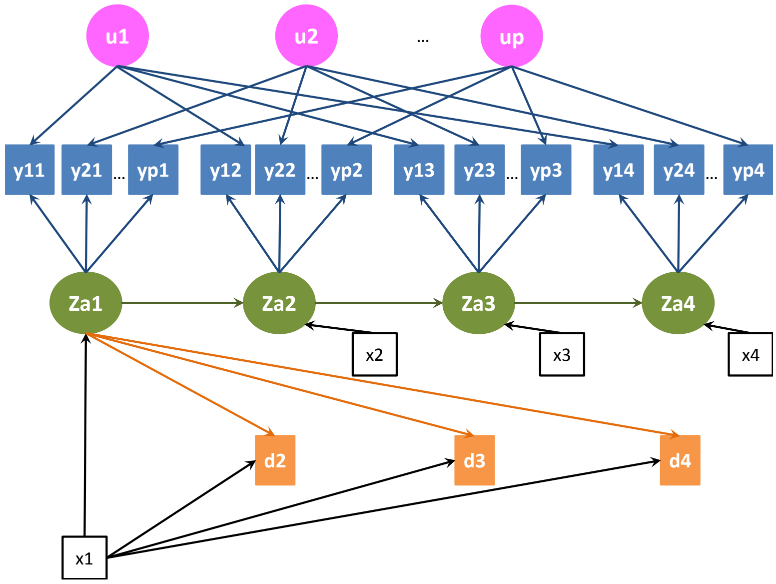


Figure 6.1: Path diagram for a model where attitude at first wave affects missingness on all waves, an example with four time points

6.4 Estimation

There are two main approaches for obtaining maximum likelihood estimates for latent variable models, the first of which depends on iterative techniques such as the Expectation-Maximisation (EM) algorithm, first introduced by Dempster et al. (1977), where the expectation of the complete likelihood is obtained with respect to the posterior distribution of the latent variables given the observed data. In many cases, the expectation can not be obtained in closed form and is thus approximated numerically. The maximisation step is then implemented using algorithms such as Newton-Raphson, and improved estimates for parameters are obtained. The above steps are repeated until convergence is attained and maximum likelihood estimates are obtained. Bartholomew et al. (2011) outline the use of EM to fit latent variable models, while Cagnone et al. (2009) use it to fit a latent variable model for multivariate longitudinal ordinal data. As models get more complex, so does the implementation of EM.

An alternative methodology for estimating parameters of a latent variable model is to adopt a Bayesian approach based on MCMC. Unlike the EM algorithm, MCMC does not require exact numerical calculation for the E-step, or pre-calculation of derivatives for the M-step, thus providing easier implementation. Patz and Junker (999a) develop an MCMC estimation technique for complex IRT models. They extend their technique to address issues such as non-response and missingness in their follow up paper (Patz and Junker (999b)). Moustaki and Knott (2005) compare the EM and MCMC estimation methods for latent variable models, where they use real examples with categorical data to illustrate this comparison.

The model specification presented here has been introduced by Dunson (2003) in a generalised linear latent variable model framework for different response types where Markov Chain Monte Carlo (MCMC) methods were used for estimation. Cagnone et al. (2009) propose a full-information maximum likelihood estimation method for the same model specification with ordinal variables. Cai (2010) develops an EM algorithm for full-information maximum marginal likelihood estimation that is computationally efficient due to the use of a dimension reduction technique of the latent variable space for the two-tier item factor analysis model, which fits into this model specification. Composite likelihood approaches have also been proposed to reduce estimation complexity for this type of models (see Vasdekis et al. (2012)). None of these papers consider dropout, though. We choose to depend on MCMC for estimation of our developed model, as it presents a flexible tool for fitting complex latent variable models.

Let $\boldsymbol{\theta}$ denote a vector of all parameters defining the model outlined by equations (6.1, 6.3, 6.4 and 6.5), including α_i and λ_i defining the measurement process; a_t, ϕ, $\boldsymbol{\beta}'_1$ and $\boldsymbol{\beta}'_t$ defining the structural model; and α_{dt}, γ and $\boldsymbol{\beta}'_{\boldsymbol{d}}$ defining the dropout mechanism, in addition to variances of errors υ_t^2 and random effects $\sigma_{u_i}^2$. For a random sample of size n, the marginal likelihood of the observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{m=1}^{n} \int_{\mathbf{z}_a} \int_{\mathbf{u}} \left\{ \prod_{i=1}^{p} \prod_{t \in \mathcal{H}_m^Y} \pi_{mit}(z_{a_t}, u_i)^{y_{mit}} (1 - \pi_{mit}(z_{a_t}, u_i))^{1-y_{mit}} \right\}$$

$$\times \left\{ \prod_{t \in \mathcal{H}_m} h_{mt}(z_{a_1}, \mathbf{x_1})^{d_{mt}} (1 - h_{mt}(z_{a_1}, \mathbf{x_1}))^{1-d_{mt}} \right\} \times h(\mathbf{z}_a, \mathbf{u}) d\mathbf{z}_a d\mathbf{u}, \qquad (6.6)$$

where $\mathcal{H}_m^Y$ is the set of time points prior to dropout for an individual m where $\mathbf{y}_{mt}$ is observed, and $\mathcal{H}_m$ is the set of time points prior to and including time of dropout for an individual m. Moreover, π_{mit} is the probability that an individual m gives a positive response to item i at time t, conditional on the attitudinal latent variable z_{a_t} and random effect u_i given by equation (6.2), h_{mt} is the probability of dropping out at time t, conditional on the first attitudinal latent variable z_{a_1} and covariates measured at first wave $\mathbf{x}_1$ as expressed by equation (6.5), and $h(\mathbf{z}_a, \mathbf{u})$ is the joint distribution of attitude latent variables and random effects. A respondent's contribution to the likelihood is thus weighted by his/her probability of being observed. At time of dropout, the contribution is merely the probability of dropping out at this time point. After dropout, data about a respondent is completely missing, and therefore there is no contribution to the likelihood.

The loglikelihood is thus

$$\log L(\boldsymbol{\theta}) = \sum_{m=1}^{n} \left[\int_{\mathbf{z}_a} \int_{\mathbf{u}} \left\{ \sum_{i=1}^{p} \sum_{t \in \mathcal{H}_m^Y} \left(y_{mit} \log \pi_{mit}(z_{a_t}, u_i) + (1 - y_{mit}) \log \left(1 - \pi_{mit}(z_{a_t}, u_i)\right) \right) \right. \right.$$

$$+ \sum_{t \in \mathcal{H}_m} \left(d_{mt} \log h_{mt}(z_{a_1}, \mathbf{x_1}) + (1 - d_{mt}) \log \left(1 - h_{mt}(z_{a_1}, \mathbf{x_1})\right) \right)$$

$$\left. \left. + \log h(\mathbf{z}_a, \mathbf{u}) \right\} d\mathbf{z}_a d\mathbf{u} \right]. \tag{6.7}$$

The above expression requires a $(T+p)-$dimensional integration, which makes its evaluation complicated especially as the number of waves T increases, thus making estimation using MCMC an appealing option.

6.4.1 Bayesian Estimation Using MCMC

In this section, we give a sketch of Bayesian estimation using MCMC. In Bayesian estimation, parameters are treated as random variables rather than fixed quantities as in a frequentist approach. Inference about unobserved parameters is based on the posterior distribution of the unobserved quantities (including parameters and latent variables) conditional on the observed data. MCMC is used to make draws from this posterior distribution. We use Win**BUGS** (**B**ayesian inference **U**sing **G**ibbs **S**ampling) (Lunn et al. (2000)) for estimation.

Let $\mathbf{v}$ denote a vector with all the unknown quantities including parameters and latent variables; such that $\mathbf{v}' = (\boldsymbol{\theta}, \mathbf{z_a}, \mathbf{u})$. The loglikelihood given by (6.7) can be written as

$$\log L(\mathbf{v} \mid \mathbf{y}, \mathbf{x}) = \sum_{m=1}^{n} \log f(\mathbf{y}_m) = \sum_{m=1}^{n} \log \int \cdots \int g(\mathbf{y}_m \mid \mathbf{v}, \mathbf{x}) h(\mathbf{v}) d\mathbf{v}. \tag{6.8}$$

The joint posterior distribution of the parameter vector $\mathbf{v}$ is

$$h(\mathbf{v} \mid \mathbf{y}, \mathbf{x}) = \frac{g(\mathbf{y} \mid \mathbf{v}, \mathbf{x}) h(\mathbf{v})}{f(\mathbf{y})} \propto g(\mathbf{y} \mid \mathbf{v}, \mathbf{x}) h(\mathbf{v}). \tag{6.9}$$

In case of binary items, $g(y_i \mid \mathbf{v}, \mathbf{x})$ is of the form (2.5). The main steps of the Bayesian approach for such a latent variable model are as outlined by Bartholomew et al. (2011) and Moustaki and Knott (2005):

1. Inference is based on the posterior distribution $h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})$, of the unknown parameters $\mathbf{v}$ conditional on the observed data $\mathbf{y}$ and covariates $\mathbf{x}$. Depending on the model fitted, the form of the distribution can be very complex.

2. The mean vector of the posterior distribution $h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})$ can be used as an estimator of $\mathbf{v}$.

3. Standard deviation of the posterior distribution $h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})$ can be used to compute standard errors of parameter estimates.

4. In general, the posterior mean $E(\psi(\mathbf{v}) \mid \mathbf{y}, \mathbf{x})$ can be used as a point estimate of a function of the parameters $\psi(\mathbf{v})$, where $E(\psi(\mathbf{v}) \mid \mathbf{y}, \mathbf{x}) = \int \cdots \int \psi(\mathbf{v}) h(\mathbf{v} \mid \mathbf{y}, \mathbf{x}) d\mathbf{v}$.

5. Analytic evaluation of the above expectation is impossible. Alternatives include numerical evaluation, analytic approximations and Monte Carlo Integration.

Markov Chain Monte Carlo (MCMC)

To avoid the integration required in the posterior expectation, Monte Carlo integration is used in which the integrals are approximated by an average of quantities calculated from sampling. Samples are drawn from the posterior distribution of all the unknown parameters $h(\mathbf{v}^{(r)} \mid \mathbf{y}, \mathbf{x})$, and the expectation over the posterior is approximated by the average over N samples:

$$E(h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})) = \frac{1}{N} \sum_{r=1}^{N} h(\mathbf{v}^{(r)} \mid \mathbf{y}, \mathbf{x}).$$

The samples drawn from the posterior distribution do not have to be independent. Samples are drawn from the posterior distribution through a Markov chain with $h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})$ as its stationary distribution. Algorithms such as the Gibbs sampler and Metropolis-Hastings are used in WinBUGS to get the unique stationary distribution. In particular, Gibbs sampling is an algorithm that produces a sequence of iterations $\mathbf{v}^0, \mathbf{v}^1, \ldots, \mathbf{v}^k$ that form a Markov chain, which eventually converges to its stationary distribution, taken to be the posterior distribution.

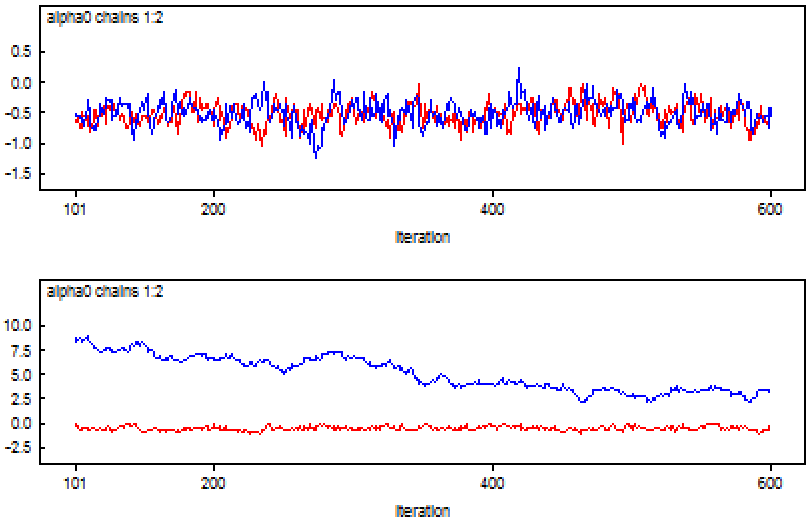
6.4.2 Choosing Prior Distributions

The posterior distribution $h(\mathbf{v} \mid \mathbf{y}, \mathbf{x})$ of the unknown parameters given the data, is obtained by multiplying the likelihood by a prior distribution as shown in equation (6.9). Thus, a prior distribution needs to be assumed for each parameter of interest of the vector $\mathbf{v}$. We assume vague or non-informative priors to emphasise the likelihood of the data rather than the prior. A normal distribution with mean 0 and a large variance taken to be 10000 is assumed for all parameters of interest defining the outlined model. Random effects $\mathbf{u}$ are assumed to be independently normally distributed, each with mean 0. Standard deviations of errors υ_t and random effects σ_{u_i} are assumed a Uniform (0.0001, 100) prior. The wide range of the uniform distribution serves as a non-informative prior. The lower limit is taken to be 0.0001, rather than 0, to avoid a possible trap in WinBUGS if the standard deviation happens to be exactly equal to the lower bound in one of the iterations, thus causing the estimation process to stop.

