THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Misspecification and flexible random effect distributions in logistic mixed effects models applied to panel survey data

Louise Marquart-Wilson

BSc BEc BAppSc(Hons)

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2016

Institute for Social Science Research

Abstract

Logistic mixed models for binary longitudinal panel data typically assume normal distributed random effects, and appropriately account for correlated data, unobserved heterogeneity and missing data due to attrition. However, this normality assumption may be too restrictive to capture unobserved heterogeneity. The motivating case study is a longitudinal analysis of women's employment participation using data from the Household Income and Labour Force Dynamics in Australia (HILDA) survey. Multimodality of the random effects was identified, potentially due to an underlying mover-stayer scenario.

This study focuses on logistic mixed models applied to the HILDA case study and simulation studies motivated by the case study, and aims to investigate:

1. robustness of random intercept logistic models to the assumed normal random effects distribution when the true distribution is multimodal
2. whether relaxing the parametric assumption of the random effects distribution can provide a practical solution to reduce the impact of distributional misspecification
3. impact of misspecification and performance of logistic mixed models in the presence of missing data due to attrition.

Random intercept logistic models applied to the case study demonstrate that the assumed normal distribution may not adequately capture the underlying heterogeneity due to a potential mover-stayer scenario. An asymmetric three component mixture of normal distributions provided a more appropriate fit, potentially representing three sub-populations: those with an extremely low, moderate, or extremely high propensity to be constantly employed.

Two simulation studies motivated by the HILDA study considered a three component mixture of normal distributions for the random intercept. The inferential impact of incorrectly assuming a normal distribution was dependent on the severity of departure of the true distribution from normality. In the first study, simulating a potential mover-stayer scenario, misspecification produced biased estimates of the intercept constant and random effect variance. More severely asymmetric and skewed multimodal distributions produced larger bias. The second study considered a range of true symmetric multimodal distributions, with increasing severity in departures from normality. The random intercept logistic model assuming normality was robust to minor deviations. However, for larger departures characterised by three distinct modes,

misspecification produced biased parameter estimates and poor coverage rates for the intercept constant, time-invariant explanatory variables and those time-varying explanatory variables exhibiting minimal within-individual variability. For both simulation studies, estimates of the random effect variance were extremely sensitive to distributional misspecification, resulting in biased parameter estimates, poor coverage rates and inaccurate standard errors.

Non-parametric estimation techniques, which leave the distribution completely unspecified, reduced the risks associated with misspecification of the random effects distribution. A novel application of the Vertex Exchange Method (VEM) was used to non-parametrically estimate the random effects distribution in logistic mixed models. The VEM was computationally intensive yet performed well to capture the univariate and bivariate random effects distribution when applied to the HILDA case study. VEM was the only method to converge when applied to the random intercept and random slope logistic mixed model. Inferential conclusions for the fixed effects parameters differed depending on the approach utilised, highlighting the practical use of sensitivity analyses to identify potential distributional misspecification of the random effects.

Distributional misspecification of the random intercept in the presence of missing data from attrition gave similar parameter estimates as for the complete case analysis, indicative of missing at random (MAR) missingness. The two simulation studies show that MAR attrition had minimal additional inferential impact on misspecifying the random intercept distribution, for a similar rate of 29.5% attrition observed in HILDA. As the negligible impact may partly be explained by the consistency of logistic mixed models in the presence of MAR missingness and by the large sample size, consideration of other missingness mechanisms and rates could be valuable. Flexible and non-parametric approaches applied to settings with attrition performed similarly as the complete case scenario.

Appropriate statistical analysis of longitudinal panel data is fundamental for researchers and policy makers to formulate and evaluate policy initiatives in health and social sciences. Hence, the need for the appropriate use and understanding of statistical models is crucial. This study provides a novel insight into the impact of assuming normality for the random effects in logistic mixed models applied to panel data where an underlying sub-population structure is suspected. For substantial departures characterised by multimodality with distinct modes, inference for the fixed effect parameters, typically the parameters of interest, can be impacted. Misspecification in the presence of MAR attrition had negligible additional inferential impact. More flexible distributions for the random effects is a practical solution to help reduce the impact of violating distributional

assumptions, and identify potential misspecification when used within a sensitivity analysis framework. VEM induced sufficient flexibility to capture multimodality of random intercepts and the complexity of the bivariate random effects in panel survey settings, including attrition. The performance of the VEM to flexibly model random effects should encourage its implementation in applications in the health and social sciences.

Declaration by author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis.

Publications during candidature

Conference abstracts

Marquart L, Haynes M and Baker P (2014) Impact of misspecified random effects distribution in multilevel models: applications to panel data, Australian Statistical Conference, Sydney 7-10 July 2014.

Marquart L, Haynes M and Baker P (2014) Impact of misspecified random effect distributions on models for panel survey data, In: Kneib T, Sobotka F, Fahrenholz J and Irmer H, *Proceedings of the 29th International Workshop on Statistical Modelling*, (207-212) Gottingen, Germany, 14-18 July 2014.

Publications included in this thesis

No publications included.

Contributions by others to the thesis

This thesis has benefitted from substantial inputs from Professor Michele Haynes on the concept, design of the project and interpretation of the results, and Dr. Peter Baker on the concept of the project. Professor Geert Verbeke significantly contributed to the concept, design, overview and interpretation of the statistical analysis presented in Chapter 7. Dr. Roula Tsonaka provided the R-code for the Vertex Exchange Method to non-parametrically estimate random intercept logistic models, which was used as the basis for extending the code to the random intercept and random slope logistic mixed model. Dr. Reza Drikvandi provided the syntax for the asymptotic gradient function diagnostic test utilised in Chapter 4, and helped with the interpretation of the diagnostic test results. This thesis also greatly benefitted from the critical reviewing, input and editorial work of Professor Michele Haynes, Professor Peter O'Rourke, Professor Geert Verbeke and Dr. Priya Parmar.

Statement of parts of the thesis submitted to qualify for the award of another degree

None.

Acknowledgements

Firstly, I would like to thank my supervisors, Michele Haynes and Peter Baker. To my primary supervisor, Michele, thank you for your encouragement, constructive feedback, support and guidance throughout this research project. I would also like to thank my academic reader, Mark Western, thank you for your insightful comments and challenging questions.

I am thankful for all the continued support and words of wisdom that I have received throughout my PhD journey. I would like to express my sincere gratitude to my mentor, Peter O'Rourke. Thank you for your continual support, encouragement and belief in me. Your guidance throughout my PhD has been a consistent source of strength, which has helped me to stay focussed, motivated and positive. Thank you for devoting so much time to helping me, for your careful attention to detail to ensure that the "numbers make sense", for asking stimulating questions and for providing in-depth feedback. I am also sincerely grateful to Professor Geert Verbeke, thank you for giving me a chance and sharing your expertise and knowledge with me, it has been an invaluable opportunity to learn from you and work with you. Special thanks to Dr. Priya Parmar for your endless support and encouragement over the years, and thank you for your valuable feedback; and to Dr. Emma Huang, thank you for your friendship and encouragement.

I would like to acknowledge the financial support I have received, and would like to thank the University of Queensland for the Australian Postgraduate Award (funded by the Australian Government) and Institute for Social Science Research for the generous top-up scholarship. I am grateful for the two travel grants that I received, both provided invaluable opportunities and enhanced my learning experience. Thank you to the Statistical Modelling Society for the student travel grant that helped support my attendance at the 29th International Workshop on Statistical Modelling in Gottingen, Germany in 2014. Thank you to the Graduate School at the University of Queensland for the GSITA that funded my travel to Belgium in 2015.

This paper uses unit record data from the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA project was initiated and is funded by the Australian Government Department of Families, Housing, Community Services and Indigenous Affairs (FaHCSIA) and is managed by the Melbourne Institute of Applied Economic and Social Research (Melbourne Institute). The findings and the views reported in this paper, however, are those of the author and should not be attributed to either FaHCSIA or the Melbourne Institute.

Finally, thanks to my friends and family. Thanks to my friends who have been constantly encouraging me and helping keep things in perspective. To my loyal, ever supportive and loving family: my parents, Cornelis and Ada, my brothers, Edric and Eugene, and my potentially soon-to-be sister in laws - Sophie and Alana, thank you for always believing in me. To my loving husband Ben, thank you for always supporting me, encouraging my independence, and helping me to always see the bigger picture. It is crazy to think, that over the past 10 years you have known me, I have been studying or thinking of continuing my study the entire time! I'm looking forward to not having my "head stuck in the books" and am looking forward to our future adventures together.

Keywords

Logistic mixed models, random effects, random effects distributional misspecification, mixture distributions, attrition, longitudinal panel survey, mover-stayer scenario, non-parametric estimation, Vertex Exchange Method

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZSRC code: 010401, Applied Statistics, 60%
ANZSRC code: 010405, Statistical Theory, 20%
ANZSRC code: 160807, Sociological Methodology and Research Methods, 20%

Fields of Research (FoR) Classification

FoR code: 0104, Statistics, 100%

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
BHPS	British Household Panel Survey
CC	Complete Cases
CI	Confidence Interval
EB	Empirical Bayes
EM	Expectation-Maximisation
GEE	Generalised estimating equation
GLM	Generalised linear model
GLMM	Generalised linear mixed model
HILDA	Household and Income Labour Dynamics in Australia
ICC	Intra-class correlation
LMM	Linear mixed model
LR	Likelihood reformulation
MAR	Missing at random
MCAR	Missing completely at random
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
MLE	Maximum likelihood estimator
MNAR	Missing not at random
NLMM	Non-linear mixed model
NLS	National Longitudinal Survey of Labour Market Experience
NPML	Non-parametric maximum likelihood
NPMLE	Non-parametric maximum likelihood estimator
PIT	Probability integral transformation
PSID	Panel Study of Income Dynamics
QMC	Quasi-Monte Carlo
SD	Standard deviation
SE	Standard error
SNP	Semi-non-parametric
SOEP	Socio-economic Panel Survey
VEM	Vertex Exchange Method

List of Symbols

Symbol	**Description**
i	index for the individual
j	index for the time-points
n_i	the number time-points observed for indiviudal i
N	the total number of individuals
p	the number of the fixed effects
q	the number of the random effects
$\boldsymbol{y_i}$	a $n_i \times 1$ vector of response variables for individual i
$\boldsymbol{x}_{ij}$	a $p-$ dimensional vector of explanatory variables with corresponding fixed effects for individual i at time j
$\boldsymbol{z}_{ij}$	a $q-$ dimensional vector of explanatory variables with corresponding random effects for individual i at time j
$\boldsymbol{b}_i$	a $q-$ dimensional vector of random effects for individual i
b_{0i}	Random intercept value for individual i
b_{1i}	Random slope value for individual i
$\boldsymbol{N}_q(\boldsymbol{\mu}, \boldsymbol{\Sigma}_q)$	a $q-$ dimensional multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ_q
$N(\mu, \sigma)$	a univariate normal distribution with mean μ and variance σ
$\boldsymbol{\Sigma}$	Variance-covariance matrix
$\boldsymbol{\mu}$	vector of means
$[a, b]$	in the interval of a and b including a and b
Pr	Probability
∞	infinity
ρ	Intra-class correlation
E	Expectation
$\mathrm{Var}(\cdot)$	Variance: $\mathrm{Var}(y) = \mathrm{E}(y - \mathrm{E}(y))^2$
$\sigma^2_{b_0}$	variance of the random intercept
$\sigma^2_{b_1}$	variance of the random slope
σ_{b_0,b_1}	the covariance between the random intercept and the random slope, b_{0i} and b_{1i}
$\phi(\cdot)$	a density function of the standard normal distribution
$f(\cdot)$	density function

$F(\cdot)$	distribution function
$f_i(y_i\|\beta,\Sigma)$	likelihood contribution for individual i
$L(\beta,\Sigma)$	Likelihood function: $L(\beta,\Sigma)=\prod_{i=1}^{N} f_i(y_i\|\beta,\Sigma)$
G	distribution function of the random effects
g	number of components in the heterogeneity model
π_λ	mixing proportion for λ^{th} component of the heterogeneity model
K	truncation factor in the Semi-non-parametric density
M	number of mass points in the non-parametric maximum likelihood estimator
$S_{-\infty}$	latent stayers in the null state
$S_{+\infty}$	latent stayers in the non-null state
$\Phi(G,H)$	directional derivative of the true distribution G in the direction of the alternative distribution H
$\Delta(G,b)$	gradient function
I	supportive region for the gradient function diagnostic test
S	total number of simulations
R_{ij}	indicator that subject i is observed at time j
D_i	drop-out time for individual i
d_{ij}	drop-out indicator for individual i at time j

1 Introduction

Appropriate statistical models for the analysis of longitudinal panel survey data enhance the validity and reliability of research findings in the health and social sciences. Panel surveys are an important source of data, providing an avenue to analyse pathways in individual outcomes and their relationships with socio-demographic characteristics and other major life events. Longitudinal panel data arises from collecting data repeatedly over time from the same panel of individuals, typically from large, nationally representative surveys. Development and use of appropriate statistical methodology for analysis of such data has not only become an important part of research, but has been fundamental for policy makers to evaluate and formulate policy initiatives. The complex sampling design of panel surveys and the correlated nature of longitudinal data, present many challenges to be considered during statistical analysis. Furthermore, statistical methods need to adequately control for unobserved variability among individuals and handle missing data due to permanent loss of respondents (attrition).

The generalised linear mixed model (GLMM) is a commonly used approach that can accommodate the aforementioned challenges and features of longitudinal panel surveys. GLMMs provide a flexible framework when applied to longitudinal data as they take into account the within-individual variability associated with having collected multiple data points per person, as well as the between-individual variability. The variability between individuals may be partly explained by measured characteristics of an individual, which are subsequently included as explanatory variables in the regression model (known as fixed effects). However, there may be other sources of heterogeneity due to unobserved individual characteristics unable to be collected by the survey. The GLMM captures the unobserved heterogeneity by incorporating individual-specific random effects into the model.

Data are often collected in categorical form in panel surveys due to administration of self-reported questionnaires. Logistic mixed effect models are a special case of GLMMs suitable for analysing longitudinal binary data and are useful for modelling the probability of being in one of two states over time, such as being in employment or non-employment. In the simplest case, the random intercept logistic model consists of a single random effect for each individual, where the random intercept represents the overall effect of all unmeasured individual-specific effects that captures the variability associated with why some individuals are more likely to be employed than others. Maximum likelihood techniques are often used to estimate the model parameters of GLMMs, which rely on statistical assumptions to obtain valid parameter estimates and standard errors, and hence, interpretation of results. In regards to the random effects, it is typically assumed that the random effects follow a Gaussian (or normal) distribution.

In practice the routinely assumed Gaussian distribution may be too restrictive. One example occurs when a key categorical variable is omitted from the model, which can result in multimodality of the random effects distribution. Multimodality can also occur when an underlying sub-population structure exists. For example, by the very nature of modelling binary response data, a subset of individuals may be observed to never change states over the observational period (a constant response profile), while another subset transitions between the two states. There are some instances whereby the constant response profile may consist of a sub-population structure as a consequence of an underlying mover-stayer scenario (Blumen et al., 1955). In this scenario, individuals with constant response profiles may consist of two sub-populations, those with a high propensity to remain in the same state (latent stayers), and those with a propensity to change states (latent movers). Thus, the random intercepts capturing the unobserved heterogeneity may be dominated by three sub-populations: one with a very low propensity of ever experiencing the outcome of interest, a more heterogeneous group with a propensity to transition between states, and one with a very high propensity of always experiencing the outcome of interest. Hence, the assumed normal distribution of the random intercept may not capture the heterogeneity of the underlying mover-stayer scenario.

As the random effects are unmeasurable, the validity of assumptions relating to the random effects distribution can be difficult to check. To help guard against the potential impact of misspecifying the random effects distribution, the parametric normality assumption can be relaxed by inducing flexibility of the assumed distribution. Flexibility can be achieved by modelling the random effects as non-Gaussian parametric distributions using computational methods (Nelson et al., 2006; Liu and Yu, 2008), however they may not adequately capture underlying multimodality. Alternatively semi-non-parametric methods (Chen et al., 2002; Vock et al., 2014) can be used to induce sufficient flexibility to capture multimodality. An approach that allows considerable flexibility is to leave the random effects distribution unspecified through the use of non-parametric methods (Laird, 1978; Heckman and Singer, 1984; Aitkin, 1999; Lesperance et al., 2014). In the context of an underlying mover-stayer scenario, discrete masses at negative and positive infinity can be incorporated into the random effects distribution to represent the two latent stayer groups with extremely low and extremely high propensity of experiencing the event (Davies et al., 1992; Berridge and Crouchley, 2011a). Alternatively, modelling the random intercepts as a finite mixture of normal distributions (Verbeke and Lesaffre, 1996) is plausible, where three components could capture the three latent sub-populations. Albeit the development of methodology to induce flexibility of the random effects distribution, implementation in practice is limited, particularly within the social sciences.

Understanding the impact of violating the assumed random effects distribution will have important implications for researchers utilising GLMMs to analyse longitudinal panel data, and additionally, for policy makers interpreting the results. Recent research has investigated the impact of misspecifying the random effects distribution in logistic mixed models (e.g. Neuhaus et al. 1992; Heagerty and Kurland 2001; Litiére et al. 2008; McCulloch and Neuhaus 2011a; Neuhaus et al. 2013), generally reporting biased estimation of parameters associated with the misspecified random effects, such as the intercept constant and variance estimates of the random effects. However there has been no general consensus about the impact on estimating fixed effects parameters, usually the parameters of primary interest. Previous literature predominately utilise simulation studies to investigate the impact of misspecifying the assumed random effects distribution, considering alternative true and assumed distributions. Limited research has considered the impact of incorrectly assuming normality when the underlying random effects distribution is multimodal, in particular, a multimodal distribution with three modes that may represent an underlying mover-stayer scenario.

The ambiguity about the impact of misspecifying the assumed random effects distribution has been further exacerbated by the lack of investigation of complexities often experienced when analysing longitudinal data, such as missing data due to attrition. This is an important area of study as the issue of misspecifying the random effects distribution in the presence of attrition can occur in practice. In the context of GLMMs, only one study has investigated the impact of misspecifying the random effects distribution in the presence of missing data. Wang (2010b) has shown that intermittent missingness and attrition can affect the power to detect variance components when the true random effects distribution is skewed yet assumed to be normal. This is an area requiring further research, as the loss of respondents due to non-response and attrition are prevalent in panel survey settings.

This study will investigate the impact of misspecifying the random effects distribution on inference for logistic mixed models in panel survey applications, focusing on multimodality of the underlying random effects distribution and on the presence of missing data due to attrition. This study will provide applications to the social sciences by modelling employment participation of working aged women using eleven years of data from the Household, Income and Labour Dynamics in Australia (HILDA) panel survey, demonstrating multimodality of the random effects as a consequence of a potential mover-stayer scenario. Through two simulation studies based on the HILDA case study, the first aim is to investigate the robustness of inference in random intercept logistic models to the normality assumption when the true distribution is multimodal, in the presence of complete and missing data following from attrition. Motivated by the underlying distribution observed in the HILDA case study, the first simulation study considers the specific departure from normality arising from a three component mixture of Gaussians to represent the mover-stayer scenario. The second simulation study investigates the robustness by simulating a range of multimodal distributions increasing in severity of departures from the assumed normal distribution. Further aims of the thesis are to determine whether flexibly modelling the random effects distribution is practically viable in panel survey settings, and to investigate the feasibility of applying a non-parametric technique, the Vertex Exchange Method (Böhning, 1985), to induce sufficient flexibly to capture underlying heterogeneity of the random effects distribution.

This study aims to address the following three research questions:

1. How robust is the assumed Gaussian distribution to multimodality of the random intercept distribution in panel survey settings due to a potential mover-stayer scenario?

2. Can the impact of multimodality of the random effects distribution be alleviated by increasing the flexibility of the assumed random effects distribution?

3. What is the additional impact of misspecified random effect distributions in the presence of missing data due to attrition?

The following chapter provides a background and overview of the current state of literature, highlighting key gaps that will be addressed by this study as a contribution to the current knowledge. The chapter begins by introducing longitudinal panel survey data and statistical methodologies commonly utilised within the social sciences to adequately account for the features and complexities of longitudinal panel data. This is followed by a focused review on GLMMs, introducing the statistical framework and the assumptions of the random effects structure. Specific focus is on the departure from the normality assumption of the random effects distribution, reviewing literature investigating the impact of misspecification, and reviewing methods used to address and detect misspecification.

Chapter 3 contains the statistical methodology that will be implemented and utilised within the thesis. Chapter 3 first introduces the methodology of GLMMs and logistic mixed models, and then describes four approaches developed to induce flexibility of the assumed random effects distribution. This is followed by the methodology underlying two diagnostic tests to identify misspecification of the random effects distribution in GLMMs. Finally, the chapter details the methodology of the simulation study and the corresponding performance measures to evaluate the impact of misspecification.

Chapter 4 will introduce a case study to investigate potential misspecification of the random effect distribution in an application of a random intercept logistic model to panel survey data. The case study investigates women's employment participation using eleven waves of data from the HILDA panel survey, and provides an example whereby the assumed normal distribution may not adequately capture the heterogeneity of a potential underlying mover-stayer scenario. By considering the missingness due to attrition, the case study will investigate the simultaneous occurrence of misspecification of the random effects distribution and attrition in practice.

Chapter 5 will investigate the impact of assuming normality of the random intercepts in a potential mover-stayer scenario. The simulation study used to investigate the impact of misspecification in the random intercept logistic model is motivated by the multimodality of the random intercept distribution characterised in the case study presented in Chapter 4. Furthermore, by replicating attrition as observed in the HILDA case study, Chapter 5 will investigate the simultaneous impact of misspecification in the presence of attrition.

Chapter 6 will investigate the robustness of the assumed normality distribution to multimodality of the true random effects distribution. By considering a variety of multimodal random intercept distributions increasing in severity of departure from the assumed normality, the simulation study will aim to identify scenarios whereby inference of random intercept logistic models is impacted by misspecified assumptions of the random effects distribution. The simulation study will investigate the impact of misspecification in the presence of attrition.

Chapter 7 will investigate the performance of a novel application of the Vertex Exchange Method (VEM) (Böhning, 1985) to non-parametrically estimate the random effects distribution in logistic mixed models. This approach to flexibly model the random effects distribution will be compared to some of the leading methods available to induce flexibility of the random effects distribution when applied to the HILDA case study. In addition to the random intercept logistic model, the VEM approach will be demonstrated to perform well in comparison to the other methods in an application to the random intercepts and random slopes logistic model.

Finally, Chapter 8 will present a discussion highlighting the relevance to panel survey applications within the social science discipline, and conclude with ideas and avenues for future research.

2 Background and Literature Review

This chapter provides a background and overview of the current state of literature, highlighting key gaps that will be addressed by this study as a contribution to the current knowledge. The chapter can be structurally divided into three parts. The first part begins by introducing longitudinal panel survey data, highlighting the features and complexities that need to be considered to ensure accurate estimation and inference when statistically analysing this rich source of data. A common issue inherent in the collection of longitudinal panel surveys is missing data, and Section 2.2 highlights different types of missingness and the hierarchy of underlying missingness mechanisms. Particular attention will focus on the type of missingness when participants drop-out of the study, known as attrition. Section 2.3 introduces statistical methodologies that are commonly utilised to model the change in response variables over time, particularly focusing on models for categorical variables with two possible values, known as binary response variables.

The second part focuses on generalised linear mixed models (GLMMs), specifically the logistic mixed model for binary response variables. Section 2.4 more formally introduces the general framework and assumptions of GLMMs, focusing on the normality assumption of the random effects distribution. However, as reviewed in Section 2.5, previous literature has identified scenarios where the normal distribution of the random effects may be too restrictive in practice. Particular attention will focus on the specific departure from normality characterised by multimodality due to an underlying sub-population structure, such as the mover-stayer scenario. The impact of misspecifying the random effects distribution on model estimates and inferential conclusions is reviewed in Section 2.6, focusing on incorrectly assuming normality for the random effects distribution in logistic mixed models. Section 2.7 reviews methods to address and detect misspecification of the random effects distribution, particularly focusing on methodology suitable to address and formally detect multimodality of the random effects distribution.

The third part begins in Section 2.8 with an overview of the current state of literature, highlighting the relevance of the literature to this study. Finally, Section 2.9 concludes by stating the research questions and how these will be addressed in this study.

2.1 Longitudinal panel surveys

The use of longitudinal panel survey data is common within the health, economics and social science disciplines. Panel surveys are a research design used to collect data from a 'panel' of the same group of units surveyed repeatedly over time (Frees, 2004; Andress et al., 2013). Typically the focus of panel surveys is to study individual changes over time. However this design can also be applied to assess changes in other units of measurement including companies, nations or other social entities (Andress et al., 2013).

Within the social sciences, panel surveys typically sample individuals within households through large population based surveys to obtain nationally representative data. This rich source of data captures time-series and cross-sectional data, providing an avenue to study characteristics that influence individual pathways, in addition to household transitions and societal trends as they evolve over time. Not only has this data become increasingly attractive in social science research, but it is often utilised to evaluate and formulate government initiatives. Predominant panel survey studies within the social sciences include the National Longitudinal Survey of Labour Market Experience (NLS, initiated in 1966) and the Panel Study of Income Dynamics (PSID, initiated in 1968) conducted in the United States. More recently household panel surveys from other countries have been initiated, including the German Socio-economic Panel Study (SOEP, initiated in 1984) and the British Household Panel Survey (BHPS, initiated in 1991).

One of the first and longest running national longitudinal panel surveys in Australia is the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA survey was an initiative of the Australian government, explicitly designed to inform policy development in the areas of economic and social participation, and family dynamics (Watson and Wooden, 2012). The design of the HILDA survey is largely similar to previous large scale, nationally representative panel surveys, especially the BHPS and the SOEP (Watson and Wooden, 2012). The HILDA survey has been conducted annually since 2001 with data collected from all members aged over 15 years in each household, with the first panel consisting of 13,969 individuals living in 7,682 households (Wooden and Watson, 2007). The reference population for HILDA was, with minor exceptions, all persons residing in private dwellings in Australia (Wooden and Watson, 2007).

As the collection of longitudinal panel survey data has become increasingly available, appropriate statistical analysis has become an important component of research in the social sciences. Statistical considerations are required to take into account key features of the panel design, including the sampling design and the correlated nature of longitudinal data. For example, in an attempt to provide a representative sample of the reference population, complex sampling designs are commonly used to select the initial panel (Vieira and Skinner, 2008). For instance, the HILDA survey used a multi-stage approach involving stratification by geographical units, in conjunction with systematic and random sampling (Wooden and Watson, 2007) to select the initial households. If interest is in population based inference, statistical weights to account for the complex sampling design are constructed and are subsequently incorporated into statistical models to obtain robust population estimates (Goldstein, 2011).