6.4.3 Assessing Convergence in MCMC

One of the main issues with MCMC estimation, is when to decide that the produced Markov chain has converged to its stationary distribution, which is the posterior distribution of the parameters given the data. A number of iterations, known as the *burn-in* session, is usually discarded by the user before starting to monitor parameters. Convergence can be checked graphically by looking at trace plots showing the full history of estimated values plotted against iteration number for each parameter. A chain is said to has converged when trace plots for parameters depict random patterns that move around the parameter space quickly indicating that the chain is *mixing* well. If the chain gets stuck in certain areas or shows a specific trend, this is an indication that it has not converged. It is common practice to run more than one chain simultaneously. In that case, one can be reasonably confident about convergence if all the chains are overlapping one another. Figure 6.2 presents two examples of trace plots from the WinBUGS manual. The top trace plot is an example of two chains for which convergence looks reasonable, while the bottom exhibits an example of two chains which have clearly not reached convergence. Trace plots need to be checked for all parameters of interest.

Figure 6.2: Examples from WinBUGS manual showing: (top) multiple chains for which convergence looks reasonable, (bottom) multiple chains which have not reached convergence



Plots of the autocorrelation function can also be used to assess convergence, where for a convergent chain, the autocorrelation decreases as the number of iterations increases. One can also check density plots for the posterior distribution to see whether parameters have reached a distribution that looks reasonably normal. However, making a decision about convergence via visual inspection can be quite subjective.

A more formal approach to assess convergence is via convergence diagnostics. These are statistics that have been developed by researchers to facilitate making such a tricky decision as convergence. An extensive review of convergence assessment techniques for MCMC is given in Brooks and Roberts (1998). Several convergence diagnostics including those proposed by Raftery and Lewis (1992), Geweke (1992), Heidelberger and Welch (1983), Gelman and Rubin (1992) and

Brooks and Gelman (1998) can be produced by CODA (Plummer et al. (2006)); an R package for analysing output obtained from WinBUGS.

One of the most widely used criteria for convergence is that proposed by Gelman and Rubin (1992). Their method relies on monitoring several sequences, with different starting points sampled from an overdispersed distribution, and detecting when the chains have *forgotten* their starting values and converged to the same stationary distribution. This method is based on analysis of variance by comparing variability between chains to variability within chains, and detects convergence when there is no much difference between the two. The method is briefly outlined below.

Suppose we run $m \geq 2$ independent simulations of length $2n$, each beginning at different starting points. The first n iterations are discarded, then for each parameter of interest θ, the between-sequence variance B and the within-sequence variance W are computed as follows:

$$B = \frac{n}{m-1} \sum_{i=1}^{m} \left(\bar{\theta}_{i.} - \bar{\theta}_{..} \right)^2 ,$$

$$W = \frac{1}{m} \sum_{i=1}^{m} s_i^2,$$

where

$$s_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (\theta_{ij} - \bar{\theta}_{i.}).$$

An estimate of the target distribution variance can be obtained by a weighted average of the two variance components B and W,

$$\widehat{Var}(\theta) = \frac{n-1}{n} W + \frac{1}{n} B,$$

which overestimates the variance in case the starting points are overdispersed, but is unbiased under stationarity (i.e. if the starting points were actually drawn from the target distribution). For any finite n, W underestimates the variance of θ because the individual sequences have not had time to range over all of the target distribution. As $n \to \infty$, both estimators $\widehat{Var}(\theta)$ and W will approach $Var(\theta)$.

Gelman and Rubin (1992) suggest monitoring convergence by estimating the factor by which the scale of the current distribution of θ will shrink as $n \to \infty$. They call it the *Potential Scale Reduction Factor* (PSRF) and estimate it by the ratio of the current variance estimate $\widehat{Var}(\theta)$ to the within-sequence variance W with some correction factor

$$\sqrt{\hat{R}} = \sqrt{\frac{\widehat{Var}(\theta)}{W} \times \frac{df}{df - 2}},$$

where df stands for the degrees of freedom of the approximate $t-$distribution for θ. If $\hat{R}$ is large, this is taken to be an indication of non-convergence. Further simulations may result in either decreasing the overestimated $\widehat{Var}(\theta)$, or increasing the underestimated W. When the PSRF is close to 1, the Markov chains are believed to have converged to a stationary distribution. It is necessary to inspect the PSRF for every parameter of interest in any model to see whether or not it has reached convergence.

Brooks and Gelman (1998) have extended Gelman and Rubin's PSRF to consider more than one parameter simultaneously. They propose a *Multivariate Potential Scale Reduction Factor* (MPSRF) that summarises all univariate measures in a single diagnostic. However, for high dimensional problems, they suggest calculating the MPSRF as an overall indicator of convergence while still inspecting the PSRF for each parameter.

6.5 Application

The model presented in this chapter, defined by equations (6.1, 6.3, 6.4 and 6.5) and illustrated by Figure 6.1, is now applied to the data on people's attitudes towards women's work that has been already introduced in Chapter 5. The same five waves of the British Household Panel Survey (1993, 95, 97, 99, 2001) are considered, with 5819 respondents who gave complete answers on the first wave subject to dropping out from the second wave onwards. However, the originally five-category ordinal variables have been dichotomised for the sake of the analysis into binary items taking one of two possible values: 0 if an individual *Strongly Agrees* or *Agrees* to an item, and 1 if his/her response is *Don't Know*, *Disagree* or *Strongly Disagree*.

Three time-invariant covariates (sex as a dummy variable for women, age at first wave and initial educational attainment) and one time-varying covariate (occupational status) are allowed to affect the attitude towards women's work at each wave. For more details about these covariates, see Section 5.2.3. The effect of covariates on the corresponding attitudes is constrained to be the same from the second wave onwards. For the first wave, the effect of covariates on the attitude is allowed to be different, as this latent variable is modelled solely as a function of covariates and not of previous attitudes (see equation (6.4)). The same four covariates measured at first wave, along with the first attitudinal latent variable, are also allowed to affect missingness indicators from wave two onwards.

Results are obtained from WinBUGS for Bayesian estimation using MCMC.

See Appendix C for WinBUGS code and initial values. The first 4000 iterations have been discarded as a burn-in period, as suggested in WinBUGS. Two chains have been run for 10000 iterations when convergence has been attained according to Brooks and Gelman (1998) multivariate diagnostic (MPSRF is estimated by 1.04), and Gelman and Rubin (1992) PSRF for each parameter individually. All univariate PSRFs are $\leqslant$ 1.03, which is taken as an indication of convergence. Convergence diagnostics were obtained from CODA package in R.

All trace plots are of the form shown in the top graph of Figure 6.2, depicting convergence for all parameters. Figure 6.3 gives a sample of trace plots for selected parameters, where some of the plots (left) exhibit very well mixing of chains while others (right) are not mixing as well but can still be considered as reasonable evidence of convergence. In general, parameters of the dropout model, including intercepts α_{dt}, dropout coefficient γ and covariates effects $\boldsymbol{\beta}_d$ on dropout indicators, are mixing very well. Parameters of measurement and structural models such as item intercepts α_i, loadings λ_i, autoregressive parameter ϕ and covariates effects $\boldsymbol{\beta}_t$ on attitudes exhibit reasonable mixing. Some of the standard deviations are mixing very well too, while others are not as good, but still acceptable as evidence of convergence. Posterior densities for all parameters look reasonably normal.

Figure 6.3: Trace plots for a sample of parameters (intercepts, loadings, regression coefficients and variances): (left) very well mixing of chains, (right) reasonable mixing of chains, attitudes towards women's work data subject to dropout

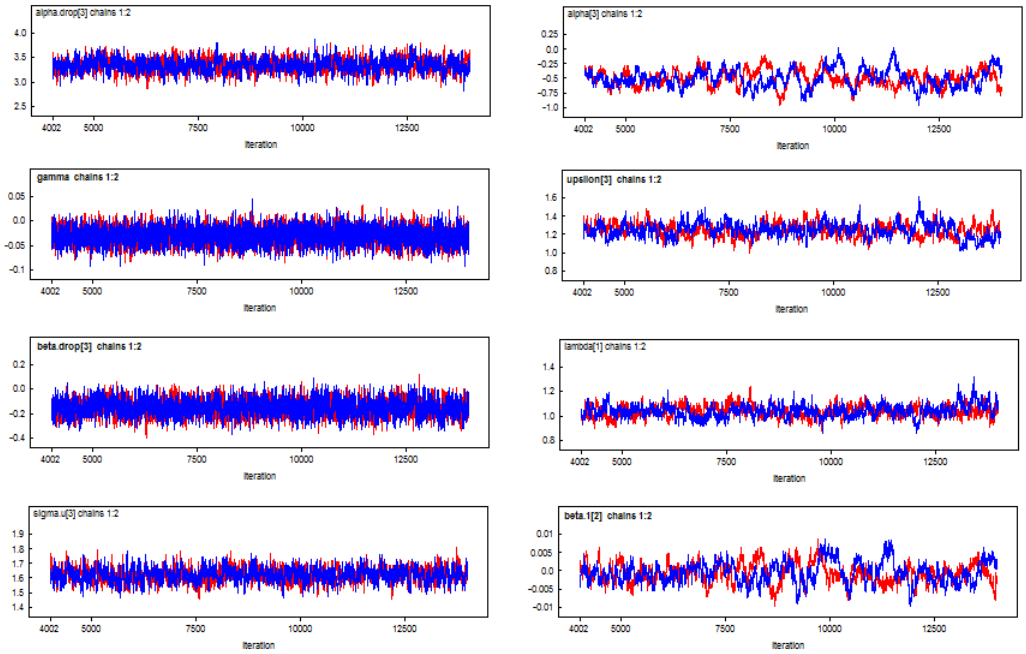


Table 6.1 shows results of the measurement and structural parts of the model, in addition to covariates effects. The first column of Table 6.1 gives parameter estimates, while the second gives their corresponding estimated standard errors. The last column gives Gelman and Rubin (1992) PSRF for each parameter. The relatively high positive estimated coefficient for the difficulty parameter α_1 of the first item [Family] indicates a high probability of a positive response to this item for an individual with a median score on the latent variable scale, given that the item-specific random effect is zero. The autoregressive parameter ϕ is estimated by 0.854 with estimated standard error of 0.010, indicating that people's perceptions about women's work are highly associated with their views on the previous wave. The unexplained variation of attitude at first wave, given by variance υ_1^2 of its

error term, is significantly higher than that of subsequent waves $\upsilon_2^2, \ldots, \upsilon_5^2$. This is an expected result since attitude at first wave is only explained by covariates, and there is no previous attitude to explain it as in case of subsequent waves.

The estimated regression coefficients of covariates on attitude at first wave , and their corresponding standard errors, show that education and employment status have a significant effect on attitude at first wave. The positive coefficient for education indicates that, people with a medium or high academic qualification tend to have more conservative views about women's work than those with no academic qualification at the beginning of the study, controlling for other covariates. The negative coefficient for occupational status indicates that those who are employed tend to score lower on the latent variable scale indicating more liberal views about women's work than those who are unemployed, controlling for other covariates. The effects of sex and age on attitude towards women's work at the first wave seem to be insignificant. From the second wave onwards, sex still has no significant effect on attitudes towards women's work, while age at the beginning of the study starts to have a negative significant effect indicating that older people have more liberal views on women's work, controlling for other covariates. Education and occupational status still have the same sort of effect on time-varying attitudes, in terms of significance and direction, as they do on attitude at the first wave.

Table 6.2 gives estimated parameters, standard errors and PSRF for the dropout model defined by equation (6.5). Having accounted for covariates, the estimated coefficient $\gamma_{(drop)}$ for the effect of first attitudinal latent variable z_{a_1} on the probability of dropout on subsequent waves is significant (at 10% level of significance), indicating that the data is subject to informative dropout, given the model specification and assumptions. The negative coefficient $\gamma_{(drop)} = -0.031$ indicates that

Table 6.1: Parameter estimates, standard errors and PSRF from MCMC after 10000 iterations for a model where attitude and covariates at first wave affect probability of dropout; attitudes towards women's work data subject to dropout

		MCMC mean		MCMC s.d.	PSRF
		Measurement model			
Family	α_1	2.861	***	(0.196)	1.02
Contribution	α_2	-0.306		(0.196)	1.02
Independent	α_3	0.503	***	(0.147)	1.01
Family	λ_1	1		—	—
Contribution	λ_2	1.038	***	(0.049)	1.01
Independent	λ_3	0.776	***	(0.035)	1.02
		Structural model			
Autoregressive parameter	ϕ	0.854	***	(0.010)	1.01
Constants	a_2	0.612	***	(0.084)	1.01
	a_3	0.393	***	(0.084)	1.00
	a_4	0.410	***	(0.083)	1.00
	a_5	0.344	***	(0.086)	1.00
Standard deviations	υ_1	2.297	***	(0.086)	1.03
of errors $\delta_1, ..., \delta_5$	υ_2	1.490	***	(0.078)	1.02
	υ_3	1.241	***	(0.073)	1.01
	υ_4	1.247	***	(0.075)	1.00
	υ_5	1.230	***	(0.083)	1.01
		Random effects			
Standard deviations	σ_{u_1}	1.347	***	(0.076)	1.00
of random effects	σ_{u_2}	1.787	***	(0.063)	1.01
	σ_{u_3}	1.622	***	(0.047)	1.00
		Covariates effects on z_{a_1}			
Sex	β_{sex}	-0.103		(0.081)	1.00
Age	β_{age}	-0.0006		(0.003)	1.01
Education	β_{edu}	0.966	***	(0.105)	1.02
Employment	β_{emp}	-0.490	***	(0.111)	1.03
		Covariates effects on $z_{a_2}, \ldots, z_{a_5}$			
Sex	β_{sex}	-0.005		(0.027)	1.00
Age	β_{age}	-0.003	***	(0.0009)	1.01
Education	β_{edu}	0.132	***	(0.033)	1.00
Employment	β_{emp}	-0.206	***	(0.045)	1.00

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

the more conservative an individual is (the higher the score on the latent variable scale), the less likely he/she is to drop out.

Table 6.2: Parameter estimates, standard errors and PSRF from MCMC after 10000 iterations for missingness mechanism in a model where attitude and covariates at first wave affect probability of dropout, attitudes towards women's work data subject to dropout

	MCMC mean		MCMC s.d.	PSRF
	Intercepts for dropout model			
$\alpha_{2(drop)}$	-2.829	***	(0.123)	1.00
$\alpha_{3(drop)}$	-3.351	***	(0.128)	1.00
$\alpha_{4(drop)}$	-3.339	***	(0.128)	1.00
$\alpha_{5(drop)}$	-3.270	***	(0.128)	1.00
	Effect of z_{a_1} on probability of dropout			
$\gamma_{(drop)}$	-0.031	*	(0.016)	1.00
	Covariates effects on probability of dropout			
$\beta_{sex\,(drop)}$	-0.193	***	(0.055)	1.00
$\beta_{age\,(drop)}$	0.021	***	(0.002)	1.00
$\beta_{edu\,(drop)}$	-0.151	**	(0.063)	1.00
$\beta_{emp\,(drop)}$	-0.234	***	(0.069)	1.00

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

Covariates measured at first wave, are assumed to affect the probability of dropping out at any subsequent wave. The effect is assumed to be the same over time. All four covariates (sex, initial age, initial educational attainment and initial occupational status) have a significant effect on the probability of dropping out at any time point, starting the second wave. Their corresponding coefficients indicate that younger, more educated and employed females are less likely to drop out.

Although results of the model presented in this chapter are not directly comparable to those of Chapter 5, as ordinal items have been dichotomised and modelling is within an IRT approach rather than SEM with the corresponding assumptions. Bayesian estimation is used here, while WLS was used to fit the models in Chapter 5. It is worth mentioning though that results are not exactly the same when it comes to which covariates are significant in the structural and dropout models.

Chapter 7

Non-Monotone Missingness

7.1 Introduction

Missing data in a longitudinal study can be classified into intermittent missingness and dropout. Dropout occurs when a subject exits the study before it comes to an end, resulting in a monotone pattern of missingness (Little and Rubin (2002)). Intermittent missingness occurs when a subject misses one or more waves of the study but shows up on a subsequent wave. This type of missingness is referred to as a non-monotone pattern of missingness. Reasons for the two types of missingness may be different. Thus, one form of missingness may be informative, while the other is not and vice versa. In this chapter, we extend the latent variable model introduced in Chapter 6 to accommodate intermittent missingness, along with dropout, in the study of multivariate longitudinal data. Item non-response is not considered here. If a respondent is observed at any wave, he/she is assumed to give full answers to all items.

Compared to the vast amount of literature on dropout in longitudinal studies, little has been written on non-monotone missingness. Troxel et al. (1998) extend selection models proposed by Diggle and Kenward (1994) for the analysis of longitudinal data subject to informative dropout, to accommodate non-monotone patterns of missingness. Their model assumes multivariate normality for continuous outcomes and allows the probability of missingness to depend on current, possibly unobserved values. A full likelihood method is used and is reported to suffer from computational difficulties when analysing more than three or four occasions. The authors attempt to overcome these difficulties in their further research (Troxel et al. (1998)) by employing a pseudo-likelihood method which reduces the multiple integration into a single dimension, and using a product of marginal likelihoods at each time point assuming independence over time.

Lin et al. (2004) propose a latent pattern-mixture model, where instead of assuming predefined patterns of missing data as in an ordinary pattern-mixture model, latent classes are used to discover joint patterns of missing data and longitudinal responses from the data itself. Their model is also developed for continuous outcomes. They assume the missingness process to be conditionally independent of the longitudinal outcomes given the latent classes. Class membership is modelled via a multinomial regression with covariates affecting the probability of belonging to a specific class. Each latent class has its own model for the continuous longitudinal outcome of interest which is represented by a linear mixed model and its own pattern of visits described by a multiplicative intensity model. A semi-parametric maximum likelihood method is used for estimation of the model parameters.

Shared parameter models, where random effects are shared both by the measurement process and the missingness mechanism, have also been used in the literature to model longitudinal data with non-monotone patterns of missingness.

Follmann and Wu (1995) provide a conditional approximation to shared random effects models for binary repeated outcome where the outcome, conditional on the random effect, follows a generalised linear model. The generalised linear model is approximated by conditioning on the missingness variables. They consider missingness variables that allow for non-monotone patterns of missingness. Minini and Chavance (2004) introduce a sensitivity parameter to represent the relationship between the measurement process and the missingness mechanism in the analysis of longitudinal binary outcomes with non-monotone patterns of missingness. Albert et al. (2002) develop a model for longitudinal binary data in which a Gaussian autoregressive latent process, rather than a random effect, is shared between the response and missing data mechanism. The binary response is modelled by a logit link as a function of covariates, conditional on the latent process. A three-state missingness variable representing whether a subject is observed, intermittent or a dropout is modelled, conditional on the latent process, via multinomial regression as a function of covariates. The shared latent process affects the probability of a positive response, the probability of an intermittent missed value and the probability of dropout thus relating the response to the missingness mechanism while allowing intermittent missingness, dropout or both to be informative. A Monte Carlo EM algorithm is used for maximum likelihood estimation of the proposed model.

The model developed in this chapter can be classified under shared parameter models where the shared parameter is a latent variable affecting both the observed variables and the missingness mechanism. However, unlike all models presented in the reviewed literature; our model is developed for a number of variables that are repeatedly measured over time rather than a single longitudinal outcome. It uses the same idea of a three-state missingness variable as the one presented in Albert et al. (2002). Items considered in this chapter are binary items analysed under item response theory.

7.2 A Latent Variable Model for Multivariate Binary Longitudinal Data Subject to Intermittent Missingness and Dropout

The model presented in this chapter is similar in structure to that presented in Chapter 6. However, it allows for intermittent missingness besides dropout in the analysis of multivariate longitudinal data. Observed items considered here are all binary, where item response theory is adopted for fitting the model. For simplicity, it is assumed that at each time point, a single continuous latent variable is sufficient to explain associations among the multiple observed binary items. Item-specific random effects are included to account for repetition of items over time. For incorporating intermittent missingness and dropout in the model, a three-state missingness variable is defined at each time point indicating whether an individual is observed, missing intermittently or has dropped out of the study. Covariates are allowed to affect both attitudes and missingness indicators.

The measurement part, in which the observed variables are explained by a latent variable at each time point in addition to random effects, and the structural part which defines relationships among latent variables over time are the same as outlined in Chapter 6 (see Sections 6.1 and 6.2). In the next section, we introduce the missingness mechanism and link it to the latent variable model via two possible specifications.

7.2.1 Missingness Mechanism

The missingness mechanism builds on the idea of a three-state nominal missingness indicator introduced by Albert et al. (2002). Let $\mathbf{r}_m = (r_{m1}, r_{m2}, ..., r_{mT})$ denote a $(T \times 1)$ vector of missingness indicators for a subject m, where a missingness indicator r_{mt} is created at each time point t; $t = 1, \ldots, T$, indicating whether a subject m is observed, missing intermittently or has dropped out:

$$r_{mt} = \begin{cases} 0, & \textit{observed}, \\ 1, & \textit{intermittent}, \\ 2, & \textit{dropout}. \end{cases} \tag{7.1}$$

It is assumed that all subjects are observed at the first wave $t = 1$, thus $r_{m1} = 0$ for all subjects m. It is also assumed that $r_{mt-1} = 2$ is an absorbing state, implying that $\Pr(r_{mt} = 2 \mid r_{mt-1} = 2) = 1$ (Albert et al. (2002)). For the last time point T, intermittent missingness is not an option as subjects can not show up on further occasions. Missingness on the last wave is thus considered as dropout, allowing for two possible values for r_{mT}; namely 0 or 2. The way the missingness indicators r_{mt} are constructed implies that they can only be recorded when the study comes to an end, thus making it possible to distinguish between intermittent missingness and dropouts.

Modelling The Missingness Indicators

When considering how to model the three-state nominal missingness indicators, a natural choice is to use the second model specification presented in Chapter 4 for modelling dropout indicators. The missingness indicators r_{mt} are assumed to measure a single continuous latent variable z_r which summarises an individual's *response propensity*. This is a real identifiable latent variable with an estimable variance, unlike the one that was used in case of dropout only, where all dropout indicators represented basically one variable; *time of dropout*. Such a specification would be presented by a path diagram like the one shown in Figure 4.2 with the main difference that the missingness indicators are as defined by equation (7.1). The drawback for such a model is that only one parameter γ can be estimated in the regression of the missingness latent variable z_r on the first attitudinal latent variable z_{a_1}, making it impossible to distinguish between the attitude's effect on intermittent missingness and dropout. This specification is therefore not considered.

An alternative specification is to drop the *response propensity* latent variable, and to model the missingness indicators directly as a function of attitudes and covariates. Suppressing the index m for an individual, the missingness indicators are modelled, conditional on the first attitudinal latent variable z_{a_1}, covariates $\mathbf{x}_1$, and that $r_{t-1} \neq 2$ for $t \geqslant 2$ via a three-state multinomial regression as,

$$p_{\ell t} = \Pr\left(r_t = \ell \mid z_{a_1}, \mathbf{x}_1, r_{t-1} \neq 2\right) = \begin{cases} \frac{1}{1+\sum_{\ell=1}^{2}\exp(\alpha_{rt\ell}+\gamma_\ell z_{a_1}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_1)}, & \ell = 0, \\ \frac{\exp(\alpha_{rt\ell}+\gamma_\ell z_{a_1}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_1)}{1+\sum_{\ell=1}^{2}\exp(\alpha_{rt\ell}+\gamma_\ell z_{a_1}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_1)}, & \ell = 1, 2, \end{cases} \tag{7.2}$$

where $\boldsymbol{\beta}_{r_\ell}$ is a vector of regression coefficients representing dependence of missingness indicators on covariates $\mathbf{x}_1$, and γ_ℓ are regression coefficients relating missingness (intermittent and dropout) to the first attitudinal latent variable z_{a_1}, thus allowing for intermittent missingness, dropout or both to be informative in case of their significance. The attitude at first wave z_{a_1} serves as a shared parameter in this context by affecting both the response process and the missingness mechanism. It is assumed that an individual's attitude at the beginning of the study z_{a_1}, affects the probability of a positive response to any of the observed items y_{i1} at time 1, and it also affects attitude on the next wave via an autoregressive parameter ϕ. On the missingness part of the model, attitude at first wave z_{a_1} affects the probability to miss one or more subsequent waves, or to drop out of the study completely at any time t. Figure 6.1 gives an illustration of this model by a path diagram for an example with four time points, if dropout indicators d_t are replaced by missingness indicators r_t.