As panel data measures each individual repeatedly over time, observations from the same individual tend to be more similar (correlated) than observations between different individuals[1]. The between-individual differences are known as heterogeneity, and may be partly explained by observed individual characteristics (observed heterogeneity) or unmeasurable individual characteristics (unobserved heterogeneity). Appropriate statistical models to account for the correlated structure of longitudinal data and different types of heterogeneity are necessary as the standard statistical analysis assumption of independence in the observations is violated. These statistical models will be discussed in more detail in Section 2.3.

In addition to the design features of panel surveys, the collection of longitudinal panel survey data is prone to challenges that also need to be considered during statistical analysis. The most common challenge is the issue of missing data. Missing data can arise when individuals do not participate at every time-point (intermittent non-response) or leave the study permanently (monotone missing or attrition). Furthermore, missing data can arise when individuals do not respond to all questionnaire items (item non-response). This results in unbalanced data, and may lead to selection bias where the individuals who choose to participate in the survey differ to those who choose not to participate. This can adversely affect the representativeness of the sample (Watson and Wooden, 2012) and can subsequently introduce bias into the model estimates and impact the validity of statistical analysis (Fitzmaurice et al., 2011). The presence of missing data and its treatment in statistical analysis adds more complexity to statistical models (Hedeker and Gibbons, 2006), and needs to be considered carefully in order to minimise bias, and the loss of information and precision.

Further complexities can have implications for statistical analyses, such as issues relating to measurement errors of the survey tool, temporal ordering of the measurements within an individual, and unbalanced data due to rotating panels and top-up samples. These issues are beyond the scope of this study, however highlight that although there are numerous advantages of panel data, there are complexities that need to be considered when analysing longitudinal panel data. This review will focus on statistical models suitable to account for the correlated structure of longitudinal data, heterogeneity and unbalanced data due to attrition. The next section more formally introduces the issue of missingness, and the underlying mechanisms for missing data.

[1]Likewise, observations from members of the same household tend to be more similar than observations from members of different households. As detailed in Section 2.3.1.1 this additional level of clustering can be accounted for, however in this study the level of clustering will be restricted to account for repeated measurements at the individual level.

2.2 Missingness in panel surveys

Following sample members in the initial panel over a long period of time, potentially over decades, is not a simple task (Andress et al., 2013). With each successive time-point, some sample members may temporarily not participate in the survey (intermittent missing), or permanently drop-out (attrition). Intermittent missingness is less likely to be related to the outcome(s) studied, hence leading to less bias in the results. However, attrition can be problematic to the representativeness of sample[2], particularly when a substantial proportion of the initial panel are lost due to attrition. For instance, over 11 years of follow-up in the HILDA survey, 62.9% of the initial sample members remained in the study, with similar rates of attrition reported for comparable time periods in other long-running nationally representative panel surveys (Watson and Wooden, 2012).

There are various reasons for sample members to drop out of the study. Not only will some sample members move away or stop participating due to health issues or death, but sample members may withdraw their cooperation from participating in the survey. As alluded to previously, the predominant concern when analysing longitudinal data with attrition, and missingness in general, is the issue of selection bias (Alderman et al., 2001). Selection bias occurs if there are non-random patterns of missingness, resulting in biased model estimates (Alderman et al., 2001). Therefore, understanding the reason for sample members to drop-out of the panel survey is crucial. However, often the reasons for missing data can not be collected, and thus, are typically not known.

Drop-out by sample members can be distinguished by how the missingness is related to the unobserved response variables and the observed data (including the observed variables and the responses). Rubin (1976) and Little and Rubin (1987) described a hierarchy of missingness consisting of three types of mechanisms: missing completely at random (MCAR), missing at random (MAR) or missing not at random (MNAR). MCAR is the most restrictive and assumes that, given the observed variables, missingness is independent of both the observed and unobserved responses. MAR is less restrictive whereby, when given the observed data, missingness is conditionally independent of the unobserved responses. Finally, MNAR occurs when missingness is not independent of unobserved response data, even after accounting for the observed data. The hierarchy was originally developed in the context of missingness in general, and later Diggle and Kenward (1994) applied the terminology in the context of attrition referring to the three mechanisms as completely random drop-out, random drop-out and informative drop-out. The notation and the methodology of the underlying missingness mechanisms is detailed in Section 3.4.2.

This section has predominately focused on individual non-response due to attrition. How-

[2]Note that as the population itself changes over time, representing a population over time is more complex than at a single time-point (Andress et al., 2013).

ever, in practice the issue of missingness is complex and can consist of different types of non-response such as item non-response and missingness due to late entry (i.e. from top-up samples). Furthermore, changes in the household composition can also lead to different types of missingness[3]. These issues are beyond the scope of this study. The issue of missingness and related methodology to address missingness is itself a vast literature, and will not be reviewed here[4]. For the remainder of this review, attention will be restricted to missingness due to attrition.

2.3 Statistical methods for analysis of longitudinal data

Longitudinal data is a special case of clustered data, where the clusters consist of repeated measurements from a single individual over several occasions (Fitzmaurice et al., 2011). Often the major focus of longitudinal data analysis is to estimate within-individual changes in the response variable over time (response trajectories) and to examine whether a set of explanatory variables (predictors)[5] influences between-individual variability. The statistical methods used to address these questions in longitudinal data are special cases of general regression methods used to analyse clustered data (Fitzmaurice et al., 2011). An array of methodology exists for analysing longitudinal data, with origins tracing back to the analysis of variance paradigm (Fitzmaurice et al., 2009). Statistical methods were originally developed for continuous responses, however for binary and categorical responses, techniques based on extensions of the generalised linear model (GLM) of Nelder and Wedderburn (1972) have been developed.

Analysis of longitudinal categorical responses is common in panel survey settings, predominately due to data collected through the administration of self-reported questionnaires. Examples of categorical variables in the HILDA survey include employment status and marital status, and ordinal categorical variables such as general health and job satisfaction measured on the five-point Likert scale. Variables which consist of two categories, known as binary variables, are often analysed within the social sciences. Not only can categorical or continuous variables be dichotomised into binary variables (for example, general health into poor or good health), but social phenomena of interest may consist of two categories, often representing yes/no or present/absent (for example, home-ownership, long-term health condition, recipient of government assistance, employed).

Three broad classes of regression to account for the clustering and heterogeneity in longitudinal data are known as: (1) marginal or population-averaged models, (2) subject-specific models, and (3) transition or response conditional models. These approaches will be briefly

[3]For example, missingness can occur when a temporary sample member is no longer living with a continuing sample member, or late entry can occur when a new sample member starts to live with a continuing sample member.

[4]For a detailed review of the issue of missingness in longitudinal data within the social sciences and more general framework, the reader is referred to Graham (2012) and Molenberghs et al. (2015), respectively, and the references within.

[5]Terminology in the literature also includes the term covariate for a set of explanatory variables. However, in this study the term covariate will be restricted to continuous explanatory variables.

introduced, however for a more thorough review the reader is referred to Fitzmaurice et al. (2011) and the references within. Not only do these models differ in how the correlation among the repeated measurements is accounted for, but also in the interpretation of the regression parameters (Fitzmaurice et al., 2011).

Marginal models explicitly model the relationship between the mean of the response and a set of predictors separately from the within-subject dependence (Hedeker and Gibbons, 2006). The generalised estimation equation (GEE) model (Zeger and Liang, 1986) is a commonly utilised marginal model. Marginal approaches are referred to as population-averaged models, as the effect of the predictors are interpreted as being averaged over the population of the subjects (Hedeker and Gibbons, 2006). In contrast, subject-specific models incorporate a parameter for each individual to capture the heterogeneity between individuals. By controlling for the subject-specific parameter and keeping the other predictors fixed, the interpretation of the effect of the predictor on the outcome is considered subject specific (Fitzmaurice et al., 2011). Transition models are also known as autoregressive models or conditional response models, as the mean response not only depends on a set of predictors but also on previous responses (Fitzmaurice et al., 2011). For example, the first-order Markov model assumes the probability of being in a given state at time j is dependent on the state occupied in the previous time point (i.e $j-1$).

As longitudinal panel surveys provide an opportunity to assess changes and trends at the individual level, subject-specific models are often utilised to provide an individual-specific interpretation by estimating individual trajectories and predictions. Here, attention is restricted to subject-specific models, in particular, to random effects models whereby the subject-specific parameter is treated as a random variable from a specified distribution. In the next section, the general framework for subject-specific models and random effects models are introduced.

2.3.1 Subject-specific models

To set the notation, consider a longitudinal panel survey consisting of N individuals for which data is collected up to n time-points (commonly referred to as waves). An individual i has observations measured repeatedly n_i times, such that y_{ij} denotes the response for individual i (for $i = 1, ..., N$) at time j (for $j = 1, ..., n_i$). For each individual and time-point, a set of p explanatory variables denoted by $\boldsymbol{x_{ij}}' = (x_{1ij}, ..., x_{pij})$ are collected. By the very nature of longitudinal data, two types of explanatory variables can be collected: variables that do not change over the study period (time-invariant), and variables that do change values over time (time-varying). Both of these types of explanatory variables are included in $\boldsymbol{x_{ij}}$.

For techniques based on extensions of the generalised linear model to analyse longitudinal categorical response variables, a link function is required to relate the response variable to the linear parameter vector. Specific link functions, known as canonical links, have convenient statistical properties (Skrondal and Rabe-Hesketh, 2004), and will be considered in this study. For the employment participation example, consider the binary response variable where $y_{ij} = 1$ when individual i at time j is employed, and $y_{ij} = 0$ when unemployed or not in the labour force. The logit link is the canonical link when modelling binary response data, and is the most commonly used link for binary response data. In the case of modelling a binary response variable over time, the following model with a logit link (known as the logistic model) takes into account the unobserved heterogeneity by incorporating an individual specific effect, α_i

$$\text{logit}(\Pr(y_{ij}) = 1) = \log(\frac{\Pr(y_{ij} = 1)}{1 - \Pr(y_{ij} = 1)}) = \alpha_i + \boldsymbol{x_{ij}}'\boldsymbol{\beta} \tag{2.1}$$

where $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ are the coefficients of the p explanatory variables. The intercept terms, α_i are allowed to vary for each individual, and are known as subject-specific effects. The coefficients $\boldsymbol{\beta}$ are common effects for all individuals, and are known as fixed effects.

2.3.1.1 Random effects models

A common approach used to account for the unmeasured heterogeneity is known as the random effects model, whereby the subject-specific parameter α_i is assumed to be a random variable from a specified distribution. The model in Equation 2.1 can be re-written such that $\alpha_i = \beta_0 + b_{0i}$, where β_0 is the fixed intercept constant and b_{0i} is the random intercept that represents the unobserved time-invariant heterogeneity. Typically the random intercept is assumed to be normally distributed ($b_{0i} \sim N(0, \sigma^2_{b_0})$).

Random effects models belong to the class of generalised linear mixed models (GLMMs) that extend GLMs by incorporating random effects into the regression model. GLMMs constitute a framework for a class of models for clustered response variables that follow a distribution from the exponential family. Examples of GLMMs include the linear mixed model for continuous response data, the logistic mixed models for binary response data, and the Poisson mixed model for count response data. The term 'mixed' refers to the fact that the GLMM includes both fixed effects[6] (parameter effects that are common among all individuals, such as $\boldsymbol{\beta}$ and β_0) and random effects (set of parameters that vary for each individual, such as b_{0i}).

The hierarchical structure of GLMMs can be described in the context of multilevel models (Goldstein, 1979). For instance, longitudinal data can be represented as a two-level model, where the repeated observations (level-1 unit) are nested within individuals (level-2 units)[7]. In

[6] Terminology regarding fixed effects can be confusing, as the definition can change across disciplines. For instance, within the multilevel model literature the term fixed effects model refers to a model that does not have any random components. As detailed in Section 2.3.2.1, within the social science and econometrics literature the term fixed effects model refers to a model whereby the subject-specific terms are considered nuisance parameters to be eliminated from the model.

[7] Higher order hierarchies could be considered. For example, consider the structure of household panel surveys, where the repeated observations (level-1 unit) are nested within individuals (level-2 unit) that are nested within households (level-3 unit).

this context, GLMMs assume that observations from an individual share a set of latent, unobserved random effects (Verbeke et al., 2010) that describe subject-specific deviations from the overall effect, and account for the correlation structure inherent in longitudinal data. The variables that have corresponding random effects are a subset of the explanatory variables included in the fixed effects. For instance, a random intercept logistic model consists of one random effect corresponding to the intercept constant (for example, Equation 2.1). However, multiple random effects can easily be incorporated. For instance, a logistic mixed model whereby both the intercept and an explanatory variable (typically capturing the time effect) have a corresponding random effect is commonly referred to as a random intercepts and random slopes logistic model[8].

2.3.2 Alternative statistical models

In addition to the aforementioned methods, alternative statistical methods are used within the economic and social science disciplines to model longitudinal and clustered data. The fixed effects regression model and the hybrid model, are two commonly implemented methods used to model change in longitudinal responses and account for unobserved heterogeneity. These two methods are briefly introduced in Sections 2.3.2.1 and 2.3.2.2, respectively. Furthermore, alternative methods have been developed to investigate the timing of events. One such model is the discrete time model and is briefly introduced in Section 2.3.2.3.

2.3.2.1 Fixed effects models

Traditionally fixed effects models and random effects models were considered different in how the individual-specific parameters capturing the time-constant unobserved variables (α_i in Equation 2.1) were treated. Fixed effects models treat α_i as a fixed, yet unknown, parameter to be estimated. In contrast, the random effects model assumes α_i is a random variable from a specified distribution. However, more recently the two approaches are considered to differ in the structure of the associations between the unobserved individual specific terms and the time-varying explanatory variables (Allison, 2009). The α_i is represented as a random variable for both approaches, however, the fixed effects model allows for correlations between α_i and the time-varying explanatory variables. This is in contrast to the random effects model, whereby the unobserved individual specific terms are assumed to be uncorrelated with all explanatory variables (due to assumptions imposing strict exogeneity and orthogonality between b_{0i} and x_{ij}, see Section 2.4 for further details).

The subject-specific parameter α_i is traditionally treated as a fixed term to be estimated using maximum likelihood. However, as the sample size increases, so does the number of parameters to be estimated. In this situation, model fitting can become difficult and estimates

[8]As detailed in Section 3.1, for models with multiple random effects the correlations between the random effects are included by modelling the structure of the variance-covariance matrix.

of the parameter coefficients $\boldsymbol{\beta}$ are no longer consistent (Agresti, 2013). An alternative estimation approach is to use conditional maximum likelihood (ML). Conditional ML treats α_i as a nuisance parameter and conditions them away using sufficient statistics[9] (McCulloch et al., 2008). Conditional on the sufficient statistic, maximising the likelihood will result in consistent estimators of the parameter coefficients $\boldsymbol{\beta}$ (Agresti, 2013). By allowing for correlations between α_i and the time-varying explanatory variables, fixed effects models control for the effects of time-invariant unobservable variables. However, the consistency of fixed effects models is sacrificed at the cost of efficiency, as fixed effects models only use data from individuals with non-constant response profiles. This can be an issue when modelling binary response data, as a subset of individuals may remain in the same category over the study period. In addition to the efficiency loss, fixed effects models do not provide estimates of the effects of time-invariant explanatory variables and use of conditional ML is restricted to models with canonical link functions (such as the logistic fixed effects model).

2.3.2.2 Hybrid models

Hybrid models (Allison, 2009) provide a framework that combines aspects of both fixed effects and random effects models. The hybrid model is also known as the covariate decomposition method (Neuhaus and Kalbfleisch, 1998; Neuhaus and McCulloch, 2006), as the time-varying explanatory variables of random effects models are separated into within-individual and between-individual components. These methods are considered as extensions of the GLMM (Bell and Jones, 2015), whereby the time-varying explanatory variables are transformed into deviations from their individual-specific means[10], that are included in the model together with the individual-specific means. However unlike in the fixed effects model, the response variable is not transformed. By decomposing the time-varying explanatory variables into within- and between-individual components, this model can provide unbiased estimates of explanatory variables correlated with the subject-specific effects. Furthermore, hybrid models can provide conditional likelihood-like inference for canonical and non-canonical link models (Neuhaus and McCulloch, 2006, 2014). Another example of a hybrid model is the conditional linear mixed model (Verbeke et al., 2001), where the random intercepts are conditioned away, but the other random effects are treated as random parameters. In theory this approach can be extended to GLMMs other than linear mixed models[11], and has the advantage that all longitudinal effects can be estimated without relying on correct model specification for modelling the cross-sectional differences between individuals.

The choice between random effects models and fixed effects models is often debated within the social sciences and economics discipline, and a detailed review is beyond the scope of this

[9] Often the sufficient statistic is the individual-specific sum of responses, i.e. $S_i = \sum_{j=1}^{n_i} y_{ij}$ for $i = 1, ..., N$

[10] Typically individual-specific means of the time-varying explanatory variables are used, however other statistics can be used (Neuhaus and McCulloch, 2014).

[11] However, the conditional distribution of the data given the sufficient statistic for the random intercepts in the GLMM may be much more complex than the one obtained for linear mixed models.

study. In regards to the issue of biased estimation of the random effects model when explanatory variables are correlated with the random effects, Clark and Linzer (2015) suggest that non-zero correlation does not necessarily imply that the fixed effects model is preferred[12]. Furthermore, although the commonly implemented Hausman test (Hausman, 1978) may indicate possible violation of the assumption that the explanatory variables are orthogonal to the random effects[13], it may not be reliable (Clark and Linzer, 2015) and should not be viewed as a decision tool to choose between fixed or random effects estimation (Bell and Jones, 2015). However, recently it has been argued that the main advantage of GLMMs is their generalisability and extendibility to handle complex data structures including multiple random effects, higher-order clustering, multiple membership, and cross classification (Bell and Jones, 2015). By extending the traditional GLMM to decompose time-varying explanatory variables within the within and between effects, Bell and Jones (2015) argue that the GLMM framework provides sufficient flexibility to analyse the longitudinal and clustered data.

2.3.2.3 Discrete time models

The above review has predominately focused on models for studying change in an outcome over time. However the other main strand of longitudinal research includes statistical models to investigate the occurrence of events (Steele, 2008). Such models include discrete time models within the event history analysis framework, whereby interest is in modelling the timing of events. In these models, the response variable is the length of time between being exposed to being at risk of an event and the event occurring (Steele, 2011). For example, in the context of employment participation, the event can be a change in employment status. As detailed in Steele (2011), discrete time models are a general model for repeated events, that can include more complex scenarios such as competing risks, multiple states and simultaneous processes. This modelling approach will not be considered further in this review, and the reader is referred to Steele (2011) for further details.

2.3.3 Assumptions regarding the underlying missingness mechanism

As alluded to in Section 2.2, statistical approaches for longitudinal data can vary in the assumption of the underlying mechanism for missingness. Understanding the reason for missing data is important as the performance and appropriateness of longitudinal models can depend on the underlying missingness model (Hedeker and Gibbons, 2006). For example, the GEE model has restrictive assumptions for missingness, assuming that underlying mechanism is MCAR. In contrast, the GLMM assumes that missingness is ignorable (MCAR or MAR), provided that the missing responses can be explained by the explanatory variables included in the mean

[12]Through simulation studies the authors show that the condition that the explanatory variables are uncorrelated with the random effects will hold only under exceptional circumstances, showing that the random effects model can be preferred (or at least perform no worse than) the fixed effects model (Clark and Linzer, 2015).

[13]The Hausman test can be used to assess whether there are similarities between the within-subject and the between-subject effects by testing if the coefficients of the time-varying explanatory variables from the random effects model are identical to the coefficients from the fixed effects model.

structure of the model[14] or explained by the available responses from a specific individual (Gibbons et al., 2010). Statistical methods have been developed to handle non-ignorable missing data (i.e. MNAR), including shared parameter models (Follmann and Wu, 1995) and pattern mixture models (Little, 1993, 1995). However, as the models for MNAR missingness rely on unverifiable assumptions[15], the use of these models is generally confined within a sensitivity analysis framework (Molenberghs et al., 2015). Sensitivity analyses considering different statistical models provides a practical way to assess the robustness of missingness assumptions, as the missingness mechanism is not usually testable (Skrondal and Rabe-Hesketh, 2014).

Consistent estimation of the alternative subject-specific models is also dependent on the underlying missingness mechanism. Recently, Neuhaus and McCulloch (2014) demonstrated that although fixed effects models (estimated using conditional ML) and hybrid models can provide consistent estimation in settings where the random effects are correlated with explanatory variables, both approaches can produce inconsistent estimation in the presence of MAR attrition[16]. Inconsistency of the conditional ML approaches in the presence of MAR attrition has also been reported previously by Rathouz (2004), however more recently, Skrondal and Rabe-Hesketh (2014) show that conditional ML methods can provide consistent estimation in settings where attrition is MNAR. Therefore, in general, conditional ML approaches requires the missingness mechanism to be MCAR[17].

As highlighted in this review, random effects models belong to the class of GLMMs which constitute a flexible framework to model changes in longitudinal responses and obtain subject-specific interpretation. GLMMs provide a powerful tool to handle the correlated nature of longitudinal data and account for unobserved heterogeneity. Furthermore, these models can provide consistent estimation in settings where the underlying missingness mechanism for attrition is MAR. The remainder of this review will focus on GLMMs, specifically the logistic mixed model for longitudinal binary responses. In the following section, a more detailed review of GLMMs and the corresponding statistical assumptions are presented. For more details regarding the statistical methodology of GLMMs the reader is referred to Section 3.1.

[14]This also assumes that specified model is correct.

[15]Similarly, the models assuming underlying MAR missingness also rely on an unverifiable assumption since MAR can not be tested.

[16]Typically the hybrid model decomposes the time-varying explanatory variables by using the individual-specific means, however as detailed in Neuhaus and McCulloch (2014) for analysis in the presence of MAR attrition these models can produce inconsistent estimates. Neuhaus and McCulloch (2014) show that by decomposing the time-varying explanatory variables using the baseline value, consistent estimates of the hybrid model can be produced in the presence of MAR attrition.

[17]With some special exceptions as highlighted above.

2.4 Assumptions and complexities of generalised linear mixed models

Often GLMMs are estimated using maximum likelihood techniques, and rely on statistical assumptions in order to obtain valid parameter estimates and standard errors, and hence, interpretation of results. In particular, GLMMs require the correct specification of both the linear predictor and the random effects structure. The GLMM assumes that conditional on the random effects, the response variables are assumed to be independent with density functions belonging to the exponential family. Furthermore, for a known link function, the conditional mean response is assumed to depend on a linear predictor containing fixed regression parameters and random effects.

In regards to the random effects structure, it is assumed that the random effects follow a specified distribution. In practice the random effects are typically assumed to follow a multivariate normal distribution. In addition to the distributional assumption of the random effects, it is assumed that the random effects have zero mean ($E(\boldsymbol{b_i}) = \mathbf{0}$) and have a common variance-covariance matrix. Furthermore, it is assumed that the random effects are independent of the explanatory variables ($\text{Cov}(\boldsymbol{b_i}, \boldsymbol{x_{ij}}) = \mathbf{0}$).

However, in practical applications of GLMMs, all assumptions are violated to some degree (McCulloch and Neuhaus, 2011a) and can subsequently have implications on inferential conclusions of the model parameters. For instance, misspecification of the conditional mean structure can occur if the link function is incorrectly specified, or, if an important explanatory variable is omitted from the linear predictor (McCulloch and Neuhaus, 2013). Research assessing the impact of misspecifying the random effects structure has predominately focused on three areas: incorrectly assuming independence to covariates (e.g. Heagerty and Kurland 2001; Neuhaus and McCulloch 2006; Huang 2009; Neuhaus and McCulloch 2014), incorrectly assuming independence to the cluster size (Neuhaus and McCulloch, 2011), and incorrectly specifying the distribution (e.g. Heagerty and Kurland 2001; McCulloch and Neuhaus 2011a; Neuhaus et al. 2013).

These aspects of random effects specification are generally considered separately in the literature. The impact of misspecifying the distributional assumption of the random effects will be detailed in Section 2.6, however literature assessing the impact of misspecifying other aspects of the random effects structure has reported biased inference of model parameters in GLMMs. For instance, substantial bias of parameter estimates can be produced when ignoring correlations between the explanatory variables and the random effects (Neuhaus and McCulloch, 2006) or characteristics of the random effects distribution (i.e. the mean (Neuhaus and McCulloch, 2006) or variance (Heagerty and Kurland, 2001)). Furthermore, in situations with missing data, ignoring any dependency between the number of observations and the random effects can produce biased parameter estimates for explanatory variables with corresponding random effects[18] (Neuhaus and McCulloch, 2011).

This study will primarily focus on investigating the impact of misspecifying the random effects distribution in GLMMs, particularly in logistic mixed models. Therefore unless stated otherwise, the term misspecification will be restricted to the distributional aspect of the random effects structure.

2.5 Non-Gaussian random effects

In some practical applications, the routinely assumed Gaussian distribution for the random effects in GLMMs may be too restrictive, particularly if heterogeneity of the random effects exists. Heterogeneity of the random effects may occur when a key time-invariant categorical variable is omitted from the model. For instance, omitting a binary time-invariant variable from the mean structure of the GLMM can result in severe polarization of the random effects distribution (Agresti et al., 2004). This can be extended to omitting a key categorical variable of three or more categories from the mean structure, resulting in a possibly multimodal random effects distribution following a finite mixture distribution (Verbeke and Molenberghs, 2009). Panel surveys can be prone to the omission of variables from the mean structure, as key variables may not be collected due to the broad scope of the study. In these scenarios, assuming a normal distribution for the random effects may not sufficiently capture the underlying heterogeneity.

Another scenario whereby the normality assumption of the random effects may be too restrictive, is when there are subgroups in the population who behave differently. By the very nature of binary outcomes, analysis of longitudinal binary responses can be complicated (Carlin et al., 2001). For instance, if a subject never exhibits an outcome over the study period (or always exhibits an outcome), it is not possible to assess the within-subject effects of time-varying variables, such as the effect of the time trend (Carlin et al., 2001). Not only do individuals with constant response profiles complicate the interpretation and comparison of the effects of time-varying variables, but the normality assumption of the random effects distribution may not be the most appropriate if a sub-population structure exists. For example, an underlying process known as the mover-stayer scenario may explain the constant response profiles. In this scenario, the individuals observed to have constant response profiles could consist of two subgroups: a subgroup of individuals known as latent stayers, who have a zero or extremely low probability of transitioning from the initial state; and, a second subgroup consisting of individuals known as latent movers, who have the propensity to transition yet were not observed to change states during the observational period. The latent movers also comprise of the group of individuals that have been observed to transition at least once between the states. In this

[18] However, Neuhaus and McCulloch (2011) argue that this type of misspecification is a form of misspecifying the random effects distribution.

scenario, the random effects may not be Gaussian distributed as the latent movers and stayers will have considerably different probabilities of exhibiting the outcome of interest.