A slight modification to the above setting results in a specification that classifies this model under selection models. The probability of an intermittent missed value or a dropout at time t, is made to depend on the unobserved attitude z_{a_t}, measured at the same wave. This means that, the probability of an intermittent missed observation or dropout will depend on the missing values themselves through the latent variable. Attitudinal latent variables z_{a_t} are shared normal latent variables that affect the probability of a positive response to each of the binary observed items, the probability of an intermittently missed observation and the probability of a dropout; at time t, and are themselves linked via a first-order autoregressive structure. The multinomial regression modelling the probability of being observed, missing intermittently or dropping out becomes

$$p_{\ell t} = \Pr\left(r_t = \ell \mid z_{a_t}, \mathbf{x}_t, r_{t-1} \neq 2\right) = \begin{cases} \frac{1}{1+\sum_{\ell=1}^{2} \exp(\alpha_{rt\ell}+\gamma_\ell z_{a_t}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_t)}, & \ell = 0, \\ \frac{\exp(\alpha_{rt\ell}+\gamma_\ell z_{a_t}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_t)}{1+\sum_{\ell=1}^{2} \exp(\alpha_{rt\ell}+\gamma_\ell z_{a_t}+\boldsymbol{\beta}'_{r_\ell}\mathbf{x}_t)}, & \ell = 1, 2, \end{cases} \tag{7.3}$$

where $\boldsymbol{\beta}_{r_\ell}$ is a vector of regression coefficients representing dependence of missingness indicators on covariates $\mathbf{x}_t$, and γ_ℓ are regression coefficients representing the dependence of missingness (intermittent and dropout) on unobserved corresponding attitudes z_{a_t}, thus allowing for intermittent missingness, dropout or both to be informative in case of their significance. In this formulation, the probability of intermittent missingness or dropping out at a given time point depends on the value of the latent attitude variable at the corresponding time point, and covariates. The effect of attitudes and covariates on each type of missingness is assumed to be the same over time. Figure 7.1 gives an illustration for this model specification by a path diagram for an example with four time points.

A special case of this specification is obtained by employing a binary variable for dropout instead of a three-state missingness indicator resulting in a selection model, where time-dependent unobserved attitudes are shared between the measurement and dropout processes. This model specification, whether considering dropout only or both dropout and intermittent missingness, provides a richer, more dynamic structure than the specifications considered in Chapter 4 where only attitude at first wave or a lagged effect of the attitude is allowed to affect the missingness mechanism. It is particularly useful in cases when the nature of the attitude of interest is prone to much change over time or when the time difference between waves is big allowing various factors to change between time points. A good example would be in psychiatric studies, when the latent variable of interest is measuring a mood or mental state for subjects. In this case, the state of the subject/patient at a certain occasion may affect whether or not they are ready to participate in that particular wave of the study or in any subsequent waves thus creating informative intermittent missingness or dropout that depends on the unobserved attitude of interest at the same time point.

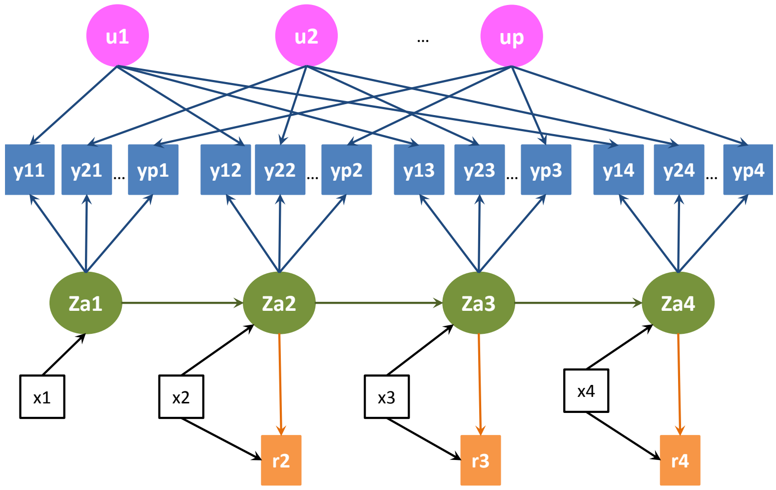


Figure 7.1: Path diagram for a model where the time-dependent attitude affects missingness on the same wave, an example with four time points

7.3 Estimation

Let $\boldsymbol{\theta}$ denote a vector of all parameters defining the model outlined by equations (6.1, 6.3, 6.4 and 7.2/7.3), including α_i and λ_i defining the measurement process; a_t, ϕ, $\boldsymbol{\beta}_1'$ and $\boldsymbol{\beta}_t'$ defining the structural model; and $\alpha_{rt\ell}, \gamma_\ell$ and $\boldsymbol{\beta}_{\boldsymbol{r}_\ell}'$ defining the missingness mechanism, in addition to variances of errors υ_t^2 and random effects $\sigma_{u_i}^2$. For a random sample of size n, the marginal likelihood of the observed data is given by

$$L(\boldsymbol{\theta}) = \prod_{m=1}^{n} \int_{\mathbf{z}_a} \int_{\mathbf{u}} \left\{ \prod_{i=1}^{p} \prod_{t \in \mathcal{H}_m^Y} \pi_{mit}(z_{a_t}, u_i)^{y_{mit}} (1 - \pi_{mit}(z_{a_t}, u_i))^{1-y_{mit}} \right\}$$

$$\times \left\{ \prod_{t \in \mathcal{H}_m} p_{1mt}(z_{a_t}, \mathbf{x}_t)^{I(r_{mt}=1)} p_{2mt}(z_{a_t}, \mathbf{x}_t)^{I(r_{mt}=2)} (1 - p_{1mt}(z_{a_t}, \mathbf{x}_t) - p_{2mt}(z_{a_t}, \mathbf{x}_t))^{I(r_{mt}=0)} \right\}$$

$$\times h(\mathbf{z}_a, \mathbf{u}) d\mathbf{z}_a d\mathbf{u}, \qquad (7.4)$$

where $\mathcal{H}_m^Y$ is the set of time points prior to dropout for an individual m where $\mathbf{y}_{mt}$ is observed, and $\mathcal{H}_m$ is the set of time points prior to and including time of dropout for an individual m. Moreover, π_{mit} is the probability that an individual m gives a positive response to item i at time t conditional on the attitudinal latent variable z_{a_t} and random effect u_i given by equation (6.2), p_{1mt}, p_{2mt} are the respective probabilities of intermittent missingness and dropout conditional on the attitudinal latent variable z_{a_t} and covariates $\mathbf{x}_t$ as expressed by equation (7.3), $I(r_{mt} = \ell)$ is an indicator function that takes the value 1 if $r_{mt} = \ell$ and 0 otherwise, and $h(\mathbf{z}_a, \mathbf{u})$ is the joint distribution of attitude latent variables and random effects.

The loglikelihood is thus

$$\log L(\boldsymbol{\theta}) = \sum_{m=1}^{n} \left[\int_{\mathbf{z}_z} \int_{\mathbf{u}} \left\{ \sum_{i=1}^{p} \sum_{t \in \mathcal{H}_m^Y} \Big(y_{mit} \log \pi_{mit}(z_{a_t}, u_i) + (1 - y_{mit}) \log (1 - \pi_{mit}(z_{a_t}, u_i)) \Big) \right. \right.$$

$$+ \sum_{t \in \mathcal{H}_m} \Big(I(r_{mt} = 1) \log p_{1mt}(z_{a_t}, \mathbf{x}_t) + I(r_{mt} = 2) \log p_{2mt}(z_{a_t}, \mathbf{x}_t)$$

$$+ I(r_{mt} = 0) \log (1 - p_{1mt}(z_{a_t}, \mathbf{x}_t) - p_{2mt}(z_{a_t}, \mathbf{x}_t)) \Big)$$

$$+\log h(\mathbf{z}_a, \mathbf{u}) \Big\} d\mathbf{z}_a d\mathbf{u} \Bigg]. \tag{7.5}$$

The above expression requires a $(T+p)-$dimensional integration, which makes its evaluation complicated especially as the number of waves T increases, thus making estimation using MCMC an appealing option. The same main steps for Bayesian estimation using MCMC, choosing prior distributions and assessing convergence, outlined in Section 6.4, are employed here.

7.4 Application

The model presented in this chapter is now applied to the data on people's attitudes towards women's work, introduced in Chapter 5. The same five waves of the British Household Panel Survey (1993, 95, 97, 99, 2001) are considered. However, it is not exactly the same dataset used in Chapter 5 in more than one aspect. Although we still employ the same three items that have been used as measures of attitude towards women's work, the originally five-category ordinal variables have been dichotomised for the sake of the analysis into binary items -as in Chapter 6- taking one of two possible values: 0 if an individual *Strongly Agrees* or *Agrees* to an item, and 1 if his/her response is *Don't Know*, *Disagree* or *Strongly Disagree*.

The sample size is different too, as we now include cases that experienced intermittent missingness. Those were previously discarded from the analysis when only dropout was considered. There are 7622 complete cases, who gave full answers to the three items of interest on the first wave considered in the analysis. A subject may miss one or more waves intermittently, or drop out of the study starting the second wave. If a subject misses the last wave, this is considered to be a dropout, since then the study has come to an end. Having observed the five waves, there are 2145 cases who have dropped out by the end of the study and 451 cases who have occasianally missed a wave (intermittent missingness).

Three time-invariant covariates (sex as a dummy variable for women, age at first wave and initial educational attainment) and one time-varying covariate (occupational status) are allowed to affect the attitude towards women's work at each wave. Details about these covariates can be found in Chapter 5. The effect of covariates on the corresponding attitudes is constrained to be the same from the second wave onwards. The same four covariates are also allowed to affect missingness indicators.

Next, we present results for a model where attitude and covariates measured at the first wave are allowed to affect the missingness mechanism, followed by results for a model where missingness is affected by time-dependent attitudes and covariates. The interpretation of the results depends on the data set and the distributional assumptions of each model, and hence cannot be generalised.

7.4.1 A Specification where Attitude Measured at First Wave Is Allowed to Affect The Missingness Mechanism

In this section, we fit a model where the first attitudinal latent variable, and covariates measured at first wave, are allowed to affect the missingness indicators from wave two onwards. This model is defined by equations (6.1, 6.3, 6.4 and 7.2), and is illustrated in Figure 6.1. Results are obtained from WinBUGS for Bayesian estimation using MCMC. See Appendix D for WinBUGS code and initial values.

The first 4000 iterations have been discarded as a burn-in period, as suggested in WinBUGS. Two chains have been run for 10000 iterations when convergence has been attained according to Brooks and Gelman (1998) multivariate diagnostic estimated by 1.08, and Gelman and Rubin (1992) PSRF for each parameter individually (each$\leqslant$ 1.05). All trace plots are of the form shown in the top graph of Figure 6.2, depicting convergence for all parameters. Figure 7.2 gives a sample of trace plots for selected parameters, where some of the plots (left) exhibit very well mixing of chains while others (right) are not mixing as well but can still be considered as reasonable evidence of convergence. Posterior densities for all parameters look reasonably normal. Convergence diagnostics were obtained from CODA package in R.

Figure 7.2: Trace plots for a sample of parameters (intercepts, regression coefficients and variances): (left) very well mixing of chains, (right) reasonable mixing of chains, attitudes towards women's work data subject to intermittent missingness and dropout

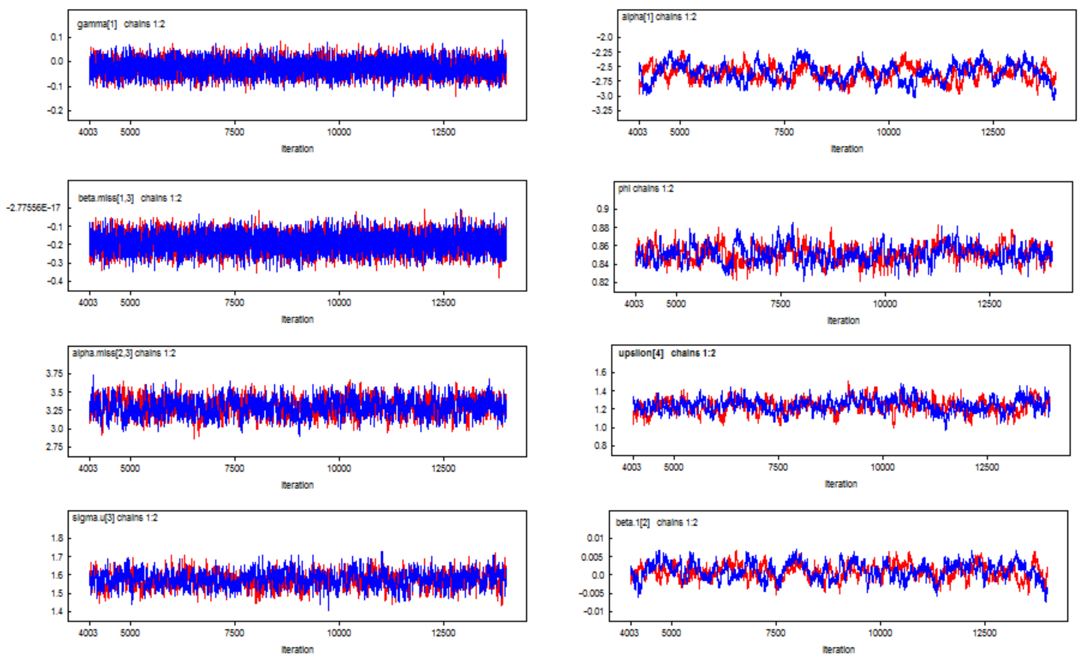


Table 7.1: Parameter estimates, standard errors and PSRF from MCMC after 10000 iterations for a model where attitude at first wave affects missingness; attitudes towards women's work data subject to intermittent missingness and dropout

	MCMC mean	MCMC s.d.	PSRF
	Measurement model		
α_1	2.608***	(0.142)	1.04
α_2	-0.548***	(0.154)	1.05
α_3	0.321***	(0.112)	1.04
λ_1	1	—	—
λ_2	1.084***	(0.045)	1.01
λ_3	0.790***	(0.030)	1.01
	Structural model		
ϕ	0.851***	(0.009)	1.00
a_2	0.617***	(0.077)	1.01
a_3	0.410***	(0.075)	1.02
a_4	0.426***	(0.075)	1.02
a_5	0.376***	(0.077)	1.02
υ_1	2.237***	(0.073)	1.01
υ_2	1.437***	(0.067)	1.01
υ_3	1.235***	(0.062)	1.01
υ_4	1.232***	(0.063)	1.02
υ_5	1.236***	(0.073)	1.00
	Random effects		
σ_{u_1}	1.288***	(0.058)	1.01
σ_{u_2}	1.784***	(0.054)	1.00
σ_{u_3}	1.573***	(0.042)	1.01
	Covariates effects on z_{a_1}		
β_{sex}	-0.050	(0.070)	1.00
β_{age}	0.001	(0.002)	1.01
β_{edu}	0.993***	(0.084)	1.00
β_{emp}	-0.380***	(0.088)	1.03
	Covariates effects on $z_{a_2}, \ldots, z_{a_5}$		
β_{sex}	-0.023	(0.024)	1.00
β_{age}	-0.003***	(0.0009)	1.01
β_{edu}	0.135***	(0.031)	1.00
β_{emp}	-0.191***	(0.041)	1.00

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

Results of the measurement and structural parts of the model, given in Table 7.1, are similar to those obtained for the model with dropout only in Chapter 6. The same comments given on Table 6.1 can be given here, too (see Section 6.5).

Table 7.2: Parameter estimates, standard errors and PSRF from MCMC after 10000 iterations for missingness mechanism in a model where attitude at first wave affects missingness, attitudes towards women's work data subject to intermittent missingness and dropout

	MCMC mean	MCMC s.d.	PSRF		MCMC mean	MCMC s.d.	PSRF
	Intercepts for intermittent				Intercepts for dropout		
$\alpha_{2(int)}$	-2.369***	(0.202)	1.00	$\alpha_{2(drop)}$	-2.750***	(0.108)	1.00
$\alpha_{3(int)}$	-2.693***	(0.207)	1.00	$\alpha_{3(drop)}$	-3.296***	(0.113)	1.00
$\alpha_{4(int)}$	-2.788***	(0.210)	1.00	$\alpha_{4(drop)}$	-3.106***	(0.112)	1.00
$\alpha_{5(int)}$	—	—	—	$\alpha_{5(drop)}$	-2.747***	(0.108)	1.00
	Effect of z_{a_1} on probability of intermittent				Effect of z_{a_1} on probability of dropout		
$\gamma_{(int)}$	-0.023	(0.029)	1.00	$\gamma_{(drop)}$	-0.024*	(0.014)	1.00
	Covariates effects on probability of intermittent				Covariates effects on probability of dropout		
$\beta_{sex\,(int)}$	-0.207**	(0.091)	1.00	$\beta_{sex\,(drop)}$	-0.190***	(0.046)	1.00
$\beta_{age\,(int)}$	-0.013***	(0.003)	1.00	$\beta_{age\,(drop)}$	0.018***	(0.002)	1.00
$\beta_{edu\,(int)}$	-0.396***	(0.111)	1.00	$\beta_{edu\,(drop)}$	-0.173***	(0.055)	1.00
$\beta_{emp\,(int)}$	-0.133	(0.112)	1.00	$\beta_{emp\,(drop)}$	-0.160***	(0.059)	1.00

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

Table 7.2 gives estimated parameters, standard errors and PSRF for the multinomial missingness model defined by equation (7.2). The table is divided into two parts; the left part gives estimates for the intermittent missingness branch of the model while the right part gives the corresponding estimates for the dropout branch. Although the estimated coefficients for the effect of the first attitudinal latent variable z_{a_1} on the probability of intermittent missingness $\gamma_{(int)}$ and dropout $\gamma_{(drop)}$ on subsequent waves is almost equal, however it is only significant for dropout (at 10% level of significance). This indicates that dropout is informative, while intermittent missingness is not according to this model specification. This is a logical result in such a study. As one might expect, missing one wave in such a longitudinal study is likely to be due to random causes such as illness or travelling, while dropout is more likely to be related to a specific attitude or behaviour. The negative coefficient $\gamma_{(drop)} = -0.024$ indicates that the more conservative an individual is (the higher the score on the latent variable scale), the less likely he/she is to drop out.

Covariates measured at first wave, are assumed to affect the probability of missing intermittently and the probability of dropping out at any subsequent wave. The effect on each probability is assumed to be the same over time. Sex, initial age and initial educational attainment have a significant negative effect on the probability of having an intermittent missingness on waves two, three and four. Older, more educated females tend to have a lower probability of missing intermittently. Thus, intermittent missingness can be considered MAR, as it depends on covariates but not on unobserved attitudes, given the model specification and data.

All four covariates (sex, initial age, initial educational attainment and initial occupational status) have a significant effect on the probability of dropping out at any time point, starting the second wave. Their corresponding coefficients indicate that younger, more educated and employed females are less likely to drop out. It is noted that the effect of age is now reversed. While younger respondents are more likely to miss an occasion intermittently, older ones are more likely to drop out of the study completely.

7.4.2 A Specification where Time-Dependent Attitudes Affect The Missingness Mechanism

This section presents results for the second model specification defined by equations (6.1, 6.3, 6.4 and 7.3) and illustrated in Figure 7.1, where time-dependent attitudinal latent variables and covariates, are allowed to affect the probability of missingness (intermittent or dropout) at the same wave. Again, the first 4000 iterations have been discarded as a burn-in period, as suggested by WinBUGS. The richer structure implied by this model required more than twice the computational time, taken by the first model specification, for each iteration. However, convergence was diagnosed earlier. Two chains have been run for another 4000 iterations after the burn-in period, when convergence has been attained according to Brooks and Gelman (1998) multivariate diagnostic (MPSRF=1.08), and Gelman and Rubin (1992) PSRF for each parameter individually (all PSRFs < 1.05). All trace plots (not shown here) are of the form shown in the top graph of Figure 6.2, depicting convergence for all parameters. Posterior densities for all parameters look reasonably normal. See Appendix E for WinBUGS code and initial values.

Table 7.3 shows results of the measurement and structural parts of the model, in addition to covariates effects. The estimated parameters and corresponding standard errors, for the measurement and structural models, seem to be almost unchanged from those obtained in the first model specification. The effects of covariates on attitudes are the same as before, too.

Table 7.4 gives estimated parameters, standard errors and PSRF for the multinomial missingness model defined by equation (7.3). Three time-invariant covariates measured at first wave (sex, age at first wave and initial educational attain-

Table 7.3: Parameter estimates, standard errors and PSRF from MCMC after 4000 iterations for a model where time-dependent attitudes affect missingness; attitudes towards women's work data subject to intermittent missingness and dropout

	MCMC mean	MCMC s.d.	PSRF
	Measurement model		
α_1	2.632***	(0.156)	1.01
α_2	-0.529***	(0.161)	1.02
α_3	0.337***	(0.118)	1.01
λ_1	1	—	—
λ_2	1.080***	(0.042)	1.00
λ_3	0.786***	(0.029)	1.00
	Structural model		
ϕ	0.849***	(0.009)	1.01
a_2	0.608***	(0.073)	1.02
a_3	0.405***	(0.073)	1.02
a_4	0.423***	(0.074)	1.02
a_5	0.370***	(0.075)	1.02
υ_1	2.248***	(0.067)	1.00
υ_2	1.449***	(0.060)	1.02
υ_3	1.241***	(0.058)	1.01
υ_4	1.246***	(0.061)	1.00
υ_5	1.244***	(0.066)	1.04
	Random effects		
σ_{u_1}	1.291***	(0.057)	1.00
σ_{u_2}	1.786***	(0.056)	1.01
σ_{u_3}	1.575***	(0.042)	1.00
	Covariates effects on z_{a_1}		
β_{sex}	-0.055	(0.071)	1.01
β_{age}	0.0008	(0.002)	1.01
β_{edu}	0.988***	(0.082)	1.00
β_{emp}	-0.383***	(0.090)	1.00
	Covariates effects on $z_{a_2}, \ldots, z_{a_5}$		
β_{sex}	-0.023	(0.025)	1.00
β_{age}	-0.003***	(0.0008)	1.02
β_{edu}	0.137***	(0.028)	1.02
β_{emp}	-0.190***	(0.039)	1.01

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

ment), in addition to employment status which is a time-varying covariate, are assumed to affect the probability of missing intermittently and the probability of dropping out at each wave. The effects of those covariates are almost unchanged in terms of direction, magnitude and significance, from those estimated under the first model specification.

Table 7.4: Parameter estimates, standard errors and PSRF from MCMC after 4000 iterations for missingness mechanism in a model where time-dependent attitudes affect missingness, attitudes towards women's work data subject to intermittent missingness and dropout

	MCMC mean	MCMC s.d.	PSRF		MCMC mean	MCMC s.d.	PSRF
	Intercepts for intermittent				Intercepts for dropout		
$\alpha_{2(int)}$	-2.315***	(0.198)	1.00	$\alpha_{2(drop)}$	-2.624***	(0.108)	1.00
$\alpha_{3(int)}$	-2.630***	(0.205)	1.00	$\alpha_{3(drop)}$	-3.157***	(0.113)	1.00
$\alpha_{4(int)}$	-2.718***	(0.209)	1.00	$\alpha_{4(drop)}$	-2.960***	(0.111)	1.00
$\alpha_{5(int)}$	—	—	—	$\alpha_{5(drop)}$	-2.592***	(0.110)	1.00
	Effect of z_{a_t} on probability of intermittent				Effect of z_{a_t} on probability of dropout		
$\gamma_{(int)}$	-0.051*	(0.029)	1.00	$\gamma_{(drop)}$	-0.040***	(0.016)	1.00
	Coavariates effects on probability of intermittent				Coavariates effects on probability of dropout		
$\beta_{sex\,(int)}$	-0.213**	(0.089)	1.00	$\beta_{sex\,(drop)}$	-0.206***	(0.046)	1.00
$\beta_{age\,(int)}$	-0.013***	(0.003)	1.00	$\beta_{age\,(drop)}$	0.018***	(0.001)	1.00
$\beta_{edu\,(int)}$	-0.371***	(0.109)	1.00	$\beta_{edu\,(drop)}$	-0.146***	(0.054)	1.00
$\beta_{emp\,(int)}$	-0.175	(0.119)	1.00	$\beta_{emp\,(drop)}$	-0.331***	(0.063)	1.00

Note: *** indicates significance at 1% , ** indicates significance at 5% , while * indicates significance at 10%.

The only difference noted in the results of this model compared to those obtained previously from the first specification is the effect of the time-dependent attitudinal latent variables z_{a_t} on the corresponding probability of intermittent missingness and dropout. Both coefficients representing these effects, $\gamma_{(int)}$ for intermittent missingness and $\gamma_{(drop)}$ for dropout, are now significant. This is an indication that both intermittent missingness and dropout are informative due to this model specification. The negative coefficients for $\gamma_{(drop)}$ and $\gamma_{(int)}$ indicate that the more conservative an individual is (the higher the score on the latent variable scale), the less likely he/she is to drop out and to miss a wave intermittently. Again, this does not conclude that attitudes towards women's work are causing dropouts or intermittent missingness. However, it reflects some sort of association between them, which makes the incorporation of a missingness mechanism essential. The latter model specification where time-varying attitudes affect the missingness seems more plausible, as it is more realistic to assume that attitudes affect missingness at the same wave, in a study where there is a two-year difference between consecutive waves resulting in various possible changes over time making the effect of an attitude at time t different from that at the beginning of the study.

Chapter 8

Contribution, Limitations and Future Research

8.1 Summary and Contribution

Chapter 1 provides an introduction to three main topics that constitute the models developed in this thesis for analysing multivariate longitudinal data subject to different forms of missingness. These are: latent variable models, analysis of longitudinal data and missing data. The three topics have been studied thoroughly in the literature, but this is the first work to integrate them together in a unified model.

A review of literature on latent variable models is given in Chapter 2, outlining the two main approaches for estimating latent variable models with categorical data. These are the underlying variable approach which regards categorical variables as manifestations of underlying continuous variables, and the item response theory (IRT) approach where distributional assumptions are directly made on categorical manifest variables. Section 2.4 discusses latent variable models for multivariate longitudinal data.