Specifically, it is likely that three populations will dominate the distribution of the random effects in a random intercept model for binary responses (Caffo et al., 2007): one with very low probability of ever experiencing the outcome, a more homogeneous group that transition between the states over time, and one with a very high probability of always experiencing the outcome. Therefore, the two groups of latent stayers will have values for the random effects that are inconsistent with the tail of the normal distribution (Barry et al., 1989). To accommodate the two subgroups of latent stayers, more flexible random effects distributions have been proposed. Singer and Spilerman (1976) suggest the spiked distribution, whereby spikes are incorporated into the parametric random effects distribution to represent the latent stayers. For instance, to explicitly model the two groups of latent stayers in a random intercept logistic model, finite probabilities at $-\infty$ and ∞ can be incorporated into the assumed normal distribution (Davies and Crouchley, 1986; Berridge and Crouchley, 2011a). Alternatively, modelling the random effects as a three component mixture distribution could capture the heterogeneity inherent in the underlying mover-stayer context. In this model, the spiked distribution could be considered a special case where two components have zero variance and extreme means. As detailed further in the following section, this issue has been a long recognised problem (Barry et al., 1989) and has been encountered in various contexts including the mover-stayer model (Goodman, 1961) and in the competing risk context.

2.5.1 Movers and stayers

Originally described by Blumen et al. (1955), the mover-stayer model extends the Markov chain model to describe two types of individuals in the study population: the stayer who has the propensity to remain in the same category during the study period; and the mover who changes states over time as described by a Markov chain with a transition probability matrix (Goodman, 1961). The concept has been adopted within the survival model framework to describe long-term survivors (e.g. Farewell, 1982) and also to the logistic random effects model to describe individuals susceptible or not susceptible to experiencing the binary outcome of interest (e.g. Davies et al., 1992; Carlin et al., 2001).

Within the competing risk context, multistate models are an extension to survival models providing a useful tool to model processes that occupy a finite number of states with changes over time, such as disease progression. Patients may experience disease progression over a period of time (observed movers) and others may stay in the same disease state (observed stayers). Multistate models assume that the movement within a state space of discrete states is governed by an underlying stochastic process and the history of the process up to time immediately prior to the current time. The transitions among the states are governed by a transition probability matrix (O'Keeffe et al., 2013). Random effects can be incorporated into multistate models to account for the unexplained heterogeneity, though it can be hard to distinguish between the unexplained heterogeneity and the effects of the process history (Cook and Lawless, 2014). The random effects provide a measure for the propensity of a patient to progress through the disease states, and often are assumed to have a finite mixture distribution to account for the latent movers and latent stayers whereby the latent stayers have a probability mass at zero (O'Keeffe et al., 2013). Often the choice of the random effects distribution can lead to different inferences (O'Keeffe et al., 2013), and other random effect distributions such as the mover-stayer gamma, mover-stayer inverse-gamma and the compound Poisson distribution have been proposed for use in the disease progression scenario (O'Keeffe et al., 2013).

These examples identify situations where the routinely assumed Gaussian random effects distribution may not be the most appropriate distribution to capture the underlying heterogeneity, particularly in scenarios whereby the true random effects distribution is multimodal due to the existence of a sub-population structure. As random effects are latent, and hence unmeasurable, the validity of assumptions relating to the random effects are difficult to check (Alonso et al., 2010a). This has led to a growing body of research focusing on assessing the validity of random effect assumptions. Areas of research have focused on: investigating the impact of misspecification on inference and prediction, development of techniques to flexibly model random effects distributions in order to relax the normality assumption, and development of diagnostic tests to identify misspecification. These areas of research are discussed in more detail in the following sections, where attention will focus on these issues within the context of true multimodal distributed random effects.

2.6 Impact of misspecifying the random effects distribution

Often the normality assumption for the random effects is taken for granted (Verbeke and Molenberghs, 2013), however as the random effects are latent, the distributional assumptions can not be directly assessed. There is a growing literature investigating the impact of misspecifying the shape of the random effects distribution in generalised linear mixed models. Earlier literature predominately focused on misspecification of the random effects distribution in linear mixed models. As the linear mixed model is a special case of the GLMM, this review will initially focus on the linear mixed models before reviewing literature investigating the impact of misspecifying the random effects distribution in GLMMs for longitudinal binary responses.

In the context of linear mixed models, both theoretical and simulation studies have consistently shown negligible inferential impact of misspecifying the random effect distributions. Maximum likelihood estimates for fixed effects and variance components obtained under the assumption of Gaussian distributed random effects have been shown to be consistent and asymptotically normally distributed under broad regularity conditions, even when the random effects distribution is misspecified (Verbeke and Lesaffre, 1996, 1997; Neuhaus et al., 2013). Although the estimates are consistent, sandwich-type corrections are required to obtain the correct asymptotic standard errors (Butler and Louis, 1992; Verbeke and Lesaffre, 1997). However, the robust standard errors for the random effects variance remain inaccurate (Verbeke and Lesaffre, 1997; Maas and Hox, 2004). Furthermore, the best predicted values of the random effects are sensitive to the shape of the assumed random effects distribution (Verbeke and Lesaffre, 1996; Zhang and Davidian, 2001; McCulloch and Neuhaus, 2011b). For situations where the assumed distribution substantially deviated from the true random effects distribution, McCulloch and Neuhaus (2011b) reported loss of performance and accuracy of predicting the random effects.[19]

Unlike the consistent estimation reported for linear mixed models when the random effects distribution is misspecified, GLMMs for non-normal responses appear to be more sensitive to distributional violations. Theoretical results for the impact on the inference of random intercept logistic models (Neuhaus et al., 1992), and more generally for GLMMs with random intercepts and random slopes (Neuhaus et al., 2013), suggests that misspecifying the random effects distribution can produce biased estimates of parameters associated with the misspecified random effects (i.e. the intercept constant and the variance components). However, as the estimation of GLMMs for non-normal responses is complicated by not having a closed form expression of the joint density, theoretical results can only be derived for the restricted scenario when all explanatory variables are unrelated to the response (i.e. $\boldsymbol{\beta} = \mathbf{0}$).

To establish results in more general scenarios, researchers have predominately utilised simulation studies to investigate the impact of misspecification in GLMMs, with some focusing on logistic mixed models. The effect of misspecifying the random effects distribution on estimation and inference can be assessed by two approaches. The first simulates a variety of true distributions and examines the performance of GLMMs under the assumption of normal distributed random effects (e.g. Litiére et al. 2008 and Neuhaus et al. 2013). Alternatively, the second approach simulates a single true distribution and assesses the impact of misspecifying the random effects distribution by considering a variety of assumed parametric distributions (e.g. Neuhaus et al. 2011 and McCulloch and Neuhaus 2011a). Neuhaus et al. (2011) argue that the first approach merely investigates the robustness of the normality assumption and not misspecification. However, Litiére et al. (2011) later argued that both methods are complementary and assess the impact of misspecification. Due to the nature of simulation studies, not only are there considerable differences between the studies in regards to the assumed and true distributions considered, but also the simulation scenario (for example, number of individuals and number of time-points) and the mean structure of the model (for example, number and type of explanatory variables). A summary of the key settings considered in simulation studies investigating misspecification in logistic mixed models is outlined in Table 2.1.

[19]For more details about the inferential impact of misspecifying the random intercept distribution in linear mixed models the reader is referred to Butler and Louis (1992), Verbeke and Lesaffre (1997) and Verbeke and Molenberghs (2009). For details regarding the impact of misspecifying the joint distribution of the random intercepts and random slopes in linear mixed models, the reader is referred to Maas and Hox (2004) and the analytic results presented by Neuhaus et al. (2013).

Table 2.1: Summary of publications performing simulation studies to investigate the impact of misspecifying distributional assumptions of (i) univariate and (ii) bivariate random effects in logistic mixed models. Details for the simulation study (Study) include the specification of the logistic mixed model (Model Specification), number of individuals (N), the number of time-points (n_i), the true distribution generated for the random effects (True Distribution) and the assumed random effects distribution considered when fitting the logistic mixed model (Assumed Distribution).

Study	**Model Specification**	**N**	n_i	**True Distribution**	**Assumed Distribution**
(i) Univariate random effects					
Butler and Louis (1992)	$\text{logit}(\Pr(y_{ij} = 1) = \beta_0 + b_i + \beta_1 x_{1ij}$ where $\beta_0 = -1$, $\beta_1 = \frac{2}{3}$, $\beta_2 = 1$ $x_{1ij} = (0, 1, 2, 3)$	100	4	$b_i \sim N(0, 1)$	(i) Ordinary logistic model ($b_i = 0$) (ii) Normal (iii) Unspecified (NPML Estimation)
Neuhaus et al. (1992)	$\text{logit}(\Pr(y_{ij} = 1) = \beta_0 + \sigma b_i + \beta_1 x_{1ij} + \beta_2 x_{2i}$ where $\beta_0 = -2$, $\beta_1 = 0.5$, $\beta_2 = 1$, $\sigma = 2$ $x_{1ij} \sim N(0, 1)$, and $x_{2i} \sim N(0, 1)$	100	5	Standardised to have a $E(b_i) = 0$ and $Var(b_i) = 1$: $b_i \sim$ Gamma(1, 0.5) $b_i \sim$ Gamma(1,16) $b_i \sim$ t-distribution with df=3 $b_i \sim$ t-distribution with df=5 $b_i \sim N(0, 1)$	Normal
Heagerty and Kurland (2001)	$\text{logit}(\Pr(y_{ij} = 1)) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{1i} x_{2ij} + b_i$ where $\beta_0 = -2$, $\beta_1 = 1$, $\beta_2 = 0.5$, $\beta_3 = -0.25$ x_{1i} 50:50 binary indicator and $x_{2ij} = (0, 0.25, 0.5, 0.75, 1)$	200	5	$b_i = \sigma(a_i - \lambda)/\sqrt(\lambda)$ for $a_i \sim$ Gamma(λ, 1) for $\lambda = 0.5, 1, 2, 4$ and $\sigma = 0.5, 1, 2, 3$	Normal
Chen et al. (2002)	$\text{logit}(\Pr(y_{ij} = 1)) = \beta_1 x_{1i} + \beta_2 x_{2ij} + b_i$ where $\beta_1 = 0.5, \beta_2 = 3$ x_{1i} 50:50 binary indicator and $x_{2ij} = (-0.2, -0.1, 0, 0.1, 0.2)$.	250	5	$b_i \sim 0.7N(-.1.5, 0.72) + 0.3N(2, 0.72)$ $E(b_i) = 0.45$ and $Var(b_i) = 3.0625$	Semi-non-parametric model (K=0,1,2)
Agresti et al. (2004)	$\text{logit}(\Pr(y_{ij} = 1)) = \beta_0 + b_i$ where $\beta_0 = 0$ or $\beta_0 = 1$	10,30	10,30	Standardised to have $E(b_i) = 0$ and $Var(b_i) = 0, 0.25, 1$: $b_i \sim N(0, \sigma_b^2)$ $b_i \sim$ Uniform $b_i \sim$ Exponential $b_i \sim$ Discrete with two equal mass points.	(i) Normal (ii) Unspecified (NPML Estimation)

Table 2.1 continued on next page

Table 2.1 – continued from previous page

Study	Model Specification	N	n_i	True Distribution	Assumed Distribution
Litiére et al. (2008)	$\text{logit}(\Pr(y_{ij}=1)) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + b_i$ where $\beta_0 = -8$, $\beta_1 = 2$ and $\beta_2 = 1$ x_{1i} 50:50 binary indicator and $x_{2ij} = (0, 1, 2, 4, 6, 8)$	25, 50, 100, 200, 400, 800, 1600	6	Standardised to have $E(b_i) = 0$ and $Var(b_i) = 1, 4, 16, 32$: $b_i \sim N(0, \sigma_b^2)$ $b_i \sim$ Uniform $b_i \sim$ Exponential $b_i \sim$ Chi-squared distribution $b_i \sim$ log-normal $b_i \sim$ Power function $b_i \sim$ discrete with 2 mass points $b_i \sim \frac{1}{2}N(-\mu, \sigma_b^2) + \frac{1}{2}N(\mu, \sigma_b^2)$ $b_i \sim 0.2 \times N(\mu_1, \sigma_1^2) + 0.8 \times N(\mu_2, \sigma_2^2)$	Normal
McCulloch and Neuhaus (2011a)	$\text{logit}(\Pr(y_{ij}=1)) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + b_i$ where $\beta_0 = -2.5$, $\beta_1 = 2$ and $\beta_2 = 1$ x_{1i} 25:75 binary indicator and x_{2ij} = equally spaced from 0 to 1	200	2, 4, 6, 10, 20, 40	Standardised to have $Var(b_i) = 1$: $b_i \sim$ Tukey $(g = 0.5, h = 0.1)$	(i) Tukey with unknown g and h (ii) Tukey with $g = 0.5$ and $h = 0.1$ (iii) Normal
(ii) Bivariate random effects					
Litiére et al. (2008)	$\text{logit}(\Pr(y_{ij}=1)) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + b_{0i} + b_{1i} x_{2ij}$ where $\beta_0 = -6$, $\beta_1 = 2$, $\beta_2 = 1$ $x_{1i} =$ 50:50 binary indicator and $x_{2ij} = (0, 1, 2, 4, 6, 8)$	50, 100	6	Standardised to have $E(\boldsymbol{b_i}) = \mathbf{0}$), and overall variance-covariance structure $\boldsymbol{V}$ with variance v_a and covariances v_b for $V1$ with $v_a = 5, v_b = 4.5$; $V2$ with $v_a = 5, v_b = 4.9$; V_3 with $v_a = 8, v_b = 6$; V_4 with $v_a = 5, v_b = 4.5$ $\boldsymbol{b_i} = (b_{0i}, b_{1i})' \sim \text{N}(\mathbf{0}, \boldsymbol{V})$ $\boldsymbol{b_i} = (b_{0i}, b_{1i})' \sim \frac{1}{2}\ \text{N}(-\boldsymbol{\mu}, \boldsymbol{D}) + \frac{1}{2}\ \text{N}(\boldsymbol{\mu}, \boldsymbol{D})$ for $\boldsymbol{\mu} = (2, 2)'$ and $\boldsymbol{D}$ with variances $d = 1, 4$ and covariances d_{12} chosen to have corr(b_{0i}, b_{1i})= 0.5 and0.9	(i) Bivariate Normal (ii) Two component mixture of Bivariate Normals
Neuhaus et al. (2013)	$\text{logit}(\Pr(y_{ij}=1)) =$ $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + b_{0i} + b_{1i} x_{2ij}$ where $\beta_0 = -2.5$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$ x_{1i} 50:50 binary indicator, $x_{2ij} = (-1, -0.5, 0, 0.5, 1)$ and x_{3ij}=$(-0.5, 1, 0, -1, 0.5)$ and is orthogonal to x_{2ij}	100	5	Standardised to have $E(b_i) =$ $\boldsymbol{b_i} \sim$ bivariate normal $\boldsymbol{b_i} \sim$ bivariate t with df= 3 $\boldsymbol{b_i} \sim$ bivariate exponential(1) $\boldsymbol{b_i} \sim$ bivariate Tukey$(g = 0.446, h = 0.05)$ $\boldsymbol{b_i} \sim$ bivariate log-normal $\boldsymbol{b_i} \sim$ bivariate Tukey$(g = 0, h = 0.109)$ $\boldsymbol{b_i} \sim$ bivariate Tukey$(g = 0, h = 0.159)$ $\boldsymbol{b_i} \sim$ bivariate Tukey$(g = 0.249, h = 0.05)$	Bivariate Normal

The impact of misspecification has been suggested to be dependent on the inferential focus (McCulloch and Neuhaus, 2011a). As detailed in the following sections, misspecification in logistic mixed models can produce biased estimators of parameters directly related to the random effects, such as the intercept constant and the random variance component. However, the impact on estimated regression coefficients of the fixed effects is more ambiguous, with differences depending on whether the explanatory variable is time-varying or time-invariant. The following review will focus on the inferential impact of incorrectly assuming normality in logistic mixed effects models, investigating each type of parameter and inferential target separately. Previous literature has predominately focused on misspecifying random effects in random intercept logistic models (e.g. Neuhaus et al. 1992; Heagerty and Kurland 2001; Chen et al. 2002; Agresti et al. 2004). However, similar impact as for the random intercept model has been reported for the more complex scenario of random intercepts and random slopes, with biased estimation typically restricted to estimates directly corresponding to the misspecified random effects (McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013).

2.6.1 Estimation of coefficients for time-varying explanatory variables

The literature assessing the impact of misspecified random effects consistently reports minimal bias in estimating the effects of time-varying explanatory variables. The robustness to the choice of the assumed distribution has been postulated to be due to the orthogonality of the time-varying explanatory variables and the between-subject variability (Chen et al., 2002). Theoretical results of Neuhaus et al. (1992) showed that maximum likelihood estimators of coefficients for time-varying explanatory variables under distributional misspecification of the random effects converge to values that minimise the Kullback-Leibler divergence (Kullback, 1959) between the correctly and incorrectly specified models. For logistic mixed models, theoretical results of Neuhaus et al. (1992) show consistent estimation of the effects of time-varying explanatory variables when $\boldsymbol{\beta} = \mathbf{0}$. These theoretical results were later extended by Neuhaus et al. (2013) to the entire class of GLMMs, including multiple random effects.

For more general scenarios, results from the simulation studies concur with the theoretical findings. The results consistently show minimal impact of misspecification on the asymptotic bias of estimating the coefficients of time-varying explanatory variables (Heagerty and Kurland, 2001; Chen et al., 2002; Litiére et al., 2008; McCulloch and Neuhaus, 2011a). Simulation studies have predominately focused on the impact of estimating effects of a continuous explanatory variable (covariate)[20] representing the time effect. For instance, Heagerty and Kurland (2001) reported relative bias less than 15% for the covariate representing time and the corresponding interaction term when assuming normality for gamma distributed random intercepts. Similarly, McCulloch and Neuhaus (2011a) reported virtually no impact on the covariate representing the time effect when assuming normality for Tukey distributed random intercepts, and Litiére et al.

[20]This term in the literature may not always be restricted to represent continuous explanatory variables, but can also refer to any explanatory variable (either continuous or categorical).

(2008) reported relative bias less than 5% for all scenarios considered, regardless of sample size or variability of the true random effect. In addition to the minimal impact on bias, Chen et al. (2002) and McCulloch and Neuhaus (2011a) report no loss in estimation efficiency under misspecification of the random intercept distribution.

2.6.2 Estimation of coefficients for time-invariant explanatory variables

It has been previously conjectured that time-invariant explanatory variables are more sensitive to distributional misspecification of the random effects than time-varying explanatory variables, as both the time-invariant explanatory variables and the random effects capture variability between individuals (Chen et al., 2002). However, the theoretical results presented by Neuhaus et al. (1992) and Neuhaus et al. (2013) imply consistent estimation for any explanatory variable not related to the response ($\boldsymbol{\beta} = \mathbf{0}$), regardless of whether they are time-invariant or time-varying.

In contrast to the theoretical studies, results from simulation studies demonstrate that bias and loss of efficiency can occur when the true and assumed random effect distributions vary substantially. For instance, Agresti et al. (2004) reported moderate loss of efficiency when assuming normality and the true distribution was a two point discrete distribution. Both Heagerty and Kurland (2001) and Litiére et al. (2008) suggest larger bias corresponding with large true random effect variability. However, for less substantial departures from the true distribution, minimal bias has been reported. For instance, relative bias of less than 5% was reported by McCulloch and Neuhaus (2011a) when assuming normality for a random effects arising from a Tukey distribution.

2.6.3 Estimation of the intercept constant

Estimation of the intercept constant can be sensitive to the assumed random intercept distribution. Results from theoretical and simulation studies demonstrate inconsistent estimation of the intercept constant when the assumed distribution is far from the underlying true random effects distribution (Neuhaus et al., 1992; Heagerty and Kurland, 2001; Chen et al., 2002; Litiére et al., 2008; McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013). Consistent with theory (Neuhaus et al., 1992, 2013), relative asymptotic bias of up to 30% was reported for incorrectly assuming normality in random intercept logistic models when the true random effects were asymmetric (Neuhaus et al., 1992). Furthermore, relative bias of up to 20% has been reported in situations assuming normality when the true random intercepts distribution was a gamma distribution (Heagerty and Kurland, 2001), and up to 40% for true asymmetric two component mixture of normal distributions (Chen et al., 2002). Not only does misspecification result in biased estimates, Neuhaus et al. (2013) reported below nominal coverage rates of 92.5% for confidence intervals of logistic mixed models with misspecified random intercepts and random slopes. Typically inference of the intercept constant is not of direct interest. However if inference is on mean estimation, McCulloch and Neuhaus (2011a) suggest that care is required, as bias of the intercept constant can transfer over to mean estimation of the outcome value.

2.6.4 Estimation of the random effects variance

Little research has investigated the impact of distributional misspecification of the random effects on estimating the variance components. Although estimation of the random effect variance may not be considered a primary inferential interest (McCulloch and Neuhaus, 2011a), it is the only measurement of the underlying variability of the random effects (Litiére et al., 2008) and can be important when considering the variability attributable to various levels of multilevel data (McCulloch and Neuhaus, 2013)[21]. Minimal impact of misspecification in the random intercept logistic model has been reported (Heagerty and Kurland, 2001; McCulloch and Neuhaus, 2011a), with Heagerty and Kurland (2001) reporting relative bias of 15% or less. However, substantial bias has been demonstrated for scenarios with large discrepancies between the assumed and true random effects distribution (Neuhaus et al., 1992; Litiére et al., 2008; McCulloch and Neuhaus, 2011a). For instance, Litiére et al. (2008) reported absolute relative bias of up to 85% when incorrectly assuming normality for highly skewed true distributions[22]. Furthermore, larger magnitudes of bias and efficiency loss were associated with larger random effects variances (Agresti et al., 2004; Litiére et al., 2008) and larger cluster size (Agresti et al., 2004).

2.6.5 Prediction of the random effects

Not only can inference of the variance component be directly impacted by misspecification, McCulloch and Neuhaus (2011b) showed that biased estimation of the variance component can subsequently impact the accuracy of the best predicted random effect values. By theory and simulations, McCulloch and Neuhaus (2011b) demonstrated modest impact of misspecification on the prediction accuracy as measured by the mean square error of prediction. However, for large discrepancies between the assumed and true distributions, severe reduction in the efficiency of predicting the random effects has been reported (Agresti et al., 2004; McCulloch and Neuhaus, 2011b). For instance, McCulloch and Neuhaus (2011b) reported a reduction in the mean square error of prediction for logistic mixed models when the assumed distribution had limited support but the true distribution had a wide range of support. Furthermore, loss of efficiency has been reported when the true distribution had large variances and for situations with large cluster sizes (McCulloch and Neuhaus, 2011b).

Although the best predicted values are relatively accurate for minor to moderate deviations

[21] For example, in a three-level GLMM of modelling an outcome measured repeatedly over time for a group of students in different schools. The variance estimates of the random effects can be important to partition the variability attributable to schools rather than students.

[22] For true random intercepts distributed as an exponential (range: 21% to 77%), chi-squared (range: 14% to 85%), log-normal (range: -56% to 39%), power-function (range: -66% to 3%) or an asymmetric mixture of two normals (range: -75% to 14%).

between the true and assumed distribution, the distribution of the best predicted values is highly sensitive to misspecification of the assumed random effects distribution (McCulloch and Neuhaus, 2011b). As identified in the context of linear mixed models (Verbeke and Lesaffre, 1996), the distribution of the predicted random effects in GLMMs reflects the assumed, rather than the true underlying shape of the random effects (Agresti et al., 2004; McCulloch and Neuhaus, 2011a,b). Thus, the distribution of the best predicted values should not be considered a reliable indicator of the true underlying distribution (McCulloch and Neuhaus, 2011b).

2.6.6 Misspecification of the random effects distribution in the presence of missingness

The ambiguity about the impact of misspecification has been further exacerbated by the lack of investigation of complexities prevalent in longitudinal data, such as missing data due to attrition. Under the assumption of MCAR or MAR attrition, maximum likelihood estimation of GLMMs can provide consistent estimation. However, this assumes that other aspects of the model are correctly specified, including the random effects distribution. This is an important area of study as bias has been reported when missing data results in informative cluster sizes (Neuhaus and McCulloch, 2011). Recently, Wang (2010b) investigated the impact of incorrectly assuming bivariate normally distributed random effects for logistic mixed models in the presence of missing data[23]. Four missingness scenarios were considered, consisting of a combination of either a MCAR or MAR underlying mechanism, for either intermittent missingness or attrition. Missingness did not have any additional impact on estimation or inference of the logistic mixed model when the random effects distribution was misspecified (Wang, 2010b). However, missingness did affect the power to detect variance components when the true random effects were positively skewed or positively skewed and leptokurtic, though not for symmetric random effects (Wang, 2010b). More research considering random effect distribution misspecification in the presence of attrition is required, particularly within applications of panel surveys whereby the loss of respondents due attrition is an inherent problem.

2.6.7 Summary

This literature review has shown that misspecification of the random effects distribution in logistic mixed models can impact estimation and inference of parameters relating to fixed and random effects. Previous literature considering the impact of distributional misspecification of the random effects in GLMMs has predominately focused on biomedical settings, particularly considering simulation studies based on clinical trial studies (e.g. Chen et al. 2002; Litiére et al. 2008; McCulloch and Neuhaus 2011a; Neuhaus et al. 2013). Previous studies have not considered misspecification within panel survey settings. The challenges of longitudinal data

[23] Wang (2010b) considered four different types of misspecification in the presence of missing data, including: incorrectly assuming normality for non-normal bivariate random effects, omitting a fixed effect from the mean structure, omitting a quadratic time effect, and ignoring a higher level of nesting.

are expected to be similar between the two settings, however there are potentially some differences that lead to gaps in the literature regarding the impact of misspecifying the random effects distribution.

Firstly, as household panel data are often collected using questionnaires, the type of variables collected are predominately categorical. However, limited research has considered the impact of misspecification on estimating categorical explanatory variables, particularly time-varying categorical variables. Furthermore, analysis of panel surveys typically includes a larger number of explanatory variables, however previous simulation studies within the biomedical setting often consider a limited number of explanatory variables (Table 2.1).

Secondly, as data from panel surveys is typically collected annually, the time between sampling is usually longer than in clinical trial settings and can result in potentially different underlying missingness patterns. As mentioned in Section 2.6.6, although longitudinal studies are typically challenged by the issue of missing data, only limited research has investigated the impact of misspecified random effect distributions in the presence of missing data, particularly attrition.