Chapter 3 first gives a literature review for missing data in cross-sectional studies followed by missing data, particularly dropout, in longitudinal studies. Selection models, pattern-mixture models and shared-parameter models are three different approaches, presented in Chapter 3, that deal with the problem of dropout in univariate longitudinal data.

In Chapter 4, two model specifications that incorporate dropout within a latent variable modelling framework to model multivariate longitudinal data, are proposed. Both model specifications allow for testing whether dropout depends on the variables of interest by modelling the probability of dropping out at a given wave as a function of the latent variables (in which case the dropout is nonrandom), and possibly covariates. The models presented are for ordinal observed variables and binary indicators for the dropout. Within a SEM framework, ordinal observed variables are modelled using underlying continuous variables and the classical factor analysis model, employing the three-step estimation procedure (thresholds, polychoric correlations, weighted least squares) as described in Jöreskog (1994, 2005). The dropout mechanism is modelled with a hazard function that may depend on the attitudinal latent variables and covariates. Two different ways of modelling relationships among the latent variables and the dropout mechanism are proposed and their advantages and disadvantages discussed. The proposed models remain within the standard framework of a general latent variable model for longitudinal data, and therefore estimation of model parameters and goodness-of-fit testing use conventional methods.

The developed models are applied in Chapter 5 to a real dataset about people's attitudes towards women's work from five waves of the British Household Panel Survey (BHPS). A sensitivity analysis is included to compare how a model that uses listwise deletion, another that ignores the dropout mechanism and a third that accounts for dropout, behave at four different levels of dropout. A model that incorporates the dropout mechanism is found to be better at detecting significant covariates, even at high levels of dropout.

In Chapter 6, a model for multivariate binary longitudinal data subject to possibly nonrandom dropout, is developed within an IRT approach. Bayesian estimation using MCMC is used to fit this model. Again, the dataset about people's attitudes towards women's work from five waves of the BHPS is analysed to illustrate this model, after the items have been dichotomised.

Chapter 7 extends the model developed for binary observed items within an IRT framework to accommodate intermittent missingness together with dropout. Two possible specifications are given for this model. An application of this model is also presented using the BHPS data.

In summary, we have developed a latent variable model to analyse multivariate longitudinal data subject to possibly nonrandom dropout and/or intermittent missingness. Latent variable models for multivariate longitudinal data, and missingness in univariate longitudinal data are present in the literature separately. Our proposed model incorporates the missingness mechanism within a latent variable model framework to account for missing data in the analysis of multivariate longitudinal data, under both SEM and IRT approaches. Different model specifications introduced in various parts of the thesis provide several modelling options. The choice of which model specification to adopt will usually depend on the application and the nature of the problem being studied. Interpretation of results will always depend on the model specification and the accompanying assumptions.

8.2 Limitations

Developing the models involved many decisions to be made about the modelling procedure: how to model the items, how to relate the latent variables, how to model dropout and why. Several model specifications were thus discussed with their possible applicability in different situations. The model specifications presented in this thesis are by no means exhaustive. Other specifications can be explored. The choice of fitting the first two model specifications within a SEM framework for ordinal items while fitting the other two specifications within an IRT approach for binary items does not restrict them to those frameworks in particular. For instance, a model with a lagged effect of attitudes on dropout can be fitted within an IRT approach for ordinal items.

Maximum likelihood for models with categorical items and continuous latent variables within a SEM framework can be implemented in MPlus with a maximum of four latent dimensions due to the increased computational capacity required for numerical integration. In our case, the number of time-dependent latent variables and item-specific random effects that required numerical integration was much higher, and thus ML was not feasible in MPlus, hence estimation using WLS.

As in most cases, some restrictions were due to the type of available data. The BHPS data set on attitudes towards women's work was used in all applications mainly for convenience. Finding suitable data sets was not a very easy task. That is why ordinal items were dichotomised for the sake of developing a model for binary items, as no suitable data set with binary items was found. The fact that the items were originally measured on a five point scale implied that the middle category had to be merged with either the *agree* or *disagree* sides. In our case, we chose to include the *don't know* with those who *disagree* or *strongly disagree*. This is quite a subjective decision, and it could be argued to include the middle category with those who *agree* instead.

Data was available for more waves, but measurement invariance did not hold beyond the waves considered here. Therefore, a decision was made to restrict the analysis to five waves.

Other covariates could be included in further analyses of this data set (e.g. socio-economic status). The employment covariate was included as a binary variable with students, retired and those out of labour market combined with those who are employed in one category against a category for the unemployed. Having a third category at least for those who are out of the labour force could be useful especially that it includes women who are looking after family/children, and who might as well have distinct views about women's work.

8.3 Future Work

Future research may extend the developed models to accommodate yet another form of missing data. That is item non-response, where at a given time point a respondent might give answers to some of the items but not the others. O'Muircheartaigh and Moustaki (1999) developed a latent variable model for cross-sectional data with item non-response with two latent dimensions, one to summarise the attitude and the other to summarise response propensity. For each observed variable, an indicator variable for responding is created, taking the value 1 if the individual responds and 0 if he or she does not respond. The attitude items are explained by the attitudinal latent variable, and the binary response items depend both on the attitudinal variable and the response propensity latent variable, thus allowing for non-ignorable missingness. Moustaki and Knott (2000b) include covariates to a similar model specification. This kind of models can be combined within our framework for modelling multivariate longitudinal data, in order to have a more general specification that accommodates dropout, intermittent missingness and item non-response.

Models presented within an IRT approach are developed for binary items, whose conditional distribution is assumed to be Bernoulli. Extensions to this model may accommodate nominal, ordinal, metric, or mixed types of items, with other possible conditional distributions of the exponential family.

Possible routines or reparameterisations to speed up the MCMC algorithm and improve mixing of chains could be investigated.

Appendix A

Mplus Code for Fitting First Model Specification (SEM Framework)

```
TITLE: Fitting first model specification of Chapter 4 to attitudes towards women's
work data (five waves of the BHPS)

DATA:
File is Woman_fully observed at wave C.dat;

VARIABLE:
Names are C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5;
usevar = C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5;
Categorical are C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5;
Missing are all (-1);

ANALYSIS:
Estimator=WLSMV;
Parameterization = delta;
!DIFFTEST=deriv15.dat; !to obtain chi-square difference test

MODEL:
!setting all loadings of random effects on items to one
u1 by C3@1 E3@1 G3@1 I3@1 K3@1;
u2 by C4@1 E4@1 G4@1 I4@1 K4@1;
u3 by C5@1 E5@1 G5@1 I5@1 K5@1;

!constraining loadings of same items to be equal over time
```

```
z1 by C3-C5 (1-3) d2;
z2 by E3-E5 (1-3) d3;
z3 by G3-G5 (1-3) d4;
z4 by I3-I5 (1-3) d5;
z5 by K3-K5 (1-3);

!constraining thresholds of same items to be equal over time
[C3$1 E3$1 G3$1 I3$1 K3$1 ] (4);
[C3$2 E3$2 G3$2 I3$2 K3$2 ] (5);
[C3$3 E3$3 G3$3 I3$3 K3$3 ] (6);
[C3$4 E3$4 G3$4 I3$4 K3$4 ] (7);

[C4$1 E4$1 G4$1 I4$1 K4$1 ] (8);
[C4$2 E4$2 G4$2 I4$2 K4$2 ] (9);
[C4$3 E4$3 G4$3 I4$3 K4$3 ] (10);
[C4$4 E4$4 G4$4 I4$4 K4$4 ] (11);

[C5$1 E5$1 G5$1 I5$1 K5$1 ] (12);
[C5$2 E5$2 G5$2 I5$2 K5$2 ] (13);
[C5$3 E5$3 G5$3 I5$3 K5$3 ] (14);
[C5$4 E5$4 G5$4 I5$4 K5$4 ] (15);

[d2$1 d3$1 d4$1 d5$1 ];

!allowing latent variables to correlate
z1 with z2-z5;
z2 with z3-z5;
z3 with z4-z5;
z4 with z5;

!estimate means of latent variables
[z2 z3 z4 z5];

!random effects independent of each other and of latent variables
u1 with u2-u3 @0;
u2 with u3 @0;
u1-u3 with z1-z5 @0;

OUTPUT:
tech4;

!Savedata:
!DIFFTEST=deriv15.dat; !to obtain chi-square difference test
```

Appendix B

Mplus Code for Fitting Second Model Specification (SEM Framework)

```
TITLE: Fitting second model specification of Chapter 4, with covariates, to atti-
tudes towards women's work data (five waves of the BHPS)

DATA:
File is Woman_fully observed at wave C_covariates.dat;

VARIABLE:
Names are C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5
sex ageC eduC empC empE empG empI empK;
usevar = C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5
sex ageC eduC empC empE empG empI empK;
Categorical are C3-C5 E3-E5 G3-G5 I3-I5 K3-K5 d2-d5;
Missing are all (-1);

ANALYSIS:
Estimator=WLSMV;
Parameterization = delta;

MODEL:
!setting all loadings of random effects on items to one
u1 by C3@1 E3@1 G3@1 I3@1 K3@1;
u2 by C4@1 E4@1 G4@1 I4@1 K4@1;
```

```
u3 by C5@1 E5@1 G5@1 I5@1 K5@1;

!constraining loadings of same items to be equal over time
z1 by C3-C5 (1-3);
z2 by E3-E5 (1-3);
z3 by G3-G5 (1-3);
z4 by I3-I5 (1-3);
z5 by K3-K5 (1-3);

zd by d2-d5@1;

!constraining thresholds of same items to be equal over time
[C3$1 E3$1 G3$1 I3$1 K3$1 ] (4);
[C3$2 E3$2 G3$2 I3$2 K3$2 ] (5);
[C3$3 E3$3 G3$3 I3$3 K3$3 ] (6);
[C3$4 E3$4 G3$4 I3$4 K3$4 ] (7);

[C4$1 E4$1 G4$1 I4$1 K4$1 ] (8);
[C4$2 E4$2 G4$2 I4$2 K4$2 ] (9);
[C4$3 E4$3 G4$3 I4$3 K4$3 ] (10);
[C4$4 E4$4 G4$4 I4$4 K4$4 ] (11);

[C5$1 E5$1 G5$1 I5$1 K5$1 ] (12);
[C5$2 E5$2 G5$2 I5$2 K5$2 ] (13);
[C5$3 E5$3 G5$3 I5$3 K5$3 ] (14);
[C5$4 E5$4 G5$4 I5$4 K5$4 ] (15);

[d2$1 d3$1 d4$1 d5$1 ];

!AR(1) structure for latent variables
z2 on z1 (16);
z3 on z2 (16);
z4 on z3 (16);
z5 on z4 (16);

zd on z1;

zd@0; !variance of zd set to zero

!covariates effects
z1 on sex ageC eduC empC;
zd on sex ageC eduC empC;
z2 on sex (17) ageC (18) eduC (19) empE(21);
z3 on sex (17) ageC (18) eduC (19) empG(21);
z4 on sex (17) ageC (18) eduC (19) empI(21);
z5 on sex (17) ageC (18) eduC (19) empK(21);
```

```
!estimate means of latent variables
[z2 z3 z4 z5];

!random effects independent of each other and of latent variables
u1 with u2-u3 @0;
u2 with u3 @0;
u1-u3 with z1-z5 @0;
u1-u3 with zd @0;
zd with z2-z5 @0;

!errors of latent variables are not allowed to correlate
z1 with z2-z5 @0;
z2 with z3-z5 @0;
z3 with z4-z5 @0;
z4 with z5 @0;

OUTPUT:
tech4;
```

Appendix C

WinBUGS Code for an IRT Model for Multivariate Binary Longitudinal Data Subject to Dropout