Thirdly, by the nature of collecting data from a panel of subjects, often the process under investigation has already initiated. In this situation, the initial observed response is typically dependent upon unobserved previous responses and unobserved variables (Crouchley and Davies, 2001). Not only will there be substantial heterogeneity in the underlying random effects, but sub-populations may exist. However substantial heterogeneity is not only confined to the panel survey setting, as clinical trials may have different response patterns depending on the treatment[24]. As mentioned in Section 2.5, multimodality of the random effects may occur in situations that are dominated by constant response profiles. However, no studies have considered the impact of misspecification in scenarios whereby the true distribution is multimodal with three or more modes.

2.7 Addressing misspecification of the random effects distribution in generalised linear mixed models

Two strands of research have emerged to address misspecification of the random effects distribution: extensions of GLMMs to flexibly model the random effects distribution, and diagnostic tests to detect distributional misspecification. A brief literature review of these two strands of research are detailed in Sections 2.7.1 and 2.7.2, respectively.

[24]As identified by Litiére et al. (2008), little variability in the responses may occur for a placebo control group, while more variability in the response may be expected in the treatment group.

2.7.1 Flexibly modelling the assumed random effects distribution

By allowing for increased flexibility of the assumed random effects distribution, the risks associated with misspecification can be reduced (Verbeke and Molenberghs, 2013). A suite of methodology has been developed to relax the normality assumption of the random effects distribution. Computational methods such as the probability integral transformation (PIT) (Nelson et al., 2006) and the likelihood reformulation (LR) (Liu and Yu, 2008) have been developed to obtain maximum likelihood estimates for GLMMs with non-Gaussian random effect distributions. Both of these approaches are based on transforming normally distributed random effects to non-normal distributions, and provide an easily implementable method to not only model random effects as standard parametric distributions, but as described below, as mixtures of parametric distributions. Furthermore, methods to fit parametric classes of densities for the random effects distribution have been developed including the class of t-distributions (Lee and Thompson, 2008) and skew extensions of the t-distribution or normal distributions (Ho and Lin, 2010). However, as standard parametric distributions and parametric classes are generally not sufficiently flexible to capture multimodal distributions, the following literature review will focus on approaches that may be suitable to capture heterogeneity inherent in the mover-stayer scenario.

One approach to induce flexibility of the assumed distribution and capture multimodality is to assume the random effects arise from a mixture of parametric distributions, such as a finite mixture of normal distributions (Verbeke and Lesaffre, 1996; Magder and Zeger, 1996; Molenberghs and Verbeke, 2005). The heterogeneity model developed by Verbeke and Lesaffre (1996) assumes that the random effects population consists of g sub-populations, such that the random effects are modelled as a mixture of g normal distributions to capture the heterogeneity. Furthermore, the model can be used to classify subjects into different components based on longitudinal profiles (Verbeke and Molenberghs, 2009). Estimation of the heterogeneity model is based on the maximization of the likelihood using the Expectation-Maximization (EM) algorithm (Dempster et al., 1977), or by the aforementioned LR or PIT computational methods. Details regarding the methodology and estimation of the heterogeneity model are provided in Section 3.2.1.

Extensions of the heterogeneity model have been developed, predominately within the Bayesian framework. One approach is a penalized Gaussian mixture distribution where the weights of the mixture components are estimated using a penalized approach and parameters of the model are estimated using Markov Chain Monte Carlo (MCMC) techniques (Komarek and Lesaffre, 2008). Another approach fits an infinite mixture model within the Bayesian framework by incorporating a Dirichlet process mixture of a normal prior as the random effects distribution (Jara et al., 2007). These approaches will not be considered further, but highlight the feasibility of Bayesian techniques to estimate the heterogeneity model.

Alternatively, flexibility may be achieved by assuming that the random effects belong to a smooth class of densities represented by the semi-non-parametric (SNP) formulation (Gallant and Nychka, 1987). Chen et al. (2002) and Vock et al. (2014) have developed approaches that utilise the SNP class of densities to fit random effects in GLMMs, and these have been shown to be sufficiently flexible to capture a range of densities including skewed, multimodal, and thick- or thin-tailed densities (Vock et al., 2014). Furthermore, by assuming a smooth density, the SNP representation of the random effects distribution can also provide an estimate of the underlying distribution (Vock et al., 2014). Details regarding the methodology of approaches utilising the SNP formulation are provided in Section 3.2.2.

An approach to allow an immense degree of flexibility is to leave the random effects distribution completely unspecified, and estimate the random effects distribution using non-parametric techniques[25]. These approaches are referred to as semi-parametric models, as the mean structure of the GLMM is parametric and is estimated using maximum likelihood, whilst the random effects distribution is estimated non-parametrically. Computational approaches to obtain the non-parametric maximum likelihood (NPML) estimator of the random effects distribution have been proposed (Laird, 1978; Heckman and Singer, 1984; Follmann and Lambert, 1989; Lesperance and Kalbfleisch, 1992; Aitkin, 1999; Rabe-Hesketh et al., 2003; Wang, 2010a; Lesperance et al., 2014), and result in a discrete distribution on a finite number of support points[26] (Lindsay, 1983). However, the NPML estimated locations and probability weights of the support points do not represent an underlying sub-population structure (Davies, 1993). Rather, the empirically determined support points provide adequate flexibility to capture the underlying random effects distribution and consistent estimation of the parameters in the mean structure of the GLMM (Davies, 1993).

Methods to obtain the NPML estimator in GLMMs vary in the underlying approach, and are detailed in Section 3.2.3 and Section 7.1. The majority of approaches to determine the NPML estimator have been restricted to GLMMs with a single random effect, with few methods developed for GLMMs with multiple random effects (e.g. Lesperance et al. 2014). In a similar context to GLMMs, Tsonaka et al. (2009) applied the Vertex Exchange Method (VEM) of Böhning (1985) to estimate the unspecified distribution of the multivariate random effects in shared parameter models. For linear mixed effects models, Baghfalaki and Ganjali (2014) proposed a computationally fast algorithm to provide an approximation to the VEM approach. The simplicity and statistical properties of VEM make it an appealing non-parametric method to estimate the random effects distribution. However, application of the VEM to flexibly model

[25] Latent class models can also be seen as a non-parametric approach, as the distribution of the random effects is assumed to be discrete with unknown probability masses and locations. In the latent class model individuals are grouped into clusters, known as latent classes, where each cluster has the same value for the random effects. Additionally, the latent class model can be considered as an extension of the heterogeneity model by allowing the components to have zero variance. For more details the reader is referred to Muthén (2004).

[26] In order to achieve the NPML estimator, the optimal number of finite support points needs to be determined. Alternatively, if the number of support points is considered to be known apriori, then it is referred to as a discrete random effects distribution estimated non-parametrically. This is discussed further in Sections 3.2.3 and 7.1.

the random effects in GLMMs has yet to be utilised.

Often the discrete nature of the non-parametric technique is considered a key limitation, particularly if interest is in estimating the random effects distribution (Butler and Louis, 1992; Vock et al., 2014). To overcome the discreteness of NPMLE, smooth non-parametric maximum likelihood estimators resulting in a continuous density have been proposed, whereby the smoothing is obtained using finite mixtures of Gaussian distributions (Magder and Zeger, 1996)[27], kernel methods (Knott and Tzamourani, 2007) or methods using a predictive recursive algorithm (Tao et al., 1999), however the degree of smoothness is often arbitrary.

As mentioned in Section 2.5, inducing flexibility to the random effects distribution has been previously considered in the context of the latent mover-stayer scenario. For instance, Barry et al. (1989) incorporated spikes at negative and positive infinity to the normal distribution for the random effects to account for the two groups of latent stayers. Previously Davies (1993) reported the similarities between the traditional mover-stayer models and non-parametric estimation of random effects. Following from the question raised by Davies and Crouchley (1986), Davies (1993) argued that the perceived goodness-of-fit of the mover-stayer model could be due to it sufficiently approximating the non-parametric estimation of the random effects distribution[28]. However, the performance of estimating the random effects in logistic mixed models with a set of explanatory variables using non-parametric techniques has not been investigated as an appropriate modelling strategy to account for potential mover-stayer scenarios.

This literature review has shown that a variety of methods is available to flexibly model the random effects distribution in GLMMs, potentially reducing the impact of misspecifying the assumed distribution on model based inference and predictions. Although a suite of methodology has been developed, implementation in practical applications is limited and the choice of approach may be dependent on the inferential focus of the random effects distribution (Vock et al., 2014). Often the price one pays for inducing flexibility of the random effects distribution is heavy computation (Huang, 2011), however with the ever increasing gains in computational power the issue of computation burden continually declines. Recently, Ghidey et al. (2010) reviewed four approaches to flexibly model the random effects distribution in linear mixed models. However, the practicality of implementing approaches to flexibly model the random effects in GLMMs has not been demonstrated, particularly in applications of longitudinal panel surveys, including settings in the presence of attrition.

[27]The smooth non-parametric maximum likelihood estimator for linear mixed models as proposed by Magder and Zeger (1996) resulted in a finite mixture of normal densities as the random effects distribution. However, unlike the heterogeneity model, the number of components is considered an unknown parameter and is subsequently estimated.

[28]In the context of GLMMs for binary responses, Davies (1993) restricts attention to a random intercept logistic model with no explanatory variables.

2.7.2 Diagnosing misspecification of the assumed random effects distribution

Detecting misspecification of the distributional assumptions of the random effects is far from straightforward (Efendi et al., 2014). This is an area of research that has recently attracted considerable attention in the literature, with several informal and formal diagnostic tools developed to assess the validity of the assumed random effects distribution in GLMMs. Generally the diagnostic tools do not have any restrictions on the type of GLMM, thus the following review focuses on the entire class of GLMMs, however all are directly implementable for the logistic mixed model. A brief review of informal and formal diagnostic tools is presented in Sections 2.7.2.1 and 2.7.2.2, respectively.

2.7.2.1 Informal diagnostic tests

Informal diagnostic tools to assess the validity of the assumed random effects distribution have been proposed, including graphical based tools based on the predicted random effects (e.g. Lange and Ryan, 1989). However, as the predicted random effects are highly sensitive to the assumed form of the distribution[29] (Verbeke and Lesaffre, 1996; Molenberghs and Verbeke, 2005; McCulloch and Neuhaus, 2011a), visual inspection of the predicted values (normal probability plots and histograms) should not be used to assess the adequacy of the normality assumption of the random effects distribution (Molenberghs and Verbeke, 2005; Verbeke and Molenberghs, 2009; McCulloch and Neuhaus, 2011a).

Recently Verbeke and Molenberghs (2013) proposed a simple graphical exploratory diagnostic tool utilising the gradient function to investigate the adequacy of the assumed random effects distribution in terms of the marginal likelihood. Not only does the tool provide evidence of misspecification of the random effects distribution, the shape of the gradient function can provide an insight into how the assumed random effects distribution can be improved to provide a better fit to the observed data (Verbeke and Molenberghs, 2013). However as the gradient function graphical tool only uses information from people with non-constant response profiles, this can provide limited evidence for binary and categorical response data with constant response profiles. Details about the methodology are presented in Section 3.3.1.

Alternatively, sensitivity analyses can provide a practical tool to informally assess the robustness of the assumed random effects distribution (Litiére et al., 2008). For instance, Agresti et al. (2004) suggested comparing the parameter estimates from a GLMM assuming normally distributed random effects and a GLMM with an unspecified distribution estimated using non-parametric techniques. Substantial differences between the parameter estimates from the two approaches would indicate a specification issue of the parametric model (McCulloch et al., 2008). Similarly, a more general sensitivity analysis framework would consider different choices

[29] In the context of linear mixed models, Verbeke and Lesaffre (1996) demonstrate that due to shrinkage the empirical Bayes predictions appear to take on the form of the assumed random effects distribution.

for the random effects distribution, such as alternative parametric distributions (McCulloch and Neuhaus, 2013) in addition to flexible approaches (Litiére et al., 2008). If the parameter estimates and standard errors differ considerably, indicative of sensitivity to the distributional assumptions of the random effects distribution, then inference of the model results should be interpreted with caution (Litiére et al., 2008). In a similar context to sensitivity analyses, Chen et al. (2002) informally test the normality of random effects in GLMMs by using information criteria to determine the optimal order of the semi-non-parametric density. A non-zero value of the optimal order would suggest non-normality in the underlying random effects distribution. Similarly, model selection techniques can be used to determine the optimal number of mixture components in the heterogeneity model (Verbeke and Lesaffre, 1996; Molenberghs and Verbeke, 2005). An optimal number of more than one component suggests potential non-normality of the random effects distribution. However, due to the complex nature of model comparisons in GLMMs, particularly when comparing models with flexible distributions for the random effects, model selection[30] may not be straightforward (Agresti et al., 2004).

2.7.2.2 Formal diagnostic tests

Implementation of formal diagnostic tests to detect violations from normality of the predicted random effects, such as the Kolmogorov-Smirnov, Anderson-Darling and Cramer-von Mises normality tests, have been advocated (e.g Hosmer et al. 2013)[31]. However, as mentioned previously, the predicted values of the random effects are sensitive to the assumed distribution. Therefore the performance of the aforementioned normality tests to diagnose departures from the normal distribution are invalid (McCulloch and Neuhaus, 2011a).

A range of diagnostic tests generally focusing on detecting misspecification of the random effects structure has been proposed and they are briefly reviewed here[32]. To examine the adequacy of the assumed random effects distribution in GLMMs with canonical links, Tchetgen and Coull (2006) proposed a diagnostic test based on the difference between marginal and conditional ML estimation of time-varying explanatory variables. In a similar context, diagnostic tests have been proposed that compare estimates between two approaches. For instance, Waagepetersen (2006) proposed a simulation-based test by generating random effects conditional on the observations. Similarly, Huang (2009) proposed a two-step parametric diagnostic test based on comparing the parameter estimates based on the observed data and reconstructed data to detect misspecification of the random effects structure in GLMMs for

[30]The choice of the most appropriate random effects distribution is one element of selecting a model (Agresti et al., 2004). Other aspects of model selection, such as variable selection for the fixed and random effects and number of random effects, is beyond the scope of this review. For further details regarding model selection for multilevel models the reader is referred to Steele (2013), and for a review about model selection in the simpler (yet related) case of linear mixed models the reader referred to Müller et al. (2013).

[31]Hosmer et al. (2013) (page 367-368) state that "(t)he best method for assessing the normality assumption of a random effect is based on the best predicted values of the random effects. One can then use standard tests and plots for normality such as the normal probaility (PP) or normal quantile (QQ) plots."

[32]For a more detailed review, the reader is referred to Verbeke and Molenberghs (2013) and the references within.

binary responses. This diagnostic test was later extended by Huang (2011) to be applied to GLMMs and non-linear mixed models, and by Lin and Chen (2015) to cumulative logit mixed models for ordinal responses. Additionally, diagnostic tests based on information equivalence under a correct model have been proposed. Alonso et al. (2008) proposed three diagnostic tests based on eigenvalues of the variance-covariance matrix to detect misspecification of the random effects structure in GLMMs. Similarly, Alonso et al. (2010b) proposed two diagnostic tests based on the information matrix test (White, 1982) to detect misspecification[33] in GLMMs.

More recently, tests based on the gradient function have been proposed by Efendi et al. (2014) and Drikvandi et al. (2016) to diagnose misspecification of the parametric assumption of the random effects distribution. Both methods have been proposed to complement the informal graphical approach developed by Verbeke and Molenberghs (2013) (Section 2.7.2.1), and test whether the fluctuations observed in the gradient function graphical tool are due to distributional misspecification of the random effects and not just random variability. Efendi et al. (2014) propose a bootstrap test based on the gradient function, however it is restricted to evaluating the gradient within an interval. Therefore, for binary response data, the diagnostic test of Efendi et al. (2014) is restricted to those subjects with non-constant response profiles. To provide a formal diagnostic test based on the gradient function across the whole support of the random effects distribution, Drikvandi et al. (2016) propose and derive the asymptotic properties of a test statistic that utilises the Cramer-von Mises measure. Further details about the diagnostic test of Drikvandi et al. (2016) are presented in Section 3.3.2.

This literature review has shown a range of informal and formal diagnostic tests that have been developed to identify potential misspecification of the random effects distribution. Albeit continual developments of formal diagnostic tools, utilisation in practice is limited, particularly within social science applications. The practicality of diagnostic tools has not been reviewed for applications to longitudinal panel data. Furthermore, the aforementioned diagnostic methods generally assume only one type of misspecification is present. As highlighted by Huang (2011), future work is required to develop diagnostic methods that can disentangle multiple sources of misspecification in GLMMs such as the structure of the mean model, link function or variance-covariance structure.

2.8 Overview

This literature review has introduced longitudinal panel surveys, highlighting some of the complexities that need to be considered when analysing longitudinal panel data (Section 2.1), including the loss of respondents due to attrition (Section 2.2). As highlighted in Section 2.3, although there are numerous approaches to estimate changes in longitudinal binary response data, logistic mixed models provide a flexible framework to obtain subject specific interpre-

[33]The diagnostic tests can detect misspecification of the random effects structure, and other types of model misspecification, such as a misspecified link function or mean structure.

tation. These models not only handle the correlated nature of longitudinal data and account for unobserved heterogeneity, but can provide consistent estimation in settings with attrition following the MAR missingness mechanism. Furthermore, consistent estimation requires the correct specification of the model, including the mean structure and the random effects structure (Section 2.4).

However, as highlighted in Section 2.5, the commonly assumed Gaussian distribution for the random effects distribution can be too restrictive in practice. This is particularly true if there exists a sub-population structure. In the context of longitudinal binary responses, individuals with constant response profiles may exhibit heterogeneity in the random effects distribution due to an underlying mover-stayer scenario. In this setting, the assumed normal distribution may not sufficiently capture the multimodality of the true distribution.

Previous literature has demonstrated that misspecifying the random effects distribution can impact inferential conclusions in logistic mixed models, and that the impact may be dependent on the inferential focus of the researcher (Section 2.6). However, no study has considered specific departure from the assumed normal distribution characterised by a trimodal distribution, and settings with substantial heterogeneity due to an underlying sub-population structure. Furthermore, no literature has considered this type of distributional misspecification in the longitudinal panel settings, including in the presence of attrition. The literature has predominately focused on biomedical settings, with limited research considering the impact of misspecification on estimating the effects of time-varying categorical explanatory variables.

A body of research has investigated potential ways to address misspecification of the random effects distribution, including methods to relax the parametric assumption of the random effects distribution (Section 2.7.1) and diagnostic tools to identify potential distributional misspecification (Section 2.7.2). These proposed methods provide a promising framework to determine the sensitivity of model based conclusions to distributional assumptions for the random effects. However, the implementation of flexible modelling techniques and diagnostic tests in practical applications of longitudinal panel survey data is limited. No review has explored the performance of these approaches, including the practicality when utilised in settings with missing data due to attrition.

2.9 Research questions

This study will contribute to the statistical and social sciences literature by focusing on evaluating assumptions and extensions of logistic mixed models to analyse longitudinal panel survey data. In particular, this study focuses on the specific departure from normality of the random effects distribution characterised by multimodality, reflecting an underlying sub-population structure such as the mover-stayer scenario.

The gaps in the literature have led to the following three research questions that this study aims to address:

1. How robust is the assumed Gaussian distribution to multimodality of the random intercept distribution in panel survey settings due to potential mover-stayer scenario?

2. Can the impact of multimodality of the random effects distribution be alleviated by increasing the flexibility of the assumed random effects distribution?

3. What is the additional impact of misspecified random effect distributions in the presence of missing data due to attrition?

The first research question will investigate the impact of misspecifying the assumed random intercept distribution in random intercept logistic models through two simulation studies presented in Chapters 5 and 6. Motivated by the underlying distribution observed in the HILDA case study (Chapter 4), the first simulation study in Chapter 5 considers the specific departure from normality arising from a three component mixture of Gaussians to represent the mover-stayer scenario. Chapter 6 presents the second simulation study to investigate the robustness of model based inference to misspecification of the random intercept distribution by simulating a range of trimodal distributions increasing in severity of departure from the assumed normal distribution.

The second research question will be explored by assessing the performance and feasibility of implementing approaches to induce sufficient flexibility of the assumed random effects distribution in logistic mixed models with univariate and bivariate random effects. Chapter 4 induces more flexibility to the random intercepts distribution by considering random intercepts that arise from a three component mixture of normal distributions. In addition to this heterogeneity model, Chapter 7 considers the end-point model, semi-non-parametric approach and two non-parametric approaches. Furthermore, Chapter 7 considers the performance of these approaches to the more complex scenario of flexibly modelling bivariate random effects distribution in a random intercept and random slope logistic model.

Finally, the third research question will be explored through the simulation studies presented in Chapters 5 and 6, and by assessing the performance of the logistic mixed models applied to settings with missing data due to attrition. The simulation studies investigate the robustness of the normality assumption in two data scenarios: the first considers misspecification in the presence of complete data, and the second considers misspecification in the presence of missing data following from MAR attrition. Chapters 4 and 7 assess the performance of logistic mixed models in the presence of missing data when assuming a normal distribution or using approaches to flexibly model the random effects distribution.

3 Statistical Methodology

The focus of this study is on investigating the impact of departures from distributional normality in the random effects of generalised linear mixed models (GLMMs). The statistical methodology relevant to the study is described in this Chapter. First, the statistical framework of GLMMs is introduced (Section 3.1), with focus on the special case of logistic mixed models for binary response variables (Section 3.1.1). The distribution of the random effects in a GLMM is typically assumed to be normal, but this can be restrictive in some applications. In Chapters 4 and 7, four alternative methods are investigated to relax the normality assumption and flexibly model the random effects of a logistic mixed model. These four methods are described in Section 3.2 and include the heterogeneity model (Section 3.2.1), the semi-non-parametric model (Section 3.2.2), non-parametric estimation (Section 3.2.3) and GLMM with endpoints (Section 3.2.4). As inference for parameters of GLMMs can be sensitive to the assumed random effects distribution, diagnostic tests to detect distributional misspecification of the random effects have recently been developed. In Chapter 4, two diagnostic tests are utilised to identify potential misspecification of the random effects distribution when applied to a random intercept logistic model. These two diagnostic tests are described in Section 3.3, and include the gradient function exploratory diagnostic tool (Section 3.3.1) and the asymptotic diagnostic test based on the gradient function (Section 3.3.2). Typically, simulation studies are utilised to assess the robustness of inference in applications of GLMMs with potentially incorrect assumptions for the random effects distribution. Chapters 5 and 6 use simulation studies to investigate the impact of incorrectly assuming normality for the random intercept distribution in panel survey settings when the true distribution is multimodal. The simulation study design utilised in Chapters 5 and 6 is described in Section 3.4, and the details of the performance measures used to assess and summarise the results from the simulation studies are described in Section 3.5.

3.1 Generalised linear mixed models

In this section the framework for generalised linear mixed models is described. Throughout the thesis discussion is restricted to two-level models focusing on longitudinal designs, where time varying observations (level one) are clustered by a higher level unit (level two). However the framework can easily be extended to higher order clustered designs.

In a longitudinal panel survey consisting of N individuals, an individual i has observations measured repeatedly n_i times, such that y_{ij} denotes the response for individual i (for $i =$

$1, ..., N)$ at time j (for $j = 1, ..., n_i$). For each individual and time-point, data for a set of p explanatory variables denoted by $\boldsymbol{x_{ij}}' = (x_{1ij}, ..., x_{pij})$ are recorded. Often time is expressed as wave number in panel surveys, and the time interval between waves can vary.

For normal and non-normal responses, generalised linear mixed models extend the generalised linear model (Nelder and Wedderburn, 1972) to incorporate random effects. It is assumed that, conditionally on the q-dimensional random effects, $\boldsymbol{b_i}$, the responses y_{ij} are independent with densities of the form:

$$f_Y(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi) = \exp\left[\phi^{-1}\left\{y_{ij}\boldsymbol{\theta_{ij}} - c(\boldsymbol{\theta_{ij}})\right\} + d(y_{ij}, \phi)\right] \tag{3.1}$$

where c is a specific function depending on the type of exponential family, d is the log-normalisation constant, ϕ is a dispersion parameter and $\boldsymbol{\theta_{ij}}$ models the mean response, μ_{ij}, through a linear predictor containing fixed regression coefficients, $\boldsymbol{\beta}$ and individual-specific random effects, $\boldsymbol{b_i}$:

$$\eta(\mu_{ij}) = \eta[E(y_{ij}|\boldsymbol{b_i})] = \boldsymbol{x'_{ij}\beta} + \boldsymbol{z'_{ij}b_i} \tag{3.2}$$

where $\eta(\cdot)$ is a known link function, and $\boldsymbol{x_{ij}}$ and $\boldsymbol{z_{ij}}$ are $p-$ and q-dimensional vectors of explanatory variables of the fixed and random effects, respectively. These models are conditional on the explanatory variables, and for ease of readability the dependence on $\boldsymbol{x_{ij}}$ will be suppressed from notation. It is assumed that the random effects are sampled from a population of individual-specific parameters with distribution function G, parameterised by the vector $\boldsymbol{\xi}$. Typically the random effects are assumed to be sampled from a q-dimensional multivariate normal distribution with zero mean and variance-covariance matrix $\boldsymbol{\Sigma}$, $\boldsymbol{b_i} \sim N_q(\boldsymbol{0}, \boldsymbol{\Sigma})$[1]. Furthermore, it is assumed that the random effects are uncorrelated with the covariates, such that $\boldsymbol{b_i}$ and $\boldsymbol{x_{ij}}$ are exogenous.

The maximum likelihood estimation of GLMMs is obtained by maximising the marginal likelihood, which requires integrating over the distribution of the random effects, such that the likelihood contribution for individual i is

$$f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi) = \int \prod_{j=1}^{n_i} f_{ij}(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi) f(\boldsymbol{b_i}|\boldsymbol{\Sigma}) db_i. \tag{3.3}$$

The likelihood is derived as

$$L(\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi) = \prod_{i=1}^{N} f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \phi). \tag{3.4}$$

[1]The assumption of normality for the random effects distribution has been carried over from linear mixed models, where it is mathematically convenient to analytically calculate the marginal likelihood (Molenberghs and Verbeke, 2005). Furthermore, a continuous distribution for the random effects is often considered more realistic than a discrete distribution, thus, the normal distribution is a common choice. Additionally, assuming a normal distribution for the random effects is typically the default in standard software packages.

However, the integral in the likelihood function typically does not have a closed form[2], making estimation more difficult (Skrondal and Rabe-Hesketh, 2008). Therefore, numerical approximation techniques for the integral are required to estimate GLMMs, such as adaptive or non-adaptive Gaussian quadrature. Optimization routines to maximise the likelihood function use optimization techniques such as the Newton-Raphson algorithm, Fisher-scoring algorithm, iterative generalized least squares or restrictive generalized least squares.

3.1.1 Logistic mixed models

Binary outcomes are common in the social sciences. For a clustered binary response, let y_{ij} be the value of the dichotomous variable, coded 0 or 1, for individual i at time j. The logistic mixed model is a widely accepted method for describing the relationship between a clustered binary response variable and explanatory variables. The logistic mixed model is written in terms of the log odds of the probability of response, denoted $p_{ij} = \Pr(y_{ij} = 1|\boldsymbol{b_i})$ and is equivalent to the GLMM with a logit link function, such that $\eta(\mu_{ij}) = \text{logit}(p_{ij}) = \log\left(\frac{p_{ij}}{1-p_{ij}}\right)$.

First consider a random intercept logistic model, that is, a logistic mixed model with a single random effect,

$$\log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \boldsymbol{x'_{ij}}\boldsymbol{\beta} + b_{0i} \tag{3.5}$$

where b_{0i} is the random intercept for individual i and is assumed to be distributed in the population as $N(0, \sigma^2_{b_0})$. The random intercept captures each individual's deviation from the overall intercept constant (β_0). Often it is of interest to express the random intercept variability in terms of the intraclass correlation (ICC). The ICC is a measurement indicating the proportion of unexplained variance at the individual level. For a random intercept logistic model assuming normally distributed random effects, the ICC can be estimated as

$$\hat{\rho} = \frac{\hat{\sigma}^2_{b_0}}{\hat{\sigma}^2_{b_0} + \frac{\pi^2}{3}} \tag{3.6}$$

where the latter term in the denominator is the variance of the underlying latent response tendency, and in a logistic model is equal to $\frac{\pi^2}{3}$ (Page 241 of Andress et al. 2013).

The random intercept logistic model is easily extended to include multiple random effects ($q \geq 2$), however we restrict our attention to random intercepts and random slopes ($q = 2$). Within the panel survey context, the random slopes are often the random coefficient for the time variable, measuring each individual's deviation from the overall wave trend. The corresponding logistic mixed model with bivariate random effects, also known as a random intercept and

[2]Closed form likelihoods can be achieved for specific cases, such as, the linear mixed model, random intercept Poisson model with a conjugate random effects distribution, and GLMMs with discrete random effects.

random slope logistic model, is given by:

$$\text{logit}[\Pr(y_{ij}=1|b_i)] = \log\left[\frac{p_{ij}}{1-p_{ij}}\right] = \boldsymbol{x'_{ij}\beta} + \boldsymbol{z'_{ij}b_i} \tag{3.7}$$

where $\boldsymbol{b_i}$ is a two dimensional vector consisting of the random intercepts (b_{0i}) and the random slopes (b_{1i}), and is assumed to follow a bivariate normal distribution $N_2(\mathbf{0}, \boldsymbol{\Sigma_b})$. The vector $\boldsymbol{z_{ij}}$ will consist of a term for the intercept and the corresponding wave term for individual i at the j^{th} wave.

3.2 Generalised linear mixed models with flexible random effects distributions

The random effects in GLMMs are typically assumed to be normally distributed. However, in some practical applications, the normality assumption may be too restrictive to adequately capture the distribution of the underlying heterogeneity. It is possible for the normality assumption of the random effects distribution to be relaxed, and alternative specifications include non-normal parametric or non-parametric distributions. In this section four alternative methodologies that flexibly capture departures from the normality assumption are described. Flexibility can be achieved by either modelling the random effects as a finite mixture of normal distributions (Section 3.2.1), or utilising semi-non-parametric (Section 3.2.2) or non-parametric (Section 3.2.3) estimation, or by incorporating endpoints at the distributional extremes to capture a potential mover-stayer scenario (Section 3.2.4).

3.2.1 Heterogeneity model

In the context of linear mixed models, Verbeke and Lesaffre (1996) and Magder and Zeger (1996) proposed to relax the normality assumption by specifying a finite mixture of multivariate normals for the random effects distribution. As shown by Molenberghs and Verbeke (2005) this specification can easily be extended to the GLMM context. The so called heterogeneity model (Verbeke and Lesaffre, 1996) assumes that the population under study consists of g sub-populations. To model the heterogeneity, the random effects are modelled as a mixture of g $q-$dimensional normal distributions, with mean vectors $\boldsymbol{\mu_\gamma}$ and covariance matrices $\boldsymbol{\Sigma_\gamma}$, where $\gamma = 1, ..., g$. Therefore, the random effects are given by

$$\boldsymbol{b_i} \sim \sum_{\gamma=1}^{g} \pi_\gamma N(\boldsymbol{\mu_\gamma}, \boldsymbol{\Sigma_\gamma}) \tag{3.8}$$

where π_γ are the mixing proportions, representing the proportion of the individuals in the total population belonging to the γ^{th} sub-population ($\sum_{\gamma=1}^{g} \pi_\gamma{=}1$). Thus, the density of the random effects is given by

$$\begin{aligned} f(\boldsymbol{b_i}) &= \sum_{\gamma=1}^{g} \pi_\gamma f_\gamma(\boldsymbol{b_i}) \\ &= \sum_{\gamma=1}^{g} \pi_\gamma (2\pi)^{-\frac{g}{2}} |\boldsymbol{\Sigma_\gamma}|^{-\frac{1}{2}} \times \exp\left\{ -\frac{1}{2}(\boldsymbol{b_i} - \boldsymbol{\mu_\gamma})' \boldsymbol{\Sigma_\gamma}^{-1} (\boldsymbol{b_i} - \boldsymbol{\mu_\gamma}) \right\} \end{aligned} \tag{3.9}$$

where $f_\gamma(\boldsymbol{b_i})$ represents the density of the γ^{th} mixture component.

Let $\delta_{i\gamma} = 1$ if the random effects $\boldsymbol{b_i}$ are sampled from the γ^{th} mixture component and 0 otherwise, such that $\pi_\gamma = \Pr(\delta_{i\gamma} = 1) = E(\delta_{i\gamma})$. The overall mean of the random effects is given by

$$\begin{aligned} E(\boldsymbol{b_i}) &= E\left(E(\boldsymbol{b_i}|\delta_{i1}, ..., \delta_{i\gamma})\right) \\ &= E\left(\sum_{\gamma=1}^{g} \boldsymbol{\mu_\gamma} \delta_{i\gamma}\right) \\ &= \sum_{\gamma=1}^{g} \pi_\gamma \boldsymbol{\mu_\gamma}. \end{aligned} \tag{3.10}$$

Furthermore, the overall covariance matrix of the random effects is given by

$$\begin{aligned} \boldsymbol{\Sigma^*} &= \mathrm{Var}[E(\boldsymbol{b_i}|\delta_{i1}, ..., \delta_{i\gamma})] + E[\mathrm{Var}(\boldsymbol{b_i}|\delta_{i1}, ..., \delta_{i\gamma})] \\ &= \mathrm{Var}\left(\sum_{\gamma=1}^{g} \boldsymbol{\mu_\gamma} \delta_{i\gamma}\right) + E\left(\sum_{\gamma=1}^{g} \boldsymbol{\Sigma_\gamma} \delta_{i\gamma}\right) \\ &= \sum_{\gamma=1}^{g} \pi_\gamma \boldsymbol{\mu_\gamma} \boldsymbol{\mu'_\gamma} + \sum_{\gamma=1}^{g} \pi_\gamma \boldsymbol{\Sigma_\gamma}. \end{aligned} \tag{3.11}$$

The resulting marginal density of the response measurements $\boldsymbol{y_i}$ is a g component mixture of marginal mixed effect models with mixing proportions π_γ:

$$\begin{aligned} f_i(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\mu}, \phi) &= \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi) f(\boldsymbol{b_i}) db_i \\ &= \sum_{\gamma=1}^{g} \pi_\gamma \int f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\beta}, \phi) f_\gamma(\boldsymbol{b_i}) db_i \\ &= \sum_{\gamma=1}^{g} \pi_\gamma f_{i\gamma}(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\Sigma_\gamma}, \boldsymbol{\mu_\gamma}, \phi) \end{aligned} \tag{3.12}$$

where $f_{i\gamma}(\boldsymbol{y_i}|\boldsymbol{\beta}, \boldsymbol{\Sigma_\gamma}, \mu_\gamma, \phi)$ is the marginal density corresponding to a generalised linear mixed model with random effects distributed as a $q-$dimensional normal distribution from the γ^{th} mixture component with mean $\boldsymbol{\mu_\gamma}$ and covariance matrix $\boldsymbol{\Sigma_\gamma}$. Furthermore, let $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma_1}, ..., \boldsymbol{\Sigma_g})'$ and $\boldsymbol{\mu} = (\boldsymbol{\mu_1}, ..., \boldsymbol{\mu_g})'$.

The heterogeneity model assumes that the number of components, g, is known. In practice several models of increasing g values can be fit, and model selection for nested models, such as the likelihood ratio test (Molenberghs and Verbeke, 2005), can be used to determine the optimal number of components[3].

3.2.1.1 Estimation

The parameters of the heterogeneity model can be estimated using maximum likelihood methods. As shown in Liu and Yu (2008), and implemented by Verbeke and Molenberghs (2013), the heterogeneity model can be estimated using the likelihood reformulation (LR) method (Liu and Yu, 2008). The LR method is based on a simple transformation to replace the conditional density on a non-normal random effects distribution by another one that can be integrated over a normal random effects distribution. The transformation used in the LR method reformulates the conditional likelihood on non-normal random effects by multiplying and dividing the conditional likelihood by a standard normal density and then reformulating the resulting likelihood for integration over normal distributed random effects using adaptive Gaussian quadrature. The LR method can be used to estimate GLMMs with non-normal $q-$dimensional random effects ($q \geq 1$), and requires that the density function of the non-normal random effects have a closed form.

To demonstrate the LR method to estimate the heterogeneity model, consider a random intercept logistic model with the random intercept assumed to follow a g component mixture of univariate normal distributions. Consider the same random intercept logistic model presented in Equation 3.5 where b_{0i} has a finite mixture density,

$$f(b_{0i}) = \sum_{\gamma=1}^{g} \frac{\pi_\gamma}{\sqrt{2\pi}\sigma_\gamma} \exp\left(-\frac{(b_{0i}-\mu_\gamma)^2}{2\sigma_\gamma^2}\right) \tag{3.13}$$

where μ_γ and σ_γ^2 are the mean and variance of the γ^{th} mixture component, with mixing proportions π_γ ($\sum_{\gamma=1}^{g} \pi_\gamma = 1$, for $\gamma = 1, ..., g$). The observed data likelihood contribution for individual i is given by:

$$L_i = \int \left[\prod_{j=1}^{n_i} f(y_{ij}|b_{0i}, \boldsymbol{\beta}, \phi)\right] f(b_{0i}) db_{0i}. \tag{3.14}$$

where $f(y_{ij}|b_{0i}, \boldsymbol{\beta}, \phi) = p_{ij}^{y_{ij}} \times (1-p_{ij})^{(1-y_{ij})}$ is the logit function, with

[3]However, determination of the number of components in a finite mixture is not a standard problem. The likelihood ratio test statistic does not necessarily follow a chi-squared distribution, as the null hypothesis is on the boundary of the parameter space. In practice it is often sufficient to select the optimal number of components by exploring how different values of g affect the inference of model parameters within the context of a sensitivity analysis (Molenberghs and Verbeke, 2005).

$p_{ij} = \left(1 + \exp(-(\boldsymbol{x'_{ij}}\boldsymbol{\beta} + b_{0i}))\right)^{-1}$. The corresponding observed data likelihood is thus,

$$L_i = \int \left[\prod_{j=1}^{n_i} \exp\left\{y_{ij}\log(p_{ij}) + (1-y_{ij})\log(1-p_{ij})\right\}\right] \times f(b_{0i})\mathrm{d}b_{0i} \tag{3.15}$$

Using the LR method of Liu and Yu (2008), the likelihood can be reformulated such that the integration is over a standard normal distribution rather than over a mixture of normal distributions. The reformulation can be done by multiplying and dividing the integrand in Equation 3.15 by a standard normal density, $\phi(a_i)$, (i.e. $a_i \sim N(0,1)$),

$$L_i = \int \left[\prod_{j=1}^{n_i} f(y_{ij}|a_i, \boldsymbol{\beta}, \phi)\right] \frac{f(a_i|\theta)}{\phi(a_i)}\phi(a_i)\mathrm{d}a_i \tag{3.16}$$

where it is noted that the non-normal random intercept in Equation 3.15, b_{0i}, has been replaced by a_i in Equation 3.16 to distinguish between the true non-normal random effects from the standard normal $\phi(a_i)$.

Algebraically the derivation of the likelihood can be shown to be expressed as a function of the conditional likelihood of the observed data (ℓ_i^A), the density of the non-normal distributed random effects (ℓ_i^B) and the standard normal density ℓ_i^C,

$$\begin{aligned} L_i &= \int \left[\prod_{j=1}^{n_i} f(y_{ij}|a_i, \boldsymbol{\beta}, \phi)\right] \frac{f(a_i|\theta)}{\phi(a_i)}\phi(a_i)\mathrm{d}a_i \\ &= \int_{-\infty}^{\infty} \exp\left(\log\left(\prod_{j=1}^{n_i} f(y_{ij}|a_i, \boldsymbol{\beta}, \phi)\right) + \log\left(f(a_i|\theta)\right) - log\left(\phi(a_i)\right)\right)\phi(a_i)\mathrm{d}a_i \\ &= \int_{-\infty}^{\infty} \exp(\ell_i^A + \ell_i^B - \ell_i^C)\phi(a_i)\mathrm{d}a_i. \end{aligned} \tag{3.17}$$

For the heterogeneity model, the likelihood contribution for the i^{th} individual can be expressed using Equation 3.17, where ℓ_i^A is the conditional log-likelihood of the observed data, ℓ_i^B is the log finite mixture density and ℓ_i^C is the log standard normal density. Given by,

$$\begin{aligned} \ell_i^A &= \sum_{j=1}^{n_i} \left\{y_{ij}\log(p_{ij}) + (1-y_{ij})\log(1-p_{ij})\right\} \\ \ell_i^B &= \log\left(\frac{\pi_1}{\sigma_1}\exp\left(-\frac{(a_i-\mu_1)^2}{2\sigma_1^2}\right) + ... + \frac{\pi_g}{\sigma_g}\exp\left(-\frac{(a_i-\mu_g)^2}{2\sigma_g^2}\right)\right) + \text{constant} \\ \ell_i^C &= \log\left(\frac{1}{\sqrt{(2\pi)}}\exp\left(-\frac{(a_i-0)^2}{2}\right)\right) = -\frac{1}{2}a_i^2 + \text{constant} \end{aligned} \tag{3.18}$$

where $p_{ij} = \left(1 + \exp(-(\boldsymbol{x'_{ij}}\boldsymbol{\beta} + a_i))\right)^{-1}$.

The reformulated likelihood can be estimated using adaptive Gaussian quadrature and implemented in SAS using the NLMIXED procedure. The intercept is omitted to avoid non-identifiability (Liu and Yu, 2008), and is estimated using the assumption that $E(b_{0i}) = 0$ as imposed by the restriction $\sum_{\gamma=1}^{g} \hat{\pi}_\gamma \hat{\mu}_\gamma = \hat{\beta}_0$. Furthermore, to impose the restriction that the mixing proportions sum to one ($\sum_{\gamma=1}^{g} \pi_\gamma = 1$), the g^{th} mixing proportion is estimated as $\hat{\pi}_g = 1 - \hat{\pi}_1 - \hat{\pi}_2 - ... - \hat{\pi}_{g-1}$. The LR method implemented for the random intercept logistic model can easily be extended to estimate a random intercepts and random slopes logistic mixed model assuming random effects are sampled from a finite mixture of bivariate normals. For further information and example SAS syntax, the reader is referred to Liu and Yu (2008).

3.2.2 Semi-non-parametric flexible random effects model

Another approach relaxing the assumption of Gaussian random effects in a GLMM is to specify the random effects density to follow a smooth class of densities as represented by the semi-non-parametric (SNP) approach of Gallant and Nychka (1987). The SNP density is a flexible class of continuous densities that includes skewed, multimodal and thick- or thin-tailed densities. Thus, assuming the random effects follow a SNP density can induce an immense amount of flexibility. Furthermore, by assuming a smooth density without jumps or oscillations, the SNP representation also provides an estimate of the random effects density.

Recently, Vock et al. (2014) developed a SAS macro, SNP_NLMM, to easily implement generalised linear mixed models and non-linear mixed models (NLMMs) with random effects assumed to follow the SNP density. The SAS macro of Vock et al. (2014) overcomes computational challenges of previously implementing SNP densities in standard software, by proposing a fast computational approach to approximate the integral required to obtain maximum likelihood estimates. The estimation technique proposed by Vock et al. (2014) fits the random effects of a GLMM or NLMM, however for the remainder of this description we will restrict our focus to GLMMs.

Following the same notation as Section 3.1, the likelihood for a GLMM with q random effects is given by,

$$L(\boldsymbol{\theta}, \boldsymbol{y}) = \prod_{i=1}^{N} \int_{-\infty}^{\infty} ... \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{b_i}, \boldsymbol{\beta}) f(\boldsymbol{b_i}, \boldsymbol{\xi}) db_{1i}...db_{qi} \tag{3.19}$$

where it is assumed the random effects $\boldsymbol{b_i} \sim f(\boldsymbol{b_i}, \boldsymbol{\xi})$ may depend on a vector of parameters $\boldsymbol{\xi}$. For brevity, the q-dimensional integral will be represented as a single integral for the remainder of this description.

Assume the random effects are centered such that $\boldsymbol{b_i} = \boldsymbol{\mu} + \boldsymbol{R}\boldsymbol{v_i}$ where $\boldsymbol{\mu}$ is a q-dimensional vector, $\boldsymbol{R}$ is a $q \times q$ lower triangular matrix, and $\boldsymbol{v_i}$ is a q-dimensional random effects matrix.

It is assumed that $\boldsymbol{v_i}$ follows a q-dimensional standard normal density, such that $\mathrm{E}(\boldsymbol{b_i}) = \boldsymbol{\mu}$ and $\mathrm{Var}(\boldsymbol{b_i}) = \boldsymbol{RR'}$. The likelihood can be rewritten in terms of $\boldsymbol{v_i}$ such that

$$L(\boldsymbol{\theta}, \boldsymbol{y}) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{v_i}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{r}) f_{v_i}(\boldsymbol{v_i}, \boldsymbol{\xi}) dv_i \tag{3.20}$$

where $\boldsymbol{r}$ is the half-vectorisation of matrix $\boldsymbol{R}$ (i.e. $vech(\boldsymbol{R})$).

As detailed in Vock et al. (2014), the SNP representation assumes that instead of $\boldsymbol{v_i}$ distributed as a q-dimensional standard normal, the $\boldsymbol{v_i}$ belong to the smooth class of densities as proposed by Gallant and Nychka (1987). The smooth class of densities, $f_{v_i} \in G$, can be expressed as an infinite Hermite series, $f_{v_i}(\boldsymbol{v}) = P^2_\infty(\boldsymbol{v})\phi(\boldsymbol{v})$ with a lower bound on the tails, where $P^2_\infty(\boldsymbol{v})$ is an infinite dimensional polynomial and $\phi(\boldsymbol{v})$ is a q-dimensional standard normal density. For modelling purposes the lower bound is ignored and the polynomial term is truncated, and as such, is referred to as a SNP density. Therefore, the random effects $\boldsymbol{v_i}$ are assumed to follow a SNP density with degree of truncation K, as given by:

$$\begin{aligned} f^K_{v_i}(\boldsymbol{v}; \boldsymbol{\xi}) &= P^2_K(\boldsymbol{v})\phi_q(\boldsymbol{v}) \\ &= \left\{ \sum_{(h_1+\ldots+h_q)\leq K} a_{h_1,\ldots,h_q}(v_1^{h_1} \ldots v_q^{h_q}) \right\}^2 \phi_q(v) \end{aligned} \tag{3.21}$$

where $\phi_q(\boldsymbol{v})$ is the q-dimensional standard normal density, a are the coefficients with $h_l \geq 0$ for $l = 1, \ldots, q$ and K is the order of the polynomial $P_K(\boldsymbol{v})$. For example, a random intercept density (i.e. $q = 1$) with $K = 2$,

$$\begin{aligned} P_K(\boldsymbol{v}) &= a_0 \boldsymbol{v_1}^0 + a_1 \boldsymbol{v_1}^1 + a_2 \boldsymbol{v_1}^2 \\ &= a_0 + a_1 \boldsymbol{v_1} + a_2 \boldsymbol{v_1}^2. \end{aligned} \tag{3.22}$$

Similarly, for a random intercept and random slope density (i.e. $q = 2$) with $K = 2$,

$$\begin{aligned} P_K(\boldsymbol{v}) &= a_{00}\boldsymbol{v_1}^0\boldsymbol{v_2}^0 + a_{01}\boldsymbol{v_1}^0\boldsymbol{v_2}^1 + a_{10}\boldsymbol{v_1}^1\boldsymbol{v_2}^0 + a_{11}\boldsymbol{v_1}^1\boldsymbol{v_2}^1 + a_{02}\boldsymbol{v_1}^0\boldsymbol{v_2}^2 + a_{20}\boldsymbol{v_1}^2\boldsymbol{v_2}^0 \\ &= a_{00} + a_{01}\boldsymbol{v_2} + a_{10}\boldsymbol{v_1} + a_{11}\boldsymbol{v_1}\boldsymbol{v_2} + a_{02}\boldsymbol{v_2}^2 + a_{20}\boldsymbol{v_1}^2 \end{aligned} \tag{3.23}$$

where $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are the vectors of centered random intercepts and random slopes respectively.

The above description of fitting random effects assuming a SNP density assumes that the truncation factor, K, is fixed. When $K = 0$ the SNP density simplifies to a q-dimensional standard normal distribution, and for $K > 0$ the value controls the departure from the standard normal, and therefore influences the flexibility for approximating the true underlying random effects density (Vock et al., 2014). Thus, K should be used as a tuning parameter, and Vock et al. (2014) suggest fitting models for several values of K and choosing the optimal model based on information criteria and/or visual inspection of the resulting densities. As the class of possible densities increases monotonically as K is increased, Vock et al. (2014) suggest values of $K \leq 2$ will be sufficient to capture most complicated densities.

Estimation of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\xi}')'$ is obtained using the maximum likelihood estimator (Vock et al., 2014). For a given K, the optimisation is the same as in standard finite dimensional maximum likelihood estimation, after approximating the integral in the likelihood (Equation 3.20). The mean and variance of the random effects are estimated as $\text{E}(\boldsymbol{b_i}) = \boldsymbol{\mu} + \boldsymbol{R} \times \text{E}(\boldsymbol{v_i})$ and $\text{Var}(\boldsymbol{b_i}) = \boldsymbol{R}\text{Var}(\boldsymbol{v_i})\boldsymbol{R}'$. For computational details of the numerical integration method proposed and utilised by Vock et al. (2014), the reader is referred to Section 4 of Vock et al. (2014). Briefly, the SNP_NLMM method uses Gaussian quadrature to approximate the integral over the random effects, with the quadrature points centered and scaled at the empirical Bayes estimates and estimated variance from assuming $q-$dimensional multivariate normally distributed random effects. The SAS macro that implements this methodology is provided as supplementary material in Vock et al. (2014), and utilises the NLMIXED procedure in SAS to numerically integrate and optimise the likelihood, using the likelihood reformulation method proposed by Liu and Yu (2008) to center the quadrature points and allow non-Gaussian random effects.

As the SNP_NLMM approach can be sensitive to the choice of starting values, Vock et al. (2014) suggest using starting values for $\boldsymbol{\beta}$, $E(\boldsymbol{b_i})$ and $\text{Var}(\boldsymbol{b_i})$ as the estimated values of the corresponding GLMM assuming random effects are a q-dimensional multivariate normal distribution. Specifically, $\boldsymbol{\mu}$ and $\boldsymbol{r}$ are set to the values that correspond to $E(\boldsymbol{b_i})$ and $\text{Var}(\boldsymbol{b_i})$ when assuming a non-mean-zero Gaussian distribution. For example, for the univariate random effects GLMM, the initial starting value for $E(b_{0i})$ would be set to $\hat{\beta}_0$ and $\text{Var}(b_{0i})$ would be set to $\hat{\sigma}^2_{b_{0i}}$ from the model assuming normal random intercepts (i.e. $b_{0i} \sim N(\hat{\beta}_0, \hat{\sigma}^2_{b_{0i}})$). The likelihood over a grid of values of $\boldsymbol{\xi}$ is evaluated, and the SNP_NLMM macro uses a small number of the parameter sets as starting values to maximise the likelihood and ensure convergence at the global maxima.

The SNP_NLMM macro is currently available to fit GLMMs with one or two dimensional random effects. As described previously, the value of K is used as a tuning parameter, and the optimal value of K is chosen based on the Akaike Information Criterion or the Bayesian Information Criterion (BIC is default). The maximum number of K, denoted K_{max} must be less than or equal to four for $q = 1$ and less than or equal to three for $q = 2$. The macro also requires the user to specify the number of grid points in each dimension of $\boldsymbol{\xi}$ (with at least 9 recommended in each of the q dimensions (Vock et al., 2014)). The SNP_NLMM approach may be sensitive to the choice of quadrature points, and re-running the optimisation with increasing quadrature points may be necessary to ensure the maximum likelihood estimate is not sensitive to the number of quadrature points (Vock et al., 2014). The default choice of the macro is for the procedure to adaptively select the number of quadrature points, however it is recommended to assess the sensitivity by considering alternative numbers of quadrature points. For further details in regards to implementation and estimation the reader is referred to Vock et al. (2014).