BUGS language for a latent variable model, with covariates for 5819 cases of the Woman data (five waves C E G I K, three variables: questions 3,4,5) - dropout, no intermittent missingness or item non-response

i -> index for individual
j -> index for item, m is used when loop starts from 2
k -> index for wave, l is used when loop starts from 2
r -> index for covariate
N -> number of individuals
p -> number of items at each wave
T -> number of waves
R -> number of covariates

```
model
{
# Latent Variable Model
  for (i in 1 : N) {
  # Structural part
  z[i,1]<-beta.1[1]*x[i,1,1]+beta.1[2]*x[i,2,1]
            +beta.1[3]*x[i,3,1]+beta.1[4]*x[i,4,1]+delta.z1[i]
  delta.z1[i]~dnorm(0,tau.delta.z1)
    for(l in 2:T){
    z[i,l]<-a[l-1]+phi*z[i,l-1]+beta.t[1]*x[i,1,1]+beta.t[2]*x[i,2,1]
              +beta.t[3]*x[i,3,1]+beta.t[4]*x[i,4,l]+delta[i,l]
```

```
      delta[i,l]~dnorm(0,tau.delta[l-1])
                   }
      for(j in 1:p){
      u[i,j] ~ dnorm(0,tau.u[j]) I(-20,20)
                   }
      # Measurement part
      for (k in 1 : T) {
      logit(prob[i, 1, k]) <- z[i,k] - alpha[1]+u[i,1]
      y[i, 1, k] ~ dbern(prob[i, 1, k])
         for (m in 2 : p) {
         logit(prob[i, m, k]) <- lambda[m-1] * z[i,k] - alpha[m]+u[i,m]
         y[i, m, k] ~ dbern(prob[i, m, k])
                            }
                         }
      #Dropout
      for (l in 2:T) {
      logit(p.drop[i,l])<- -alpha.drop[l-1]+gamma*z[i,1] +beta.drop[1]*x[i,1,1]
                  +beta.drop[2]*x[i,2,1] +beta.drop[3]*x[i,3,1]+beta.drop[4]*x[i,4,1]
      d[i,l] ~ dbern(p.drop[i,l])
                      }
      }
#Priors
   alpha[1] ~ dnorm(0,0.0001)
   upsilon.z1 ~ dunif(0.0001,100)
   tau.delta.z1<-1/(upsilon.z1*upsilon.z1)
   phi ~ dnorm(0,0.0001)
   gamma ~ dnorm(0,0.0001)
   for (j in 1:p) {
   sigma.u[j] ~ dunif(0.0001,100)
   tau.u[j]<-1/(sigma.u[j]*sigma.u[j])
                   }
   for (m in 2 : p) {
   alpha[m] ~ dnorm(0, 0.0001)
   lambda[m-1] ~ dnorm(0,0.0001)
                      }
   for (l in 1:T-1) {
   a[l] ~ dnorm(0, 0.0001)
   alpha.drop[l] ~ dnorm(0,0.0001)
   upsilon[l] ~ dunif(0.0001,100)
   tau.delta[l]<-1/(upsilon[l]*upsilon[l])
```

```
                }
   for(r in 1:R) {
   beta.1[r] ~ dnorm(0,0.0001)
   beta.t[r] ~ dnorm(0,0.0001)
   beta.drop[r] ~ dnorm(0,0.0001)
                }
}
```

Initial values 1 ⇒ list (alpha = c(0,0,0), lambda = c(0,0), a = c(0,0,0,0), upsilon.z1 = 0.5, phi = 0.5, upsilon = c(0.5,0.5,0.5,0.5), gamma = 0.5, alpha.drop = c(0,0,0,0), sigma.u = c(0.5,0.5,0.5), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0), beta.drop = c(0,0,0,0))

Initial values 2 ⇒ list (alpha = c(0.5,0.5,0.5), lambda = c(0.5,0.5), a = c(0.5,0.5,0.5,0.5), upsilon.z1 = 1,phi = 1, upsilon = c(1,1,1,1), gamma = 1, alpha.drop = c(0.5,0.5,0.5,0.5), sigma.u = c(1,1,1), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0), beta.drop = c(0,0,0,0))

Appendix D

WinBUGS Code for an IRT Model for Multivariate Binary Longitudinal Data Subject to Dropout and Intermittent Missingness, Attitude at First Wave Affecting Missingness

BUGS language for a latent variable model, with covariates for 7622 cases of the Woman data (five waves C E G I K, three variables: questions 3,4,5) - dropout and intermittent missingness, no item non-response, attitude at first wave affecting missingness mechanism

i -> index for individual
j -> index for item, m is used when loop starts from 2
k -> index for wave, l is used when loop starts from 2
r -> index for covariate
N -> number of individuals
p -> number of items at each wave
T -> number of waves
R -> number of covariates
Q -> number of states for missingness (0:observed, 1:intermittent, 2:dropout) changed into 1,2,3 for waves 1,2,3,4
For the last wave 0: observed, 1:dropout

```
model
{
```

```
# Latent Variable Model
  for (i in 1 : N) {
  # Structural part
  z[i,1]<-beta.1[1]*x[i,1,1]+beta.1[2]*x[i,2,1]
          +beta.1[3]*x[i,3,1]+beta.1[4]*x[i,4,1]+delta.z1[i]
  delta.z1[i]~dnorm(0,tau.delta.z1)
    for(l in 2:T){
    z[i,l]<-a[l-1]+phi*z[i,l-1]+beta.t[1]*x[i,1,1]+beta.t[2]*x[i,2,1]
            +beta.t[3]*x[i,3,1]+beta.t[4]*x[i,4,l]+delta[i,l]
    delta[i,l]~dnorm(0,tau.delta[l-1])
              }
    for(j in 1:p){
    u[i,j] ~ dnorm(0,tau.u[j]) I(-20,20)
              }
    # Measurement part
    for (k in 1 : T) {
    logit(prob[i, 1, k]) <- z[i,k] - alpha[1]+u[i,1]
    y[i, 1, k] ~ dbern(prob[i, 1, k])
      for (m in 2 : p) {
      logit(prob[i, m, k]) <- lambda[m-1] * z[i,k] - alpha[m]+u[i,m]
      y[i, m, k] ~ dbern(prob[i, m, k])
                      }
                  }
  #Missing data mechanism
  for (l in 2:T-1) {
  #conditional probabilities, d=3 is an absorbing state
  if.branch[i,l]<- 1+step(d[i,l-1] -3) #1 if d[l-1] 1,2 and 2 if d[l-1]=3
  prob.miss.branch[i,l,1,2]<-0
  prob.miss.branch[i,l,2,2]<-0
  prob.miss.branch[i,l,3,2]<-1
    for (q in 1:Q) {
    #linear predictor
    eta[i,l,q]<- -alpha.miss[l-1,q]+gamma[q]*z[i,1] +beta.miss[1,q]*x[i,1,1]
            +beta.miss[2,q]*x[i,2,1] +beta.miss[3,q]*x[i,3,1]+beta.miss[4,q]*x[i,4,1]
    expeta[i,l,q]<-exp(eta[i,l,q])
    #probabilities (link function)
    prob.miss.branch[i,l,q,1]<-expeta[i,l,q]/sum(expeta[i,l,1:Q])
    prob.miss[i,l,q]<- prob.miss.branch[i,l,q, if.branch[i,l] ]
                  }
    #stochastic part
```

```
    d[i,l] ~ dcat (prob.miss[i,l,1:Q])
                  }
  #Wave T
  if.branch[i,T]<- 1+step(d[i,T-1] -3) #1 if d[T-1] 1,2 and 2 if d[T-1]=3
  #prob.miss.branch[i,T,1,2]<-0
  prob.drop.branch[i,T,2]<-1
  logit(prob.drop.branch[i,T,1])<- -alpha.miss[T-1,Q]+gamma[Q]*z[i,1]
                        +beta.miss[1,Q]*x[i,1,1]+beta.miss[2,Q]*x[i,2,1]
                        +beta.miss[3,Q]*x[i,3,1]+beta.miss[4,Q]*x[i,4,1]
  prob.drop[i,T]<- prob.drop.branch[i,T, if.branch[i,T] ]
  d[i,T] ~ dbern(prob.drop[i,T])
}
#Priors and constraints
  alpha[1] ~ dnorm(0,0.0001)
  upsilon.z1 ~ dunif(0.0001,100)
  tau.delta.z1<-1/(upsilon.z1*upsilon.z1)
  phi ~ dnorm(0,0.0001)
  for(q in 2:Q) {
  gamma[q]~dnorm(0,0.0001)
                }
  for (j in 1:p) {
  sigma.u[j] ~ dunif(0.0001,100)
  tau.u[j]<-1/(sigma.u[j]*sigma.u[j])
                }
  for (m in 2 : p) {
  alpha[m] ~ dnorm(0, 0.0001)
  lambda[m-1] ~ dnorm(0,0.0001)
                  }
  for (l in 1:T-1) {
  a[l] ~ dnorm(0, 0.0001)
  upsilon[l] ~ dunif(0.0001,100)
  tau.delta[l]<-1/(upsilon[l]*upsilon[l])
  alpha.miss[l,1]<-0 #constraint
                }
  for (l in 1:T-2) {
    for(q in 2:Q) { alpha.miss[l,q]~dnorm(0,0.0001)}
                  }
    alpha.miss[T-1,2]<-0 #constraint
    alpha.miss[T-1,3] ~ dnorm(0,0.0001)
  for(r in 1:R) {
```

```
    beta.1[r] ~ dnorm(0,0.0001)
    beta.t[r] ~ dnorm(0,0.0001)
    beta.miss[r,1]<-0 #constraint
    for(q in 2:Q) {beta.miss[r,q] ~ dnorm(0,0.0001)}
                  }
}
```

Initial values 1 ⇒ list (alpha = c(0,0,0), lambda = c(0,0), a = c(0,0,0,0),
upsilon.z1 = 0.5, phi = 0.5, upsilon = c(0.5,0.5,0.5,0.5), gamma = c(NA,0.5,0.5),
alpha.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,NA,0), .Dim = c(4,3)),
sigma.u = c(0.5,0.5,0.5), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0),
beta.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,0,0), .Dim=c(4,3)))

Initial values 2 ⇒ list (alpha = c(0.5,0.5,0.5), lambda = c(0.5,0.5), a = c(0.5,0.5,0.5,0.5),
upsilon.z1 = 1,phi = 1, upsilon = c(1,1,1,1), gamma = c (NA,1,1),
alpha.miss = structure (.Data = c(NA,0.5,0.5,NA,0.5,0.5,NA,0.5,0.5,NA,NA,0.5),
.Dim = c(4,3)),
sigma.u = c(1,1,1), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0),
beta.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,0,0), .Dim=c(4,3)))

Appendix E

WinBUGS Code for an IRT Model for Multivariate Binary Longitudinal Data Subject to Dropout and Intermittent Missingness, Time-Dependent Attitudes Affecting Missingness

BUGS language for a latent variable model, with covariates for 7622 cases of the Woman data (five waves C E G I K, three variables: questions 3,4,5) - dropout and intermittent missingness, no item non-response, time-dependent attitudes affecting missingness mechanism

i -> index for individual
j -> index for item, m is used when loop starts from 2
k -> index for wave, l is used when loop starts from 2
r -> index for covariate
N -> number of individuals
p -> number of items at each wave
T -> number of waves
R -> number of covariates
Q -> number of states for missingness (0:observed, 1:intermittent, 2:dropout) changed into 1,2,3 for waves 1,2,3,4
For the last wave 0: observed, 1:dropout

```
model
{
```

```
# Latent Variable Model
  for (i in 1 : N) {
  # Structural part
  z[i,1]<-beta.1[1]*x[i,1,1]+beta.1[2]*x[i,2,1]
           +beta.1[3]*x[i,3,1]+beta.1[4]*x[i,4,1]+delta.z1[i]
  delta.z1[i]~dnorm(0,tau.delta.z1)
    for(l in 2:T){
    z[i,l]<-a[l-1]+phi*z[i,l-1]+beta.t[1]*x[i,1,1]+beta.t[2]*x[i,2,1]
             +beta.t[3]*x[i,3,1]+beta.t[4]*x[i,4,l]+delta[i,l]
    delta[i,l]~dnorm(0,tau.delta[l-1])
               }
    for(j in 1:p){
    u[i,j] ~ dnorm(0,tau.u[j]) I(-20,20)
               }
    # Measurement part
    for (k in 1 : T) {
    logit(prob[i, 1, k]) <- z[i,k] - alpha[1]+u[i,1]
    y[i, 1, k] ~ dbern(prob[i, 1, k])
      for (m in 2 : p) {
      logit(prob[i, m, k]) <- lambda[m-1] * z[i,k] - alpha[m]+u[i,m]
      y[i, m, k] ~ dbern(prob[i, m, k])
                     }
                   }
  #Missing data mechanism
  for (l in 2:T-1) {
  #conditional probabilities, d=3 is an absorbing state
  if.branch[i,l]<- 1+step(d[i,l-1] -3) #1 if d[l-1] 1,2 and 2 if d[l-1]=3
  prob.miss.branch[i,l,1,2]<-0
  prob.miss.branch[i,l,2,2]<-0
  prob.miss.branch[i,l,3,2]<-1
    for (q in 1:Q) {
    #linear predictor
    eta[i,l,q]<- -alpha.miss[l-1,q]+gamma[q]*z[i,l] +beta.miss[1,q]*x[i,1,1]
             +beta.miss[2,q]*x[i,2,1] +beta.miss[3,q]*x[i,3,1]+beta.miss[4,q]*x[i,4,l]
    expeta[i,l,q]<-exp(eta[i,l,q])
    #probabilities (link function)
    prob.miss.branch[i,l,q,1]<-expeta[i,l,q]/sum(expeta[i,l,1:Q])
    prob.miss[i,l,q]<- prob.miss.branch[i,l,q, if.branch[i,l] ]
                   }
    #stochastic part
```

```
    d[i,l] ~ dcat (prob.miss[i,l,1:Q])
                  }
  #Wave T
  if.branch[i,T]<- 1+step(d[i,T-1] -3) #1 if d[T-1] 1,2 and 2 if d[T-1]=3
  #prob.miss.branch[i,T,1,2]<-0
  prob.drop.branch[i,T,2]<-1
  logit(prob.drop.branch[i,T,1])<- -alpha.miss[T-1,Q]+gamma[Q]*z[i,T]
                         +beta.miss[1,Q]*x[i,1,1]+beta.miss[2,Q]*x[i,2,1]
                         +beta.miss[3,Q]*x[i,3,1]+beta.miss[4,Q]*x[i,4,T]
  prob.drop[i,T]<- prob.drop.branch[i,T, if.branch[i,T] ]
  d[i,T] ~ dbern(prob.drop[i,T])
}
#Priors and constraints
  alpha[1] ~ dnorm(0,0.0001)
  upsilon.z1 ~ dunif(0.0001,100)
  tau.delta.z1<-1/(upsilon.z1*upsilon.z1)
  phi ~ dnorm(0,0.0001)
  for(q in 2:Q) {
  gamma[q]~dnorm(0,0.0001)
                }
  for (j in 1:p) {
  sigma.u[j] ~ dunif(0.0001,100)
  tau.u[j]<-1/(sigma.u[j]*sigma.u[j])
                }
  for (m in 2 : p) {
  alpha[m] ~ dnorm(0, 0.0001)
  lambda[m-1] ~ dnorm(0,0.0001)
                  }
  for (l in 1:T-1) {
  a[l] ~ dnorm(0, 0.0001)
  upsilon[l] ~ dunif(0.0001,100)
  tau.delta[l]<-1/(upsilon[l]*upsilon[l])
  alpha.miss[l,1]<-0 #constraint
                }
  for (l in 1:T-2) {
    for(q in 2:Q) { alpha.miss[l,q]~dnorm(0,0.0001)}
                  }
    alpha.miss[T-1,2]<-0 #constraint
    alpha.miss[T-1,3] ~ dnorm(0,0.0001)
  for(r in 1:R) {
```

```
   beta.1[r] ~ dnorm(0,0.0001)
   beta.t[r] ~ dnorm(0,0.0001)
   beta.miss[r,1]<-0 #constraint
   for(q in 2:Q) {beta.miss[r,q] ~ dnorm(0,0.0001)}
                 }
}
```