3.2.3 Non-parametric maximum likelihood estimation of the random effects distribution

Instead of making any assumptions about the distribution of the random effects, non-parametric maximum likelihood estimation allows immense flexibility of the random effects distribution by leaving the distribution completely unspecified. In the context of random intercept mixed models, it has been shown that the non-parametric maximum likelihood estimator (NPMLE) of the unspecified distribution is a discrete distribution with a finite number of M support points located at ν_m $(m = 1, ..., M)$ with probability weights π_m $(\sum_{m=1}^{M} \pi_m = 1)$ (Simar, 1976; Laird, 1978; Lindsay, 1983). For a GLMM with multiple random effects (i.e. $q > 1$), the location of the support points $\boldsymbol{\nu_m} = (\nu_1, ..., \nu_q)'$ is in a q-dimensional space (Aitkin, 1999). Despite the discreteness of the resulting random effects estimate, non-parametric maximum likelihood estimation does not assume the random effects are discrete. The resulting estimate is an approximation of the true underlying distribution, regardless of it being continuous (normal or non-normal), discrete, or both continuous with discrete components (Skrondal and Rabe-Hesketh, 2004). The likelihood corresponding to the NPML estimate of the GLMM with the random effects estimated on M support points is given by,

$$L(\boldsymbol{\theta}^M, \boldsymbol{\pi}_M, \boldsymbol{\nu}_M) = \prod_{i=1}^{N} \sum_{m=1}^{M} \pi_m \prod_{j=1}^{n_i} f(y_{ij} | \boldsymbol{b_i} = \boldsymbol{\nu_m}, \boldsymbol{\beta}, \phi) \tag{3.24}$$

In order to achieve the NPMLE, the optimal number of support points M needs to be determined. Many approaches utilised to achieve the NPMLE are based on the concept of the directional derivative that has been used and discussed in many contexts (e.g. Wynn 1970; Fedorov 1972; Wu 1978; Lindsay 1983; Follmann and Lambert 1989; Lesperance and Kalbfleisch 1992, among others). One approach, as utilised by Rabe-Hesketh et al. (2003), is to introduce support points one at a time until the likelihood is maximised as determined by the directional derivative, also known as the Gateaux derivative (Heckman and Singer, 1984). Another approach, as utilised by Tsonaka et al. (2009) in the context of shared parameter models, is to start with a large grid of support points and remove or merge support points using the directional derivative-based Vertex Exchange Method (Böhning, 1985) until the log-likelihood is maximised with respect to the random effects distribution. These two methods are described in further detail in Sections 3.2.3.1 and 3.2.3.2, respectively.

3.2.3.1 Non-parametric maximum likelihood estimation of the random effects distribution utilising the Gateaux derivative

The approach proposed by Rabe-Hesketh et al. (2003) is similar to the algorithm proposed by Simar (1976) and adapted by Heckman and Singer (1984) (among others). The approach starts with a single support point, and estimates the model parameters jointly with both the location and probability weight of the support point using maximum likelihood estimation. Additional support points are introduced one at a time using an implementation of the directional derivative-based algorithm, the Vertex Direction Method (VDM) (Wynn, 1970; Fedorov, 1972; Wu, 1978; Lindsay, 1983). Keeping the model parameters fixed at the current estimate, the Gateaux derivative calculates the log-likelihood when an additional support point of small probability weight is moved across a fine grid of values for the random effects. If the log-likelihood increases at any of the considered locations, a new support point is introduced. The estimation of the model with an additional support point is based on maximum likelihood with the starting values set at the previous model estimates (parameter coefficients and support point estimates) with the starting value of the additional support point set at the location corresponding to the greatest increase in the log-likelihood.

More formally, consider the maximised likelihood for a random intercept model with M support points, $L(\hat{\boldsymbol{\theta}}^M, \hat{\boldsymbol{\pi}}_M, \hat{\boldsymbol{\nu}}_M)$ (Equation 3.24). To determine whether $L(\hat{\boldsymbol{\theta}}^M, \hat{\boldsymbol{\pi}}_M, \hat{\boldsymbol{\nu}}_M)$ is the NPMLE, assess whether the inclusion of an additional support point will result in a larger maximum likelihood. For a fine grid of support points consider changing the discrete distribution along the path $((1-\lambda)\hat{\boldsymbol{\pi}}_M, \lambda)'$ with support points located at $(\hat{\boldsymbol{\nu}}_M, \nu_{M+1})'$. Therefore, if $\lambda = 0$, then the current solution consisting of M support points is the NPMLE, however if $\lambda = 1$ then an additional support point at location ν_{M+1} is included. This can be formally assessed by considering the directional derivative, given by

$$\Delta(\nu_{M+1}) = \lim_{\lambda \to 0} \frac{\ln L(\hat{\boldsymbol{\theta}}^M, ((1-\lambda)\hat{\boldsymbol{\pi}}_M, \lambda)', (\hat{\boldsymbol{\nu}}_M, \nu_{M+1})') - \ln L(\hat{\boldsymbol{\theta}}^M, \hat{\boldsymbol{\pi}}_M, \hat{\boldsymbol{\nu}}_M)}{\lambda}. \tag{3.25}$$

As defined by the general mixture maximum likelihood theorem (Lindsay, 1983), the model with M support points is considered the NPMLE if and only if $\Delta(\nu_{M+1}) \leq 0$ for all ν_{M+1}. If a location can be found (i.e. if $\Delta(\nu_{M+1}) > 0$), the directional derivative implies that including an additional support point will improve the maximum likelihood. This procedure of assessing whether an additional support point can be included, and estimating the model with $M + 1$ support points is repeated until no additional location can be found to increase the likelihood (i.e. when $\Delta(\nu_{M+1}) \leq 0$).

As detailed by Rabe-Hesketh et al. (2003), the model parameters $\boldsymbol{\theta}$ are estimated using the Newton-Raphson algorithm. To ensure correct parametrisation of the model, the assumption of zero mean random effects (i.e. $E(b_i) = 0$) is imposed by estimating the location for $M - 1$ support point locations. Furthermore, to ensure that the probability weights sum to one, the restriction is imposed by estimating $M-1$ probability weights (i.e. $\pi_M = 1 - \sum_{m=1}^{M-1} \pi_m$). The variance-covariance matrix is not estimated directly but is based on the estimated locations and probability weights of the support points. Therefore, the variances and covariances are based on the q-dimensional discrete probability distribution, estimated as $\sum_{m=1}^{M} \boldsymbol{\nu_m} \boldsymbol{\nu_m}' \pi_m$. The approximate standard errors are estimated by inverting the observed information matrix. As the information matrix includes terms for the mass-point parameters, the standard errors for the coefficients take into account the uncertainty of the locations and masses (Rabe-Hesketh et al., 2003). However the standard errors do not take into account the uncertainty of the number of support points, and thus, are conditional on the number of support points.

Non-parametric maximum likelihood estimation using the Gateaux derivative to obtain the NPMLE is implemented in STATA using the GLLAMM package. The default stopping rule of the Gateaux derivative implemented in GLLAMM is if $\Delta(\nu_{M+1}) \leq 10^{-5}$ for all locations of ν_{M+1} along a user-specified fine grid of locations spanning across a wide range of values. For further details regarding implementation in GLLAMM and interpretation of the resulting output, the reader is referred to Rabe-Hesketh et al. (2003), Skrondal and Rabe-Hesketh (2004) and the GLLAMM website (http://www.gllamm.org).

3.2.3.2 Non-parametric maximum likelihood estimation of the random effects distribution utilising the Vertex Exchange Method

Böhning (1999) has shown that a two phase procedure can be utilised to obtain the NPMLE. The first phase is to use the VEM algorithm to estimate the probability masses $\boldsymbol{\pi}$ in a fixed grid of support points, $\boldsymbol{\mu}$. In the second phase, the locations of $\boldsymbol{\mu}$ are refined by using the EM algorithm with the estimated distribution from the first phase as the initial starting values. However, the EM step is computationally slow and has been shown to have minimal additional improvement on the model fit if the original gird is sufficiently dense (in terms of maximising the log-likelihood) (Böhning, 1999). Therefore, Tsonaka et al. (2009) proposed that the random effects distribution can be estimated using only the VEM algorithm. The resulting estimate derived by VEM will provide an approximate NPMLE of the random effects distribution, G (Tsonaka et al., 2009). The VEM algorithm is described in more detail below.

The VEM was proposed by (Böhning, 1985) as an alternative to the VDM (Wynn, 1970; Fedorov, 1972; Wu, 1978; Lindsay, 1983). The VEM algorithm starts with a very dense grid of support points and within an iterative procedure either merges or omits support points as determined by the directional derivative. To illustrate the VEM algorithm, consider a fixed, pre-specified grid of equally spaced and equally weighted C support points in the one-dimensional case, $\mu_1, ..., \mu_C$ (such as a random intercept distribution in a random intercept logistic model). Therefore, the VEM algorithm starts by assuming each of the C support points have a probability weight, $\pi_c = \frac{1}{C}$ for $c = 1, ..., C$. In each iteration, the VEM algorithm maximises the log-likelihood $l(G|\theta)$ by moving weight from a ‘bad’ support point to a ‘good’ support point.

The 'good' and 'bad' support points correspond to the locations that maximise and minimise the directional derivative, $D(G^0, G_{\mu_c})$, over the grid of C support points, denoted respectively μ^+ and μ^- with corresponding weights $\pi_{\mu+}$ and $\pi_{\mu-}$ (Böhning, 1985, 1999; Tsonaka et al., 2009) .

More formally, at the t^{th} iteration of the estimation procedure, for each grid point μ_c ($c = 1, ..., C$) the directional derivative is given by

$$\begin{aligned} D(\hat{G}^{(t)}, G_{\mu_c}) &= \lim_{s \to 0} \frac{\ell\left((1-s)\hat{G}^{(t)} + sG_{\mu_c}\right) - \ell(G^{(t)})}{s} \\ &= \sum_{i=1}^{N} \frac{f(y_i|\mu_c; \hat{\theta}^{(t)})}{\sum_{c=1}^{C} \hat{\pi}_c^{(t)} f(y_i|\mu_c; \hat{\theta}^{(t)})} - N \end{aligned} \tag{3.26}$$

where $\hat{G}^{(t)}$ is the current estimate of G at iteration t as estimated by the weights $\hat{\pi}_c^{(t)}$. Of the C evaluations of the directional derivative, define $\mu^- = \arg\min_{\mu_c} D(\hat{G}^{(t)}, G_{\mu_c})$ and $\mu^+ = \arg\max_{\mu_c} D(\hat{G}^{(t)}, G_{\mu_c})$, corresponding to the support points that minimise and maximise the directional derivative. Once these two support points have been identified, weight of the support points are exchanged by moving weight from μ^- in the direction of μ^+. The weights of μ^- and μ^+, are updated by $\hat{\pi}_{\mu-}^{(t+1)} = (1-s^*)\hat{\pi}_{\mu-}^{(t)}$ and $\hat{\pi}_{\mu+}^{(t+1)} = s^*\hat{\pi}_{\mu-}^{(t)} + \hat{\pi}_{\mu+}^{(t)}$, with step length $s^* \in [0,1]$. In order to determine the optimal step length, as defined by $s^* = \arg\max_s [l(\hat{G}^{(t+1)}(s)|\hat{\theta}^{(it)}) - l(\hat{G}^{(t)}|\hat{\theta}^{(t)})]$ for $s \in [0,1]$, the line search method is utilised (Tsonaka et al., 2009). If the optimal step length is 1 ($s^* = 1$) such that $\hat{\pi}_{\mu-}^{(t+1)} = 0$, then the updated grid will be reduced by one support point (i.e. μ_- will be removed from the fixed grid $\mu_1, ..., \mu_C$). The VEM algorithm continues the iterative process until convergence.

Convergence is reached when the following two conditions are met: (i) the change in the log-likelihood from the current estimate of G and the previous estimate at $t-1$, $\ell(\hat{G}^{(t)}) - \ell(\hat{G}^{(t-1)})$ is less than the stopping criteria defined by $\epsilon' \left| l(\hat{G}^{(t)}) \right| + \epsilon'$ where ϵ' is small (i.e. $\epsilon' = 10^{-7}$); and (ii) the maximum directional derivative over the grid of C support points is small, such as $\max_{\mu_c} D(\hat{G}^{(t)}, G_{\mu_c}) < 10^{-3}$ which guarantees that $l(\hat{G}^{(t)}) - l(\hat{G}^{(t-1)}) < 10^{-3}$ (Tsonaka et al., 2009). This reflects that the estimate of G that maximises the log-likelihood can equivalently be characterised by the three conditions: (1) $\hat{G}$ maximised $L(G)$, (2) $\hat{G}$ minimises $\sup_\theta D(\boldsymbol{\theta}; G)$, and (3) $\sup_\theta D(\boldsymbol{\theta}, \hat{G}) = 0$ (Theorem 4.1 in Lindsay, 1983).

As the aim of non-parametric maximum likelihood estimation in GLMMs is to simultaneously find maximum likelihood estimates of the parameter coefficients ($\boldsymbol{\theta}$) and estimate the random effects distribution, Tsonaka et al. (2009) proposed a two step optimisation procedure that utilises the Vertex Exchange Method to estimate the random effects distribution. In the first step, the random effects distribution G is estimated using the VEM, for $\boldsymbol{\theta}$ fixed at its current estimate ($\hat{\boldsymbol{\theta}}$). In the second step, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm (Nocedal and Wright, 2006) is used to update $\hat{\boldsymbol{\theta}}$ by maximising the likelihood at the estimated random effects distribution from the first step ($\hat{G}$). These two steps are repeated in an iterative process until convergence, as defined above, is reached. Due to an alternative reparameterisation of the logistic mixed model (detailed in Section 7.2.1) the random effects are not restricted to have zero mean, and are estimated as $\hat{\boldsymbol{b}}_{\boldsymbol{i}} = \sum_{c=1}^{C} \pi_c \boldsymbol{\mu_c}$. Furthermore, the variance-covariance matrix is based on the estimated locations and probability weights of the support points. Therefore, the variances and covariances are based on the q-dimensional discrete probability distribution, estimated as $\sum_{c=1}^{C}(\boldsymbol{\mu_c} - \hat{\boldsymbol{b}}_{\boldsymbol{i}})(\boldsymbol{\mu_c} - \hat{\boldsymbol{b}}_{\boldsymbol{i}})'\pi_c$. The approximate standard errors for the parameter coefficients are estimated by inverting the Hessian of the log-likelihood evaluated at the estimates $\hat{\boldsymbol{\theta}}$ and $\hat{G}$.

Non-parametric maximum likelihood estimation using VEM to obtain an approximation to the NPMLE can be implemented in R using code adapted from Tsonaka et al. (2009). The code was originally developed for estimating the random effects distribution in shared parameter models, and Dr. Tsonaka provided the R code to estimate the random effects in a random intercept logistic model. The R code has been extended by the candidate to estimate logistic mixed models with bivariate random effects. The relevant R code is available upon request.

3.2.4 Random intercept logistic model with endpoints

As detailed in Section 2.5, if a latent mover-stayer scenario is suspected, the goodness-of-fit of GLMMs can be improved by incorporating spikes into the parametric random effects distribution to represent the stayers (Singer and Spilerman, 1976). The so-called "spiked" distribution is a combination of the mover-stayer model (Goodman, 1961) and mixture models. It is often considered a more parsimonious approach to account for latent stayers than non-parametric maximum likelihood estimation (particularly in regards to the number of estimable parameters) (Davies and Crouchley, 1986; Berridge and Crouchley, 2011a).

To allow for latent stayers in a two-level random intercept logistic model, the assumed normal distribution for the random intercepts is supplemented with endpoints at positive and/or negative infinity. By considering endpoints at negative and positive infinity, two types of stayers can be accounted for. The stayers susceptible to remain in state $y_{ij} = 0$ are represented by a spike at negative infinity (denoted $S_{-\infty}$) and the stayers susceptible to remain in state $y_{ij} = 1$ are represented by a spike at positive infinity (denoted $S_{+\infty}$). The probabilities of these two subgroups of stayers are denoted $\Pr[S_{-\infty}]$ and $\Pr[S_{+\infty}]$, respectively. Therefore, the random intercept distribution consists of a homogeneous group of 'movers' represented by a normal distribution, and one or two spikes to represent the latent stayers. Let $y_{i1}, ..., y_{in_i}$ denote the sequence of binary response measurements for the i^{th} individual. The likelihood of the random intercept logistic model with two endpoints at positive and negative infinity is given by,

$$
\begin{aligned}
L(\boldsymbol{\beta}, \sigma_{b_{0i}}^2 | \boldsymbol{y}) = \prod_{i=1}^{N} \Bigg\{ & \Pr[S_{-\infty}] \left[\prod_{j=1}^{n_i} (1 - y_{ij}) \right] + \Pr[S_{+\infty}] \left[\prod_{j=1}^{n_i} (y_{ij}) \right] \\
& + \frac{L_i}{(1 - \Pr[S_{-\infty}] - \Pr[S_{+\infty}])} \Bigg\}
\end{aligned}
\tag{3.27}
$$

where L_i is the likelihood contribution for individual i, as given by

$$
\begin{aligned}
L_i(\boldsymbol{\beta}, \sigma_{b_{0i}}^2 | \boldsymbol{y}) &= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} f(y_{ij} | u_{0i}, \boldsymbol{\beta}, \phi) f(u_{0i}) du_{0i} \\
&= \int_{-\infty}^{\infty} \prod_{j=1}^{n_i} \frac{\left[\exp(\boldsymbol{x'_{ij}} \boldsymbol{\beta} + u_{0i}) \right]^{y_{ij}}}{1 + \exp(\boldsymbol{x'_{ij}} \boldsymbol{\beta} + u_{0i})} f(u_{0i}) du_{0i}
\end{aligned}
\tag{3.28}
$$

where u_{0i} is the random intercept and is assumed to be normally distributed ($u_{0i} \sim N(0, \sigma_{u_{0i}}^2)$). The probabilities $\Pr[S_{-\infty}]$ and $\Pr[S_{+\infty}]$, are parameterised as

$$
\Pr[S_{-\infty}] = \frac{\zeta_0}{1 + \zeta_0 + \zeta_1} \tag{3.29}
$$

and

$$
\Pr[S_{+\infty}] = \frac{\zeta_1}{1 + \zeta_0 + \zeta_1} \tag{3.30}
$$

where $\zeta_0 > 0$ and $\zeta_1 > 0$ are end-point parameters to be estimated.

Random intercept logistic models with endpoints can be estimated when implemented using SABRE (Barry et al., 1989), which is currently available as a stand-alone package, as a library in R, or as a plug-in for STATA. The likelihood in Equation 3.27 is maximised using a Newton-Raphson algorithm and using Gaussian quadrature to numerically evaluate the integral in Equation 3.28 (Barry et al., 1989). Implementation of spiked distributions in SABRE is restricted to random intercept logistic models and random intercept Poisson models. For the random intercept Poisson model, the end-point is restricted to the negative infinity value to account for latent stayers in the null state ($y_{ij} = 0$) (Berridge and Crouchley, 2011a).

The reader is referred to the SABRE website (http://sabre.lancs.ac.uk/) for further details about the implementation of SABRE in each of the three software platforms.

3.3 Misspecification diagnostic tools

As outlined in Section 3.1, inference for GLMMs is typically based on the marginal model of $\boldsymbol{y_i}$, obtained by integrating over the distribution of the random effects, G (Fitzmaurice et al., 2009). Therefore, as the assumed distribution of the random effects is crucial in the calculation of the marginal model, it is important to assess the adequacy of the fit of the resulting log-likelihood $\ell(\hat{G}) = \sum_{i=1}^{N} \ln[f_i(\boldsymbol{y_i}|G)]$ to the data (Verbeke and Molenberghs, 2013).

Two diagnostic tools are implemented in Chapter 4 to assess for distributional misspecification of the random effects and these are detailed below. Section 3.3.1 describes the gradient function exploratory diagnostic tool of Verbeke and Molenberghs (2013), a graphical tool to assess whether the assumed random effects, G, adequately fits the data, or if there exists any other random effects distribution that improves the fit. However, as the exploratory tool is an informal diagnostic tool, Drikvandi et al. (2016) have recently proposed a powerful diagnostic test to supplement the graphical diagnostic function tool. The diagnostic test of Drikvandi et al. (2016) is also based on the gradient function, and formally tests for misspecification of the random effects distribution, as described in Section 3.3.2.

3.3.1 Gradient function exploratory diagnostic tool

The gradient function exploratory tool proposed by Verbeke and Molenberghs (2013) uses the directional derivative to check whether $\ell(\hat{G})$ adequately fits the data or whether there exists any other random effects distribution, denoted H, that yields a larger log-likelihood than $\ell(\hat{G})$ (i.e. $\ell(H) > \ell(\hat{G})$). Consider two random effects distributions, G and H, the directional derivative of the log-likelihood evaluated at G into the direction of H is defined as:

$$\begin{aligned} \Phi(G,H) &= \lim_{\alpha \to 0} \frac{\ell[(1-\alpha)G + \alpha H] - \ell(G)}{\alpha} \\ &= \frac{\partial \ell[(1-\alpha)G + \alpha H]}{\partial \alpha}|_{\alpha=0} \end{aligned} \tag{3.31}$$

where α is an infinitesimal weight assigned to the distribution H. Equation 3.31 represents the infinitesimal change in the log-likelihood when the random effects distribution G is replaced by the mixture $(1-\alpha)G + \alpha H$. If $\Phi(G,H) \leq 0$ for all H then there exists no better random effects distribution than G. Furthermore, it can be shown that,

$$\begin{aligned} \frac{1}{N}\Phi(G,H) &= \frac{1}{N}\sum_{i=1}^{N} \frac{f_i(\boldsymbol{y_i}|H) - f_i(\boldsymbol{y_i}|G)}{f_i(\boldsymbol{y_i}|G)} \\ &= \frac{1}{N}\sum_{i=1}^{N} \frac{f_i(\boldsymbol{y_i}|H)}{f_i(\boldsymbol{y_i}|G)} - 1 \\ &= \int \Delta(G,\boldsymbol{b})\mathrm{d}H(b) - 1 \end{aligned} \tag{3.32}$$

where the gradient function,$\Delta(G,\boldsymbol{b})$, is the average of likelihood ratios given by:

$$\Delta(G,\boldsymbol{b}) = \frac{1}{N}\sum_{i=1}^{N} \frac{f_i(\boldsymbol{y_i}|\boldsymbol{b})}{f_i(\boldsymbol{y_i}|G)}. \tag{3.33}$$

For each value $b \in R^q$, the gradient function is interpreted as an average of likelihood ratios. The likelihood ratios in the gradient function measure how much more likely $\boldsymbol{y_i}$ is observed for individual i to have the random effects $\boldsymbol{b_i}$ correspond to $\boldsymbol{b}$ than to be sampled from G.

There are three properties of the gradient function and the directional derivative that imply that, if there is no random effects distribution H that provides a better fit than $\hat{G}$, then $\Delta(\hat{G}, \boldsymbol{b})$ will not exceed 1 and will equal 1 for all support points of $\hat{G}$ within the region I. The supportive region, I, is located in the interval $[b^*_{min}, b^*_{max}]$ where b^*_{min} and b^*_{max} are the minimum and maximum b^*_i values for individuals $i = 1, ..., N$. The b^*_i corresponds to the value of $\boldsymbol{b}$ that maximises the log-likelihood value for individual i (i.e. the unique mode). Therefore, the supportive region will cover the whole real line if there exists any individuals with constant response profiles (i.e. for binary or categorical response data, those individuals who do not change outcomes over the observation period). For example, in a random effects logistic regression some individuals may have all binary response equal to zero or all equal to one. In this situation, in order to avoid the extremes of $\pm\infty$, the supportive region $[b^*_{min}, b^*_{max}]$ will be based on only those individuals with non-constant response profiles. Verbeke and Molenberghs (2013) propose that a graph of the gradient function $\Delta(\hat{G}, \boldsymbol{b})$ can be used to assess the fit of an assumed random effects distribution $\hat{G}$. If the plot of $\Delta(\hat{G}, \boldsymbol{b})$ over a fine grid of $\boldsymbol{b}$ values does not exceed 1, and is equal to 1 within the supportive region I, then it suggests no other random effects distribution H can provide a better fit to the data. To identify true deviations from a gradient function of 1, pointwise confidence interval limits can be obtained about $\Delta(\hat{G}, \boldsymbol{b})$. As the asymptotic distribution of the gradient function is normal, the confidence interval limits of $\Delta(\hat{G}, \boldsymbol{b})$ are obtained based on the central limit theorem with the variance estimated as the sample variance of likelihood ratio contributions of $f_i(\boldsymbol{y_i}|\boldsymbol{b})/f_i(\boldsymbol{y_i}|G)$.

The graphical representation of the gradient function and corresponding pointwise confidence limits are used to identify potential distribution misspecification of the random effects. In the case of severe distributional misspecification, the gradient function and confidence bands will clearly exceed 1 within the support region I. Furthermore, the shape of the gradient function gives some indication of how the shape of the random effects distribution can be adapted to provide a better fit. For example, an increase in the likelihood can be achieved by replacing the random effects distribution with H for areas where the $\Delta(\hat{G}, \boldsymbol{b}) > 1$. Therefore, the model can be improved, in terms of log-likelihood, by moving probability mass from areas with small gradient function ($\Delta(\hat{G}, \boldsymbol{b}) < 1$) to areas with large gradient function ($\Delta(\hat{G}, \boldsymbol{b}) > 1$). The graphical representation of the gradient function to assess for distributional misspecification can be produced using SAS, implemented by using the syntax presented in Supplementary material of Verbeke and Molenberghs (2013).

3.3.2 Asymptotic diagnostic test based on the gradient function

A powerful diagnostic test based on the gradient function has been developed by Drikvandi et al. (2016) to formally test for misspecification of the random effects distribution for a general class of mixed models. The diagnostic test of Drikvandi et al. (2016) supplements the graphical diagnostic gradient function tool proposed by Verbeke and Molenberghs (2013), providing a formal way to determine whether the fluctuations observed in the gradient plot (detailed in Section 3.3.1) are due to distributional misspecification of the random effects, not just random variability (Drikvandi et al., 2016).

The test statistic is constructed by utilising the theoretical properties of the gradient function and using the Cramér-von Mises measure. For a general mixed model (including linear, generalised linear and non-linear mixed models), using similar notation as in Section 3.1 and 3.3.1, it is assumed that the response vector $\boldsymbol{y_i}$ has density $f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\xi}')'$ represents all B unknown parameters in the model (i.e. the parameters corresponding to the covariates and the random effects as parametrised through the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$, respectively).

Assuming the conditional distribution $f_i(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\theta})$ is correctly specified, the test uses the gradient function to assess the specification of random effects distribution. The null hypothesis of the test is that the assumed random effects distribution, G, is correctly specified. Following the derivation of the gradient function by Verbeke and Molenberghs (2013), the test statistic is constructed based on the Cramér-von Mises measure of distance between the gradient function $\Delta(G, \boldsymbol{b})$ (as defined in Equation 3.33) and the value of one. The test statistic is defined by $T(\boldsymbol{\theta}) = \int_{R^q} (\Delta(G, \boldsymbol{b}) - 1)^2 dG(\boldsymbol{b})$, evaluated for all possible values of $\boldsymbol{b}$ in the support of G. Under the null hypothesis it can be shown that $T(\boldsymbol{\theta}) = 0$. As $\boldsymbol{\theta}$ is unknown, the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ obtained under the null hypothesis can be used to derive the test statistic, thus,

$$T(\hat{\boldsymbol{\theta}}) = \int_{R^q} (\hat{\Delta}(\hat{G}, \boldsymbol{b}) - 1)^2 d\hat{G}(\boldsymbol{b}) \tag{3.34}$$

where $\hat{G}$ and $\hat{\Delta}(\hat{G}, \boldsymbol{b})$ are the estimated random effects distribution and gradient function, respectively, obtained by replacing $\boldsymbol{\theta}$ with the suitable estimator, $\hat{\boldsymbol{\theta}}$.

Let $\boldsymbol{\theta_0} = (\theta_{01}, ..., \theta_{0B})'$ denote the vector containing the true values for the B parameters. Drikvandi et al. (2016) derive the asymptotic distribution of $T(\hat{\boldsymbol{\theta}})$ under the null hypothesis, and show that

$$T(\hat{\boldsymbol{\theta}}) = \sum_{k=1}^{r} \lambda_k \chi_k^2 + o_p(1) \tag{3.35}$$

where $o_p(1) = \hat{\Delta}(\hat{G}, \boldsymbol{b}) - 1$, χ_k^2 (for $k = 1, ..., r$) are independent chi-squared random variables with one degrees of freedom, and $\lambda_1 \geq ... \geq \lambda_r$ are the r non-zero eigenvalues of $\boldsymbol{A}'Q(\boldsymbol{\theta_0})\boldsymbol{A}$, where $\boldsymbol{A}$ is the square root of the inverse Fisher Information matrix of the model parameters and $Q(\boldsymbol{\theta_0})$ is the $B \times B$ matrix with the (l, l')-th element

$$Q_{ll'}(\boldsymbol{\theta_0}) = \int_{R^q} \left(\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\partial}{\partial \theta_{0l}} \frac{f_i(\boldsymbol{y_i}|\boldsymbol{b})}{f_i(\boldsymbol{y_i}|\boldsymbol{G})} \right) \right) \times \left(\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} E\left(\frac{\partial}{\partial \theta_{0l'}} \frac{f_i(\boldsymbol{y_i}|\boldsymbol{b})}{f_i(\boldsymbol{y_i}|\boldsymbol{G})} \right) \right) dG(b). \tag{3.36}$$

The critical values of $T(\hat{\boldsymbol{\theta}})$ can be computed analytically and are based on the exact distribution of a weighted sum of independent chi-squared random variables (as derived by Imhof (1961)) (See Appendix B of Drikvandi et al. (2016)). As asymptotic results are often less adequate for data with small to moderate sample sizes (i.e. $N = 100$ to 300), Drikvandi et al. (2016) also derived a parametric bootstrap procedure to approximate $T(\hat{\boldsymbol{\theta}})$. However, as the sample sizes in panel surveys are typically large, the focus in this study will be restricted to the asymptotic test.

The test statistic is calculated using a quasi-Monte Carlo (QMC) integration method,

$$\begin{aligned} T(\hat{\boldsymbol{\theta}}) &= \int_{R^q} (\hat{\Delta}(\hat{G}, \boldsymbol{b}) - 1)^2 d\hat{G}(b) \\ &= \frac{1}{K} \sum_{k=1}^{K} (\hat{\Delta}(\hat{G}, b_k) - 1)^2 \end{aligned} \tag{3.37}$$

where $b_k = \hat{G}^{-1}(c_k)$ where c_k $(k = 1, ..., K)$ are the quasi-Monte Carlo integration nodes over the q-dimensional unit cube $C^q = [0, 1)^q$. The QMC approach is also used to approximate $\hat{Q}_{ll'}(\hat{\boldsymbol{\theta}})$ in order to calculate the eigenvalues, such that

$$\hat{Q}_{ll'}(\hat{\boldsymbol{\theta}}) = \frac{1}{KN^2} \sum_{k=1}^{K} \sum_{i=1}^{N} \sum_{i'=1}^{N} \left(\frac{\partial}{\partial \theta_l} \frac{f_i(\boldsymbol{y_i}|b_k)}{f_i(\boldsymbol{y_i}|G)} |_{\theta_l = \hat{\theta}_l} \right) \left(\frac{\partial}{\partial \theta_{l'}} \frac{f_{i'}(\boldsymbol{y_{i'}}|b_k)}{f_{i'}(\boldsymbol{y_{i'}}|G)} |_{\theta_{l'} = \hat{\theta}_{l'}} \right) \tag{3.38}$$

where the derivatives of the ratio of the conditional and marginal distributions are calculated by

$$\frac{\partial}{\partial \theta_l} \frac{f_i(\boldsymbol{y_i}|b_k)}{f_i(\boldsymbol{y_i}|G)} = \left(\frac{\partial}{\partial \theta_l} \log f_i(\boldsymbol{y_i}|b_k) - \frac{\partial}{\partial \theta_l} \log f_i(\boldsymbol{y_i}|G) \right) \frac{f_i(\boldsymbol{y_i}|b_k)}{f_i(\boldsymbol{y_i}|G)}. \tag{3.39}$$

The asymptotic test statistic can be estimated in SAS using NLMIXED and IML procedures, as implemented using the SAS syntax presented in Drikvandi et al. (2016).

Unlike the graphical diagnostic tool of Verbeke and Molenberghs (2013) that provides information about potential misspecification within a supportive region I, the test statistic $T(\hat{\boldsymbol{\theta}})$ appropriately evaluates the gradient function at all possible values of b in the whole support region of G. This is particularly advantageous in the case of logistic mixed models, where the graphical diagnostic tool could only be evaluated for people with non-constant response profiles.

3.4 Simulation study design

Simulation studies are commonly utilised to explore the robustness of GLMMs to misspecification of the random effects distribution. These techniques allow researchers to empirically estimate the sampling distribution of the parameters of interest, providing an avenue to assess the performance and accuracy of model estimates under different model assumptions and violations. Typically, simulation studies are designed to reflect the complex scenarios observed in practice. In Chapters 5 and 6, the impact of incorrectly assuming normality of a multimodal random intercept distribution in panel survey settings is assessed via simulation. In the simulation study, data are generated based on the HILDA case study presented in Chapter 4, and reflect similar characteristics as the HILDA panel survey, including missing data due to attrition. The simulation study generates a large number of datasets using the same random intercept logistic model considered in the HILDA case study, with the random effects generated from various three component mixture distributions of normals to represent different types of multimodal distributions. To each of these simulated datasets, the impact of misspecifying the random effects distribution on the parameter estimates and standard errors are assessed by fitting a random intercept logistic model assuming normal distributed random intercepts. The design of the simulation studies considered in Chapters 5 and 6 is described in more detail below. As detailed in Section 3.4.1, simulating longitudinal data requires a data generating model to simulate the responses for each individual and time-point. Furthermore, as the simulation studies presented in Chapters 5 and 6 aim to assess the impact of misspecified distributional assumptions of the random effects in the presence of missing data, attrition can be simulated by utilising a drop-out generating model as described in Section 3.4.2.

3.4.1 Data generating model

The data generating model is used to create S simulated datasets that will subsequently be analysed to assess the performance of statistical procedures. The aim is to create simulated datasets that have similar properties to the original data. Let N_s denote the number of individuals in the s^{th} simulated dataset ($s = 1, ..., S$), and assume that each individual has complete data and is observed at all n time-points. The aim of the data-generating model is to simulate n responses for each individual. In this study, the focus is on generating binary responses, and as such, the response vector is generated from the logistic mixed model,

$$\text{logit}\left(\Pr(y_{ij} = 1)\right) = \boldsymbol{x_{ij}}'\boldsymbol{\beta^0} + \boldsymbol{z_{ij}}'\boldsymbol{b_i} \tag{3.40}$$

where $\boldsymbol{x_{ij}}$ is a matrix of p covariates for individual i at time j, $\boldsymbol{\beta^0}$ is a vector of true parameters corresponding to the fixed effects, $\boldsymbol{z_{ij}}$ is the design matrix of the random effects and $\boldsymbol{b_i}$ is a vector of random effects for the i individual. For a random intercept logistic model, the $\boldsymbol{z_i}$ will be a vector of ones and the corresponding $\boldsymbol{b_i}$ would be the vector of random intercepts (b_{0i}).

There are numerous ways to simulate longitudinal data, and the method utilised in this study adapts the general framework detailed in Chapter 12 of Wicklin (2013) to simulate longitudinal binary data. The framework consists of the following steps:

1. Create design matrices for the fixed and random effects
2. Construct a diagonal matrix containing the variance and covariance components for the q random effects
3. Simulate the response vector as determined by a logistic mixed effects model.

The first step requires generating the matrices containing the variables of the fixed and random components in the mixed effects model. The design matrix for the fixed effects can either be generated from a multivariate normal distribution with data motivated mean and variance-covariance matrix, or by utilising the structure of the original data. To maintain the correlation structure of the explanatory variables across individuals and across time, the design matrix used for simulation studies in this study is obtained by utilising resampling techniques. For each iteration of the simulation, a random sample of N_s individuals was selected without replacement from the N individuals in the original data with complete case data. The explanatory variables of the selected N_s individuals were then used in the design matrix, $\boldsymbol{x_{ij}}$. Similarly, the design matrix of the random effects $\boldsymbol{z_{ij}}$ can either be manually created or could be obtained as the corresponding variables of the selected N_s individuals.

The second step is to generate the individual-specific random effects. The diagonal matrix containing the components of variance-covariance matrix allows correlation between the q random effects. Creating the diagonal matrix can either be manually specified or generated using estimated parameters from a fitted mixed effects model. The true random effects b_i are assumed to be distributed as either a continuous or discrete distribution, with the variance-covariance matrix inducing the variability and correlation.

Once the steps of creating the design matrices, $\boldsymbol{x_{ij}}$ and $\boldsymbol{z_{ij}}$, and the random effects ($\boldsymbol{b_i}$) have been constructed, the response vector can be simulated according to the logistic mixed model. For each of the N_s individuals, the linear predictor of that individual at time j is the sum of the linear predictor of the fixed effects ($\boldsymbol{x_{ij}}'\boldsymbol{\beta^0}$), and the random effects. ($\boldsymbol{z_{ij}}'\boldsymbol{b_i}$). The linear predictor η_{ij} is given by:

$$\eta_{ij} = \boldsymbol{x_{ij}}'\boldsymbol{\beta^0} + \boldsymbol{z_{ij}}'\boldsymbol{b_i} \tag{3.41}$$

where true parameter vector $\boldsymbol{\beta^0}$ used to generate the linear predictor of the fixed effects is either manually specified or set as the parameter estimates from a previous model fit. The linear predictor η_{ij} is used to generate the response y_{ij} as a Bernoulli random variable with expected value μ_{ij}, such that $y_{ij} \sim \text{Bernoulli}(\mu_{ij})$ where $\mu_{ij} = \exp(\eta_{ij})/(1 + \exp(\eta_{ij}))$ is the inverse logit transformation of η_{ij}.

3.4.2 Drop-out generating model

The drop-out generating model is used to simulate missing observations in the datasets generated in Section 3.4.1. After creating a simulated dataset consisting of response data for N_s individuals at n time-points, the drop-out generating model is applied to simulate missingness by setting responses y_{ij} to missing. The missing data mechanism characterises the reasons for missing data. We will restrict discussion to generating monotone missingness assuming the missing at random mechanism.

To describe the drop-out generating model, we first need to introduce the notation for missing data. Let R_{ij} be an indicator variable for whether individual i is observed at time j $(j = 1, ..., n)$:

$$R_{ij} = \begin{cases} 1 & \text{if } y_{ij} \text{ is observed for individual } i \text{ at time } j \\ 0 & \text{if } y_{ij} \text{ is not observed for individual } i \text{ at time } j \end{cases} \tag{3.42}$$

Therefore $R_{ij} = 1$ if individual i is observed at time j, and $R_{ij} = 0$ if individual i is missing at time j. The $n \times 1$ missing pattern indicator vector for individual i is given by, $\boldsymbol{R_i'} = (R_{i1}, R_{i2}, ..., R_{in})$.

If missing data are only due to monotone missingness, then the time that individual i drops out of the study is the smallest index j for which $R_{ij} = 0$, denoted m. The drop-out indicator d_{ij} follows on from the definition of R_{ij}, where $d_{ij} = 1 - R_{ij}$ for $j \leq m$. For all time-points after individual i has dropped out at time m, the drop-out indicator is set to missing (i.e. $d_{ij} = .$ for time-points $m+1, ..., n$). For individuals observed at all time-points, $d_{ij} = 0$ for $j = 1, ..., n$.

Let $p_{ij}(\alpha)$ denote the conditional probability of drop-out at time j for individual i, given the available data observed up to time $j-1$, such that:

$$p_{ij}(\alpha) = \Pr(d_{ij} = 1 | d_{i(j-1)} = 0, y_{11}, ..., y_{i(j-1)}, \boldsymbol{x_{ij}}; \boldsymbol{\alpha}) \tag{3.43}$$

where $y_{i1}, ..., y_{i(j-1)}$ is the history of the responses up to time $j-1$, and $\boldsymbol{x_{ij}}$ and $\boldsymbol{\alpha}$ are the covariates and corresponding regression coefficients. We assume that all individuals are observed at the first time-point, such that $d_{i1} = 0$ and $p_{i1}(\alpha) = 1$. The probability of drop-out, $(p_{ij}(\alpha))$, is estimated from observed data by fitting an ordinary logistic model,

$$\text{logit}(p_{ij}(\alpha)) = \boldsymbol{\alpha}'\boldsymbol{w_{ij}} \tag{3.44}$$

where $\boldsymbol{w_{ij}}$ is a vector of covariates which may contain values of $\boldsymbol{x_{ij}}$ and current or previous response values $y_{i1}, ..., y_{i(j-1)}, y_{ij}$. According to the dependence of the missingness mechanism on the response pattern, $\boldsymbol{y_i}$, the missingness mechanism can be classed as either MCAR, MAR or MNAR. For example, if drop-out is assumed to be MCAR, then the conditional probability of drop-out is not related to the response process and only dependent on the covariates. A MAR drop-out process is dependent on the observed response components ($\boldsymbol{y_i^O} = (y_{i1}, ..., y_{i(j-1)})'$) and covariates. In the scenario of MNAR drop-out, the probability of drop-out not only depends on the previous response value but also on the current or future response values and the covariates.

The conditional probability of drop-out for individual i at time j is given by:

$$p_{ij}(\alpha) = \frac{\exp(W'_{ij}\boldsymbol{\alpha})}{1 + \exp(W'_{ij}\boldsymbol{\alpha})}. \tag{3.45}$$

To simulate drop-out, for each individual i at each time-point j with complete data, the conditional probability of drop-out is calculated based on Equation 3.45 using coefficients $\hat{\boldsymbol{\alpha}}$ estimated from the observed data. The conditional probability of drop-out $p_{ij}(\alpha)$ is compared to a random draw from a uniform distribution, $u_{ij} \sim U[0, 1]$ (Bonate, 2011). If $u_{ij} < p_{ij}(\alpha)$ then the individual is dropped for that time point and subsequent time-points, such that the response for individual i at time j and subsequent time-points $(j + 1, ..., n)$ is set to missing.

3.5 Simulation study performance measures

After the simulations have been generated and analysed, performance measures are used to evaluate the adequacy of the model by comparing the simulated results with the true results. For each of the S simulations generated using the methods described in Section 3.4, a coefficient estimate of interest $\hat{\beta}_s$ (for $s = 1, ..., S$) from the logistic random intercept model is produced. Performance criteria are used to compare summary measures of $\hat{\beta}_s$ with the true value β^0, and give an assessment of bias, coverage and accuracy. For example, the average estimate of interest ($\bar{\hat{\beta}} = \sum_s^S \hat{\beta}_s/S$) is commonly used as a summary measure of the S simulations. The performance criteria that were used are the percentage bias, coverage of the confidence intervals and the standard error ratio. These performance criteria and limits indicating acceptable performance are described in more detail below.

3.5.1 Percentage bias of parameter estimates

Percentage bias is defined as the bias as a percentage of the true parameter value, where bias is the difference between the mean of the simulation estimates and the true parameter value. The percentage bias is calculated as:

$$((\bar{\hat{\beta}} - \beta^0)/\beta^0) \times 100. \tag{3.46}$$

Criteria for acceptable performance is a percentage bias within -10% and 10% (Marshall et al., 2010).

3.5.2 Coverage rates of confidence intervals

The coverage rate is defined as the proportion of 95% confidence intervals that contain the true parameter value for converged models. Of the S Monte Carlo simulations, the coverage rate is the proportion of times the interval $\hat{\beta}_s \pm 1.96 \times SE(\hat{\beta}_s)$ includes β^0, where $SE(\hat{\beta}_s)$ is the standard error of the estimate of interest (for $s = 1, ..., S$). Criteria for acceptable performance is that the coverage should be within 2× SE of the nominal coverage probability (P) (Burton et al., 2006), where $\mathrm{SE}(P) = \sqrt{(P(1-P)/S)}$. Therefore, for 1000 Monte Carlo simulations, coverage rates of the 95% confidence intervals should be within 93.6% and 96.4% to be considered appropriate.

3.5.3 Standard error ratio

The standard error ratio is defined as the ratio of the mean of the standard error estimates to the standard deviation of the parameter estimates over the S Monte Carlo simulation iterations,

$$\frac{\frac{1}{S}\sum_{s=1}^{S}\mathrm{SE}(\hat{\beta}_s)}{\sqrt{\frac{1}{S-1}\sum_{s=1}^{S}(\hat{\beta}_s - \bar{\hat{\beta}})^2}}. \tag{3.47}$$

The criteria for acceptable accuracy is defined as standard error ratios within 0.9 and 1.1, indicating that the model-based standard error estimates accurately describe the variability of the coefficient estimators (Neuhaus et al., 2013).

4 Case Study of Women's Employment Participation using the HILDA survey

4.1 Introduction

Raising women's employment participation rate has become an important policy priority in many developed economies (Jenkins, 2006). In Australia, there have been a number of policy and regulatory changes to further support women's participation in the workforce, including initiatives and reforms regarding parental leave to facilitate women returning to paid work after child-birth (Tannous and Smith, 2013). The reforms may have been successful in increasing the levels of participation in the labour force for working aged women of child-bearing potential. In the last 50 years the employment participation rate for females aged 25-44 years increased from 36.6 percent in August 1966 to 71.4 percent in December 2014 (ABS, 2007, 2014). However, gender differences between the workforce engagement are apparent (Tannous and Smith, 2013). For example, in December 2014 women aged 25-44 years accounted for over 30% of the population aged 25-64 years that were not in paid work, with the comparable figure for men being 14% (ABS, 2014). The differing patterns of employment for women have been attributed to caring for children, other care roles and household responsibilities (House of Representatives Standing Committee on Employment and Workplace Relations, 2009).

Given the changes in Australian employment trends over the past 50 years, understanding women's labour force participation patterns is paramount for policy makers. Apart from modelling trends, research is required to understand the determinants of labour force participation over time. These areas of research have recently attracted increasing interest because of concerns that an ageing population will put downward pressure on labour supply, subsequently impacting material living standards and public finances (Jaumotte, 2003). An increase in female participation has been suggested to partially resolve this problem (Burniaux et al., 2003). Longitudinal studies can provide a rich source of data to address these research questions, providing an opportunity to trace employment patterns and investigate inter-relationships among employment outcomes and events over the life-course. One panel survey increasingly used to assess relationships between employment trends and life course events in Australia is the annual Household Income and Labour Force Dynamics in Australia (HILDA) survey. The HILDA survey started in 2001 and collects annual and monthly employment data for individuals aged over 15 years of age in over 7000 households across Australia. Researchers have utilised the HILDA survey to examine, for example, the effect of health (Cai and Kalb, 2006; Austen and Ong, 2010; Honey et al., 2014; Frijters et al., 2014), children (Tannous and Smith, 2013; Parr, 2012) and family dynamics (Baxter and Renda, 2011) on employment participation and employment transitions of working aged Australians.

A challenge when analysing panel data is to control for the impact of unobserved heterogeneity among individuals in order to obtain valid inferences of model parameters (Hsiao, 2007). One approach is to use fixed effects models, where the time-constant heterogeneity is allowed to be correlated with the explanatory variables. Fixed effect models rely on differencing out the time-constant explanatory variables as well as the unobserved heterogeneity, removing the issue of endogeneity. This approach is useful when the effects of covariates on within-individual change over time are of interest. For example, Tannous and Smith (2013) used fixed effects logistic models to assess the association between child-birth and working part-time using 10 years of HILDA data, and Frijters et al. (2014) used fixed effect logistic models to assess the effect of mental health on employment using waves 2 to 11 of the HILDA data.

However, if the unobserved between-individual effects of time-invariant variables are also of interest, then generalised linear mixed models (GLMMs) are a commonly used approach to the analysis of longitudinal panel data. Furthermore, GLMMs can provide consistent estimation in settings where the underlying mechanism for missing data is MAR[1]. In GLMMs, the unobserved heterogeneity is assumed to be captured by a random variable, known as a random effect. These models are useful as they accommodate the dependence of repeated observations within individuals and also the between individual variability through the inclusion of fixed and individual-specific random effects, respectively. For example, using monthly calendar HILDA data over a seven year period, Baxter and Renda (2011) used random effect logistic models to assess differences between lone and couple mothers leaving or entering employment, whilst Feeny et al. (2012) used a random intercept logistic model to examine whether employment outcomes of Australian labour market programme participants vary according to receiving housing assistance over six waves of the HILDA survey. The parameters of interest for GLMMs are often estimated using maximum likelihood, typically under the assumption of Gaussian distributed random effects with mean zero and fixed variance-covariance matrix.

However, this assumption of normality may not be appropriate in practice. For example, multimodality of the random effects may occur if a key categorical variable is omitted from the model or latent sub-populations exist. In this chapter we demonstrate the potential multimodality of the random intercept distribution in an application of a random intercept logistic model to assess employment participation of working aged women using data from HILDA. For this case study, and to illustrate the methods developed in this study, employment participation is considered as a dichotomous variable representing employed (full or part-time)

[1] See Section 2.3.3 for further details.

and non-employed (including unemployed and not in labour force). Over 11 waves of HILDA, 10% of women were observed to always be non-employed, 45% of women were observed to transition at least once between the two employment states, and the remaining 45% were observed to always be employed. In this case the extreme response pattern may be influenced by an underlying mover-stayer scenario, whereby over the study period some women have a high propensity to remain in the same employment state (stayers), whilst others are susceptible to change employment states (movers). In a mover-stayer scenario (Blumen et al., 1955), the people that have been observed to transition states are movers, and the people that have been observed to stay in the same state will consist of both the latent stayers and any movers that did not change states over the study period (Lindsey, 1997). Therefore, the random intercepts capturing the unobserved heterogeneity may be dominated by three sub-populations: one with an extremely low propensity of ever experiencing the outcome, a more heterogeneous group transitioning between states over time, and one with an extremely high propensity of always experiencing the outcome. Hence, the assumed normal distribution of the random intercept may not appropriately capture the heterogeneity of the latent mover-stayer scenario.

To guard against the impact of misspecifying the random effect distribution, the parametric normality assumption can be relaxed by using semi-parametric (Chen et al., 2002; Vock et al., 2014) or non-parametric methods (Laird, 1978; Heckman and Singer, 1984; Aitkin, 1999; Lesperance et al., 2014). Alternatively, flexibility can be achieved by modelling the random effects as non-Gaussian distributions using computational methods such as the probability integral transformation (Nelson et al., 2006) and the likelihood reformulation (Liu and Yu, 2008). In the case of an underlying mover-stayer scenario, statistical methodology has been developed to identify and quantify the latent movers and stayers. One such methodology incorporates discrete masses at negative and positive infinity in the random effects distribution to represent the two stayer subgroups (Davies et al., 1992; Berridge and Crouchley, 2011a). In the specific case where the mover-stayer is dominated by stayers in the null state, Carlin et al. (2001) proposed a Bayesian discrete mixture to allow the subgroup of individuals immune to a binary outcome to be modelled by a discrete mass at negative infinity, while random variability within the susceptible subgroup is modelled by a logistic mixed model. Alternatively, modelling the random intercepts as a finite mixture of normal distributions (Magder and Zeger, 1996; Verbeke and Lesaffre, 1996) may be plausible, where three components potentially capture the three latent sub-populations.

In this chapter we undertake a longitudinal analysis of women's employment participation using eleven waves of the HILDA survey. To investigate potential misspecification of the random intercept distribution, we fit two random intercept logistic models assuming different random effect distributions and assess the fit using diagnostic tests. The first model assumes the random intercept is normally distributed, whilst the second model assumes a three component mixture of normal distributions. By considering two different assumed random effect distributions, this chapter demonstrates potential multimodality of the random intercept distribution and investigates the practicality of implementing flexible random effects in panel survey applications. The aim of this case study is not to quantify or address the potential latent mover-stayer scenario, nor to comprehensively analyse determinants of women's employment participation. The focus of this chapter is to demonstrate potential misspecification of the random effect distribution in an application of a random intercept logistic model to panel survey data.

4.2 Data and variables

The case study models employment participation of women from 2001 to 2011 using 11 waves of data from the HILDA survey. The HILDA survey is a nationally representative household panel survey conducted annually, with the first wave of data collection starting in 2001. The primary focus of HILDA is to collect information about economic and subjective well-being, and labour market and family dynamics in Australia. The details of the survey are given in Watson and Wooden (2012), and are briefly described here. In the first wave, 7,682 households of all in-scope households were interviewed, resulting in a sample of 15,127 eligible persons aged 15 years or older. Of those, data for 13,969 sample members were collected through the successful completion of personal interviews. In addition to the personal interviews, respondents were required to return a self-completion questionnaire.

Women of child-bearing age between 30 and 44 years at June 2001 from the HILDA survey were selected to represent working aged women in Australia and to avoid women in their early careers. As commonly experienced in longitudinal surveys, the HILDA survey was subject to the problem where some sample members are lost at each successive wave, either at one time-point (intermittent missing) or lost permanently (attrition). Of the total 2340 women aged 30 to 44 with valid employment data at the first wave, 1359 (58.1%) had complete employment history for all 11 waves, 413 (17.6%) had intermittent missing and 568 (24.3%) dropped out of the survey (attrition). Investigation to explore potential reasons for the observed missingness and relationship to the response variable, employment status, is beyond the scope of this study. However, there is evidence suggesting employment status is related to likelihood of responding (conditional on making contact) to the HILDA survey (Watson and Wooden, 2009). Using the first four waves of the HILDA survey, Watson and Wooden (2009) identified that even though employed people were easier to contact, they were less likely to respond if they worked full-time hours (35 or more hours per week). If missingness is not related to the response variable, and thus ignorable, maximum likelihood estimation will produce consistent estimates as the missing data generating mechanism is missing at random (MAR) (Molenberghs and Verbeke, 2005; Skrondal and Rabe-Hesketh, 2008). Therefore, if the missingness is ignorable, it would be expected that analysis restricted to the complete cases, or analysis of unbalanced data due to missingness will produce similar results. To explore whether missingness is related to the response variable, the following analysis will focus on two sub-samples: the 1359 women with complete cases (those with complete employment history); and the 1927 women with monotone

Table 4.1: Number of respondents (n) and cumulative attrition rate as percentage of original HILDA sample (%) of 1927 women aged between 30 and 44 years at Wave 1

Wave	Respondents (n)	Cumulative Attrition (%)
1	1927	0
2	1759	8.7
3	1647	14.5
4	1580	18.0
5	1540	20.1
6	1508	21.7
7	1474	23.5
8	1450	24.8
9	1421	26.3
10	1391	27.8
11	1359	29.5

missingness (those who experienced attrition or had complete case data). Of the 1927 women at the first wave with monotone missingness, the number of women observed at each wave and cumulative attrition rate are shown in Table 4.1, whereby 29.5% had dropped-out by the eleventh wave.