Initial values 1 ⇒ list (alpha = c(0,0,0), lambda = c(0,0), a = c(0,0,0,0), upsilon.z1 = 0.5, phi = 0.5, upsilon = c(0.5,0.5,0.5,0.5), gamma = c(NA,0.5,0.5), alpha.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,NA,0), .Dim = c(4,3)), sigma.u = c(0.5,0.5,0.5), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0), beta.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,0,0), .Dim=c(4,3)))

Initial values 2 ⇒ list (alpha = c(0.5,0.5,0.5), lambda = c(0.5,0.5), a = c(0.5,0.5,0.5,0.5), upsilon.z1 = 1,phi = 1, upsilon = c(1,1,1,1), gamma = c (NA,1,1), alpha.miss = structure (.Data = c(NA,0.5,0.5,NA,0.5,0.5,NA,0.5,0.5,NA,NA,0.5), .Dim = c(4,3)), sigma.u = c(1,1,1), beta.1 = c(0,0,0,0), beta.t = c(0,0,0,0), beta.miss = structure (.Data = c(NA,0,0,NA,0,0,NA,0,0,NA,0,0), .Dim=c(4,3)))

Bibliography

Albert, P., D. Follmann, S. Wang, and E. Suh (2002). A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics 58*, 631–642.

Allison, P. (2012). Handling missing data by maximum likelihood. In *SAS Global Forum: Statistics and Data Analysis.*

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological methodology 13*(1), 61–98.

Arminger, G. and U. Küsters (1988). *Latent trait and latent class models*, Chapter Latent trait models with indicators of mixed measurement level, pp. 51–73. Springer US.

Bartholomew, D. J., M. Knott, and I. Moustaki (2011). *Latent Variable Models and Factor Analysis: a Unified Approach.* London: Wiley.

Bartholomew, D. J. and P. Tzamourani (1999). The goodness of fit of latent trait models in attitude measurement. *Sociological Methods & Research 27*(4), 525–546.

Bentler, P. (1990). Comparative fit indexes in structural models. *Psychological Bulletin 107*(2), 238–246.

Bijleveld, C. C. J. H., A. Mooijaart, L. J. T. van der Kamp, and W. A. van der Kloot (1998). *Longitudinal Data Analysis: Designs, Models and Methods*, Chapter Structural equation models for longitudinal data, pp. 207–268. London: SAGE Publications Ltd.

Brooks, S. and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics 7*(4), 434–455.

Brooks, S. P. and G. O. Roberts (1998). Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing 8*, 319–335.

Cagnone, S., I. Moustaki, and V. Vasdekis (2009). Latent variable models for multivariate longitudinal ordinal responses. *British Journal of Mathematical and Statistical Psychology 62*, 401–415.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika 75*, 581–612.

Dantan, E., C. Proust-Lima, L. Letenneur, and H. Jacqmin-Gadda (2008). Pattern mixture model and latent class model for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics 4*(1), 1–26.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*(1), 1–38.

Diggle, P. J., P. J. Heagerty, K. Liang, and S. L. Zeger (2013). *Analysis of Longitudinal Data* (2nd ed.). Oxford: Oxford University Press.

Diggle, P. J. and M. Kenward (1994). Informative drop-out in longitudinal data analysis. *Applied Statistics 43*, 49–93.

Donnelly, K., J. M. Twenge, M. A. Clark, S. K. Shaikh, A. Beiler-May, and N. T. Carter (2015). Attitudes toward women's work and family roles in the united states, 1976-2013. *Psychology of Women Quarterly*. DOI:101177/0361684315590774.

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

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

Fieuws, S. and G. Verbeke (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics 62*, 424–431.

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

Forero, C. G., A. Maydeu-Olivares, and D. Gallardo-Pujol (2009). Factor analysis with ordinal indicators: A monte carlo study comparing dwls and uls estimation. *Structural Equation Modeling 16*(4), 625–641.

Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical science 7*(4), 457–472.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. *Bayesian Statistics 4*, 169–193.

Heidelberger, P. and P. D. Welch (1983). Simulation run length control in the presence of an initial transient. *Operations Research 31*(6), 1109–1144.

Henderson, R., P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics 1*, 465–480.

Holman, R. and C. A. W. Glas (2005). Modelling non-ignorable missing-data mechanisms with item response theory models. *British Journal of Mathematical and Statistical Psychology 58*, 1–17.

Hooper, D., J. Coughlan, and M. Mullen (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods 6*(1), 53–60.

Jansen, I., C. Beunckens, G. Molenberghs, G. Verbeke, and C. Mallinckrodt (2006). Analyzing incomplete discrete longitudinal clinical trial data. *Statistical Science 21*, 52–69.

Jöreskog, K. G. (1990). New developments in lisrel: analysis of ordinal variavariables polychoric correlations and weighted least squares. *Quality and Quantity 24*, 387–404.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika 59*, 381–389.

Jöreskog, K. G. (2005). Structural equation modelling with ordinal variables using LISREL. Technical report, Scientific Software International.

Jöreskog, K. G. and I. Moustaki (2001). Factor analysis of ordinal variables: a comparison of three approaches. *Multivariate Behavioral Research 36*(3), 347–387.

Jöreskog, K. G. and D. Sörbom (1996). *LISREL 8: User's Reference Guide.* Scientific Software International.

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

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

Little, R. J. and D. B. Rubin (2002). *Statistical analysis with missing data, Second Edition.* John Wiley & Sons.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association 88*, 125–134.

Lunn, D., A. Thomas, N. Best, and D. Spiegelhalter (2000). Winbugs – a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing 10*, 325–337.

Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika 47*(2), 149–174.

Masters, G. N. (1988). The analysis of partial credit scoring. *Applied Measurement in Education 1*(4), 279–297.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological) 42*(2), 109–142.

McElvey, R. and W. Zavoina (1975). Astatistdata model for the analysis of ordered level dependent variables. *Journal of Mathematical Sociology 4*, 103–120.

Millsap, R. E. and J. Yun-Tein (2004). .assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research 39*(3), 479–515.

Minini, P. and M. Chavance (2004). Sensitivity analysis of longitudinal binary data with non-monotone missing values. *Biostatistics 5*(4), 531–544.

Molenberghs, G., M. G. Kenward, and E. Lesaffre (1997). The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika 84*, 33–44.

Molenberghs, G., H. Thijs, I. Jansen, C. Beunckens, M. G. Kenward, C. Mallinckrodt, and R. J. Carroll (2004). Analyzing incomplete longitudinal clinical trial data. *Biostatistics 5*, 445–464.

Molenberghs, G. and G. Verbeke (2001). A review on linear mixed models for longitudinal data, possibly subject to dropout. *Statistical Modelling 1*, 235–269.

Moustaki, I. (1996). A latent trait and a latent class model for mixed observed variables. *British Journal of Mathematical and Statistical Psychology 49*, 313–334.

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

Moustaki, I. and M. Knott (2000b). Weighting for item non-response in attitude scales by using latent variable models with covariates. *Journal of the Royal Statistical Society, Series A 163*, 445–459.

Moustaki, I. and M. Knott (2005). *Computational aspects of the EM and Bayesian estimation in Latent Variable Models*, Chapter New Developments in Categorical Data Analysis for the Social and Behavioral Sciences, pp. 103–124. Psychology Press.

Moustaki, I. and F. Steele (2005). Latent variable models for mixed categorical and survival responses, with an application to fertility preferences and family planning in bangladesh. *Statistical Modelling 5*, 327–342.

Muthén, L. K. and B. O. Muthén (1998–2011). *Mplus User's Guide* (Sixth ed.). Los Angeles, CA: Muthén & Muthén.

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

Muthén, B., T. Asparouhov, A. M. Hunter, and A. F. Leuchter (2011). Growth modeling with nonignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods 16*, 17–33.

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

Olinsky, A., S. Chen, and L. Harlow (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research 151*, 53–79.

O'Muircheartaigh, C. and I. Moustaki (1999). Symmetric pattern models: a latent variable approach to item non-response in attitude scales. *Journal of the Royal Statistical Society, Series A 162*, 177–194.

Patz, R. J. and B. W. Junker (1999a). A straightforward approach to markov chain monte carlo methods for item response models. *Journal of Educational and Behavioral Statistics 24*(2), 146–178.

Patz, R. J. and B. W. Junker (1999b). Applications and extensions of mcmc in irt: Multiple item types, missing data, and rated responses. *Journal of Educational and Behavioral Statistics 24*(4), 342–366.

Plummer, M., N. Best, K. Cowles, and K. Vines (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News 6*(1), 7–11.

Rabe-Hesketh, S. and A. Skrondal (2012). *Multilevel and Longitudinal Modeling Using Stata, Third Edition. Volume I: Continuous Responses.* College Station, Texas: Stata Press.

Raftery, A. L. and S. Lewis (1992). How many iterations in the gibbs sampler? *Bayesian Statistics 4*(2), 763–773.

Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics 59*, 829–836.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*, 581–592.

Rubin, D. B. (1978). Multiple imputations in sample surveys. *Proceedings Survey Research Methods Section, American Statistical Association*, 20–34.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: Wiley.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association 91*, 473–489.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika monograph supplement No. 17.*

Skrondal, A. and S. Rabe-Hesketh (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models.* Boca Raton, FL: Chapman & Hall/CRC.

Steen, K., G. Molenberghs, G. Verbeke, and H. Thijs (2001). A local influence approach to sensitivity analysis of incomplete longitudinal ordinal data. *Statistical Modelling 1*, 125–142.

Steiger, J. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research 25*, 173–180.

Taylor, M. F., J. Brice, N. Buck, and E. Prentice-Lane (2010). *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices.* Colchester: University of Essex.

Ten Have, T. R., A. R. Kunselman, E. P. Pulkstenis, and J. R. Landis (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics 54*, 367–383.

Troxel, A. B., D. P. Harrington, and S. R. Lipsitz (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics 47*(3), 425–438.

Troxel, A. B., S. R. Lipsitz, and D. P. Harrington (1998). Marginal models for the analysis of longitudinal measurements with nonignorable non-monotone missing data. *Biometrika 85*(3), 661–672.

Vasdekis, V., S. Cagnone, and I. Moustaki (2012). A composite likelihood inference in latent variable models for ordinal longitudinal responses. *Psychometrika 77*, 425–441.

Verbeke, G. and G. Molenberghs (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.

Verbeke, G., G. Molenberghs, H. Thijs, E. Lesaffre, and M. Kenward (2001). Sensitivity analysis for nonrandom dropout: A local influence approach. *Biometrics 57*, 7–14.

Wu, M. C. and K. R. Bailey (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics 45*, 939–955.

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

Yang-Wallentin, F., K. G. Jöreskog, and H. Luo (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling 17*(3), 392–423.