The HILDA survey contains detailed information about labour force participation and history. Labour force participation details were collected for every wave of the HILDA survey using both personal interviews and self-completion questionnaires. In this case study, the binary response variable (y_{ij}) represents employment status for individual i at wave j which equals 1 for women in part-time or full-time employment (employed), and 0 for women who are unemployed or not in the labour force (not employed). The employment status was based on the derived detailed current labour force status variable, whereby women not in the labour force includes those who are or are not marginally attached. Marginal attachment to the labour force is defined as a person who is not in the labour force, who wants to work and either (i) is actively seeking work but are not available to start work; or (ii) is not actively seeking work but is available to start work within four weeks (ABS, 2001). A person is not marginally attached to the labour force if they are not in the labour force and either (i) is not wanting work; or (ii) is wanting work though not actively seeking work and is not available to start working within four weeks (ABS, 2001).

The analysis is restricted to a small number of key explanatory variables, including the woman's age, current marital status, highest level of education achieved and the presence of young and dependent children. These variables are similar to standard predictors used in analyses of women's employment participation (e.g. Jenkins 2006; Parr 2012; Tannous and Smith 2013). The respondent's age is a continuous variable and is included as a linear term in the model (x_{1ij}). Marital status for individual i at wave j is a three category variable with categories for married or de-facto (reference); separated, divorced or widowed (x_{2ij}); and single (x_{3ij}). The highest educational qualification attained at the first wave for individual i is categorised as Bachelor degree or higher (reference); Year 12 or Diploma and Certificate (x_{4i}); or Year 11 and less (x_{5i}). Dependent children for individual i at wave j is categorized as no dependent children (reference); youngest dependent child aged 4 years old or younger (x_{6ij}); and youngest dependent child aged 5 to 24 years old (x_{7ij}). The respondent's age captures cohort differences in employment among women, wave effects and also potential employment experience. Family commitments that may potentially impact the capacity of a woman to enter paid employment are captured by marital status and age of youngest dependent child. The highest level of education attained is used as an index of human capital.

4.3 Statistical models and estimation

As discussed previously in Section 2.3, a challenge when analysing panel data is to control for the impact of unobserved heterogeneity among individuals in order to obtain valid inferences of model parameters (Hsiao, 2007). The analysis of the employment participation data also requires a model to take into account the dependency of repeated observations within individuals over time. One approach that accounts for unobserved heterogeneity and correlation within individuals, is the generalised linear mixed model detailed in Chapter 2 and Section 3.1. For the analysis of binary employment data, a logistic random effects model can be used to explore sources of individual-to-individual variability in the propensity to be employed. The following random intercept logistic model is used to model employment participation of working aged women over the 11 waves of HILDA:

$$\begin{aligned}\text{logit}[\Pr(y_{ij} = 1|b_i)] = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \\ \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6ij} + \beta_7 x_{7ij} + b_i \end{aligned} \tag{4.1}$$

where b_i is the random intercept, β_0 denotes the intercept coefficient and β_1 to β_7 are the corresponding fixed effect parameter coefficients of the explanatory variables x_{1ij} to x_{7ij} (as defined in Section 4.2). The random intercept in Equation 4.1 captures the unobserved heterogeneity that is unable to be captured by the fixed effects. The model in Equation 4.1 is analysed assuming the random intercept is normally distributed with zero mean and variance σ_b^2 ($b_i \sim N(0, \sigma_b^2)$).

However, the assumed normal distribution for the random intercept may not appropriately capture the heterogeneity if there exists an underlying mover-stayer scenario. To further investigate potential misspecification of the random effect distribution, the normality assumption is relaxed by estimating the model assuming random intercepts are distributed as a finite mixture of normal distributions (Verbeke and Lesaffre, 1996; Verbeke and Molenberghs, 2009). Finite mixtures of normals are very flexible and can capture a range of distributions, including multimodal distributions (Verbeke and Molenberghs, 2013). Hence, the potential multimodality of the latent mover-stayer scenario could be captured by the three component mixture of normal distributions. Therefore, the logistic random intercept model in Equation 4.1 is also fitted assuming the random effects are distributed as

$$b_i \sim \pi_1 N(\mu_1, \sigma_1^2) + \pi_2 N(\mu_2, \sigma_2^2) + \pi_3 N(\mu_3, \sigma_3^2) \tag{4.2}$$

where π_k, μ_k and σ_k^2 are the mixing proportions, component means and component variances for the $k = 1, 2, 3$ components, respectively. The likelihood reformulation method (Liu and Yu, 2008) can be used to obtain maximum likelihood estimates for mixed models assuming non-Gaussian random effect distributions. As detailed in Section 3.2.1, the likelihood reformulation method is based on transforming the conditional density on non-normal random effects to one that can be integrated over a normal random effects distribution. To ensure correct parametrisation of the three component mixture of normals, the assumption $E(b_i) = 0$ was imposed by the restriction $E(b_i) = \pi_1\mu_1 + \pi_2\mu_2 + \pi_3\mu_3 = 0$, such that the $\hat{\beta}_0$ was estimated as the expected value of the random intercepts. The restriction of the mixing proportions summing to one was imposed by $\pi_1 + \pi_2 + \pi_3 = 1$, such that the $\hat{\pi}_3$ was estimated as $1 - \hat{\pi}_1 - \hat{\pi}_2$. The total random effect variance, denoted σ_b^2, was estimated based on the variance of a finite mixture, such that $\hat{\sigma}_b^2 = \sum_{k=1}^{3} \hat{\pi}_k(\hat{\sigma}_k^2 + \hat{\mu}_k^2) - (\sum_{k=1}^{3} \hat{\pi}_k\hat{\mu}_k)^2$.

To assess the model fit of the assumed random effect distribution, either normal or a mixture, the gradient function exploratory diagnostic tool (Verbeke and Molenberghs, 2013) was used to identify potential random effect distributional misspecification. As detailed in Section 3.3.1, the gradient function diagnostic tool is a graph of the gradient function ($\Delta(G, \boldsymbol{b})$) as a function over the random effect ($\boldsymbol{b}$). The graphical representation of the gradient function and corresponding pointwise confidence limits are used to assess the fit of the assumed random effect distribution, giving an indication of how the distribution can be adapted to improve the model fit. The gradient function diagnostic tool only provides information about the shape of the random effect distribution within a supportive region, located in the interval of b^*_{min} and b^*_{max}. However, as there are extreme response profiles when modelling a binary response, the region for which the gradient function can provide information about the random effect distribution covers the whole real line, from $-\infty$ to ∞. Therefore, to avoid the extremes of $\pm\infty$, the supportive region with limits b^*_{min} and b^*_{max} used to assess the fit for the logistic model in Equation 4.1 is estimated based on women with non-constant response profiles (i.e. the observed movers) as detailed in Section 3.3.1.

The graphical diagnostic tool of Verbeke and Molenberghs (2013) is an exploratory tool to identify potential misspecification. To formally diagnose misspecification of the random effects distribution, the asymptotic diagnostic test (Drikvandi et al., 2016) was utilised. As detailed in Section 3.3.2, the diagnostic test of Drikvandi et al. (2016) is also based on the gradient function, and supplements the graphical diagnostic tool to formally test whether any fluctuations identified by the graphical tool is due to distributional misspecification of the random effect and not just due to random variability. The test statistic $T(\hat{\boldsymbol{\theta}})$ appropriately evaluates the gradient function at all possible values of b, within the whole support region. Therefore, in the case of the random intercept logistic model, the diagnostic tool will be evaluated on the whole real line, utilising information for people with non-constant and constant response profiles.

All analyses were undertaken using SAS (Version 9.4, SAS Institute, Cary NC). The logistic models assuming normal random intercepts were fit using the SAS procedure NLMIXED with adaptive Gaussian quadrature and 20 quadrature points. The logistic models assuming random intercepts distributed as a three component mixture of normals were estimated using the likelihood reformulation method implemented using SAS procedure NLMIXED (Appendix A contains the relevant SAS syntax). The graphical gradient function diagnostic tool and the asymptotic diagnostic test to assess the goodness of fit of the assumed random intercept distribution were both implemented in SAS. All analyses were performed for the two sub-samples: the 1359 women with complete case data (Complete Case) and the 1927 women with monotone missingness (Monotone Missing).

4.4 Results

Descriptive statistics of the response and the explanatory variables at the first wave for the two data analysis sub-groups are presented in Table 4.2, where continuous variables are summarised by means and standard deviations, whilst categorical explanatory variables are summarised as frequency and percentage. At the first wave the employment rates were similar for the two data analysis sub-groups, with 69.7% and 68.0% of women employed in the complete case and monotone missing data sub-groups. The demographic characteristics for the two data analysis sub-groups are similar, with a slightly larger proportion of women with complete case data having Bachelor degrees or higher (28.4%) than the monotone missing sub-group (25.4%). The average age of the women in the sample was 37 years of age (SD=4.2), with the majority of women married or in a defacto relationship (77%), completed Year 12 or tertiary diplomas and certificates (40%) and having the youngest dependent child aged 5 to 24 years (43%).

The model estimates assuming random intercepts as a normal and finite mixture of normal distributions, respectively, for the complete cases and monotone missing data, are presented in Table 4.3. Due to boundary issues, a formal comparison of the log-likelihood using a likelihood ratio test will not follow standard rules. Therefore, comparison of the residual deviance ($-2ll$) will be used as an indication of model fit. The residual deviance of the random intercept logistic model applied to the complete cases decreased from 9697 for the assumed normal to 9691 for the assumed mixture. Similarly, the deviance decreased from 11543 to 11526 for the monotone missing data. This suggests that the fit of the model improved marginally when random intercepts were fitted as a three component mixture of normals. Although the assumed random effects distributions were very different, the inference for the fixed parameters for each of the assumed distributions in the two missing data scenarios were similar. The estimated random effect variance (σ_b^2) was larger for the assumed normal random effects distribution

Table 4.2: Employment status and demographic characteristics at wave 1 for the 1359 women with complete case data (Complete Case) and the 1927 women with monotone missing data (Monotone Missing)

Demographic	Complete Case (N=1359)	Monotone Missing (N=1927)
Employment Status - n(%)		
Not Employed	412 (30.3%)	616 (32.0%)
Employed	947 (69.7%)	1311 (68.0%)
Age- mean (SD)	37.3 (4.2)	37.2 (4.2)
Marital Status - n(%)		
Married/defacto	1049 (77.3%)	1487 (77.3%)
Separated/Divorced/Widowed	167 (12.3%)	238 (12.4%)
Single	141 (10.4%)	199 (10.3%)
Highest Education - n(%)		
Bachelor or higher	386 (28.4%)	489 (25.4%)
Year 12/Diploma/Certificate	546 (40.2%)	779 (40.4%)
Year 11 or less	427 (31.4%)	658 (34.2%)
Dependent Children - n(%)		
None	330 (24.3%)	481 (25.0%)
Youngest aged < 5	446 (32.8%)	607 (31.5%)
Youngest aged 5-24	583 (42.9%)	839 (43.5%)

than for the mixture distribution for both missing data scenarios. However, larger standard errors for σ_b^2 were estimated for the mixture random intercept for both missing data scenarios. The estimated coefficients and standard errors for the parameters in the linear predictor and the random component were similar for the complete case data and monotone missing data scenarios, for both the assumed random intercept distributions.

The estimated random intercept distributions for the two assumed distributions for women with complete cases and monotone missing data are shown in Figure 4.1 (a) and (b), respectively. The estimated random intercept variance (σ_b^2) is large for both analysis subgroups and both assumed random effect distributions. The substantial heterogeneity may be explained by the extreme response pattern influenced by the potential underlying mover-stayer scenario. Of the 1359 women with complete employment data for all 11 waves, 103 (7.6%) were continuously non-employed, 625 (46%) transitioned between the two employment states, and the remaining 631 (46.4%) were continuously employed. Including the 568 women who dropped out of the HILDA survey, the proportion of women always employed (48.5%) was similar, however the proportion never employed (12.6%) or transitioning (38.9%) differed.

Assuming the random effects are a finite mixture of normal distributions also supports the extreme response patterns, where the components in the mixture correspond to three sub-populations (Table 4.3). For the women with complete case data, the first component with the predicted proportion of 12.4% and fitted average intercept $\hat{\mu}_1 = -5.44$ represents the women

Table 4.3: Parameter estimates (standard errors) for random intercept logistic model assuming normal or three component mixture of normal random effects applied to the HILDA case study for women with complete case data and monotone missing data

Parameter (Coefficient)	**Complete Case** ($N = 1359$)		**Monotone Missing** ($N = 1926$)	
	Normal	Mixture	Normal	Mixture
Constant (β_0)	1.20 (0.50)	1.07 (0.46)	1.35 (0.46)	0.61 (0.43)
Age (β_1)	0.09 (0.01)	0.09 (0.01)	0.09 (0.01)	0.10 (0.01)
Marital Status				
Married/defacto	ref	ref	ref	ref
Sep/Div/Wid (β_2)	-0.28 (0.14)	-0.31 (0.14)	-0.25 (0.13)	-0.31 (0.14)
Single (β_3)	-0.12 (0.28)	-0.12 (0.26)	-0.19 (0.25)	-0.07 (0.25)
Highest Education				
Bachelor or higher	ref	ref	ref	ref
Year 12/Dip/Cert (β_4)	-1.53 (0.28)	-1.52 (0.25)	-1.64 (0.25)	-1.56 (0.25)
Year 11 or less (β_5)	-2.77 (0.29)	-2.82 (0.27)	-2.92 (0.26)	-2.81 (0.26)
Dependent Children				
None	ref	ref	ref	ref
Youngest < 5 (β_6)	-2.33 (0.16)	-2.33 (0.15)	-2.35 (0.15)	-2.28 (0.14)
Youngest 5-24 (β_7)	-0.39 (0.12)	-0.40 (0.12)	-0.44 (0.12)	-0.40 (0.12)
Random Effect				
Variance (σ_b^2)	11.82 (0.86)	9.07 (1.29)	11.81 (0.79)	9.27 (0.90)
μ_1		-5.44 (0.66)		-4.79 (1.31)
σ_1		1.12 (0.42)		2.10 (0.67)
π_1		0.12 (0.04)		0.17 (0.10)
μ_2		-0.70 (0.34)		-0.28 (0.46)
σ_2		1.50 (0.51)		1.53 (0.87)
π_2		0.55 (0.14)		0.53 (0.25)
μ_3		3.29 (0.64)		3.26 (0.54)
σ_3		0.84 (0.66)		0.62 (0.54)
π_3		0.32 (0.11)		0.30 (0.16)
$-2ll$	9697	9691	11543	11526

with the propensity to be continuously non-employed. The second component with predicted proportion of 55.3% and fitted average intercept $\hat{\mu_2} = -0.70$, represents the women transitioning between the two employment states. The third component with predicted proportion of 32.3% and fitted average $\hat{\mu_3} = 3.29$ represents the women with the propensity to be continuously employed. A similar random effects distribution was estimated for the monotone missing subgroup (Table 4.3), with 17.2% in the first component with fitted average $\hat{\mu_1} = -4.79$. The second component with predicted proportion of 53% and fitted average of $\hat{\mu_2} = -0.28$, and the third component with predicted proportion of 29.9% and fitted average $\hat{\mu_3} = 3.26$. In both missing data scenarios, the estimated proportions for the first component were larger than the observed proportion of women continuously non-employed, whilst the estimated proportions for the third component were lower than the observed proportion of women continuously employed.

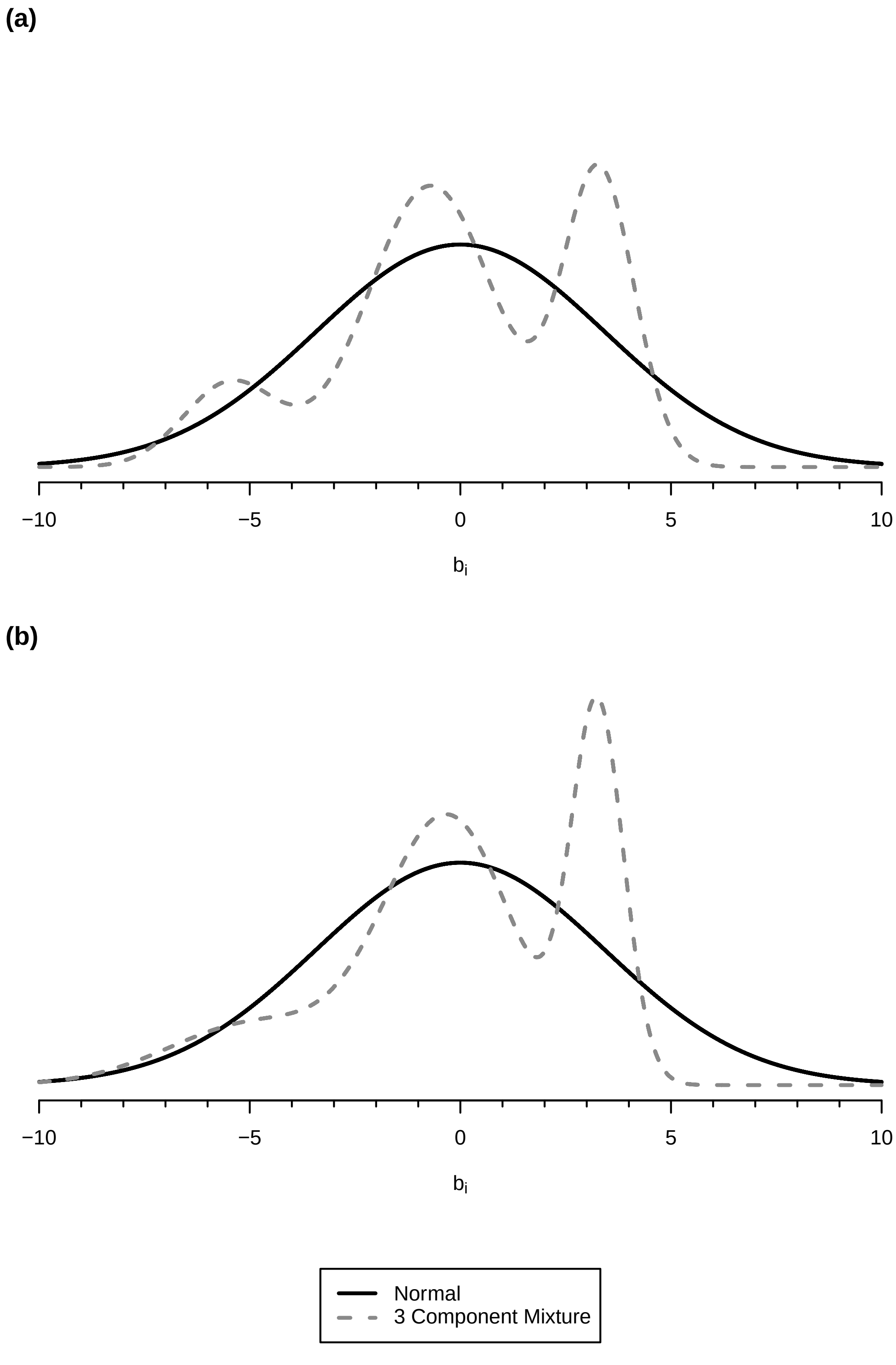
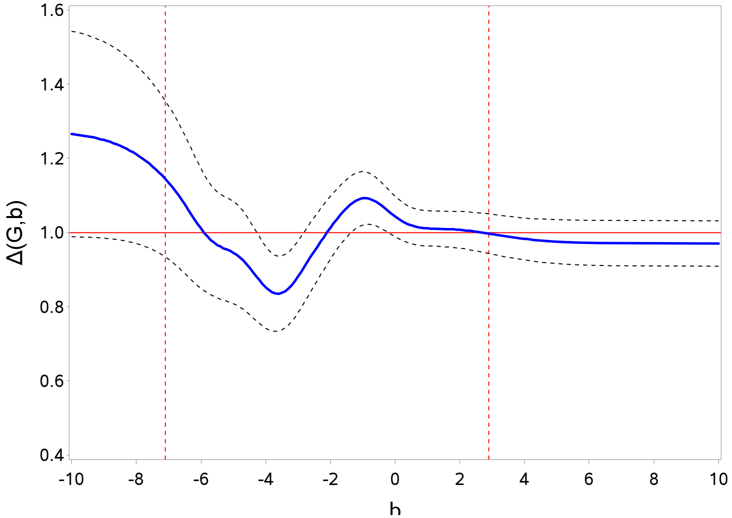


Figure 4.1: Estimated random intercept distribution for the random intercept logistic model assuming normal and three component mixture of normal distributions when applied to (a) women with complete case data and (b) women with monotone missing data.

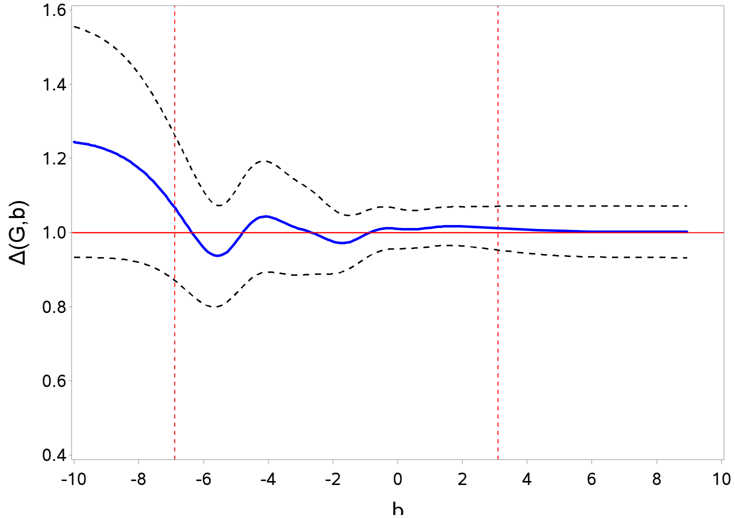
To assess the model fit of the assumed random effects distribution, the gradient function relating to the fitted random intercept logistic model for both data scenarios (Complete Case and Monotone Missing) and the assumed distributions (Normal and Mixture) are shown in Figure 4.2. Figure 4.2 shows a plot of the gradient function ($\Delta(G, \boldsymbol{b})$) over the values of the random intercept ($\boldsymbol{b}$) (solid blue line) and its corresponding 95% pointwise confidence interval band (dashed black lines). The gradient function provides information about the shape of the random effect distribution within the range of b^*_{min} and b^*_{max} (dashed red vertical lines). The shape of the gradient function gives an indication of how the random effects distribution can be adapted, with gradient function of 1 indicating optimal fit (solid red horizontal line).

To identify potential random effect distribution misspecification, Figure 4.2(a) shows a plot of the gradient function for the random intercept logistic model applied to the complete case data assuming normally distributed random intercepts. Within the supportive region of [-7.1, 2.9] (dashed red vertical lines), the gradient function deviates from 1 with the confidence bands not including the value of 1 in two regions. This suggests that the assumed random intercept distribution might be misspecified, and that fit of the model can be improved by moving probability from the region [-4, -2] to the region [-2, 0]. To formally test whether the assumed normal distribution is misspecified, the asymptotic diagnostic tool is used to determine whether the fluctuations observed in the gradient function is due to misspecification and not just random variability. The asymptotic test produces a test statistic of $T = 0.005$, which results in a p-value of 0.025 ($U = 443302$). The significant asymptotic test suggests that the assumed normal distribution for the random intercept is inadequate. Fitting the random intercept logistic model to the complete data assuming a three component mixture of normal distributions appeared to improve the model fit (Figure 4.2(b)). The gradient function shows some fluctuation around 1, yet the confidence bands contain 1 within the supportive range of [-6.9, 3.1]. To investigate this formally, the asymptotic test produces a test statistic of $T = 0.0006$ which corresponds to a p-value of 0.95 ($U = 248030$). Therefore, the gradient function graphical tool and the insignificant asymptotic test concludes there is no evidence that a substantial improvement can be achieved by further refinement of the random effect distribution when assuming the random intercepts are distributed as a three component mixture of normals. Thus, the three component mixture of normals is adequate for the random intercept.

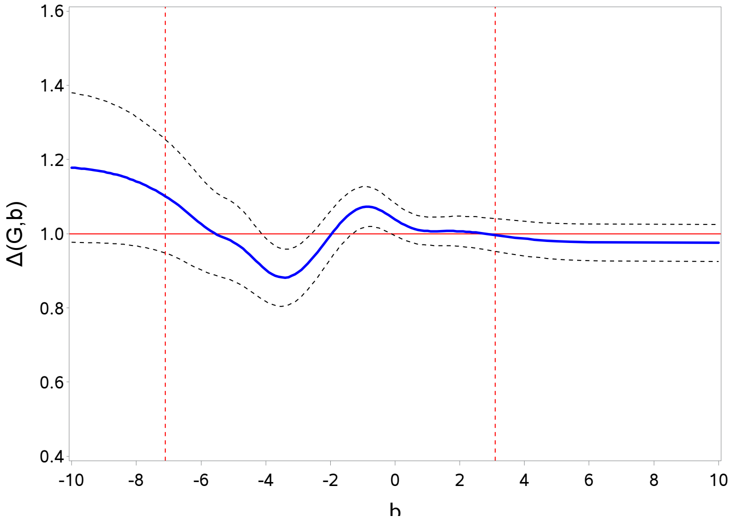
As for the complete case data scenario, the gradient function diagnostic tool for the monotone missing data suggested similar potential misspecification for the assumed normal distribution (Figure 4.2(c) and (d)). Figure 4.2(c) shows the gradient function for fitting the random intercept logistic model assuming normal distributed random effects. The gradient function within the supportive region of [-7.1, 3.1] deviates substantially from 1 with the confidence bands not including 1 in two regions. The gradient function suggests that the random intercept logistic model for the monotone missing data can be improved by moving probability from the region [-4, -2] to the region [-2, 0]. The inadequacy of assuming normally distributed



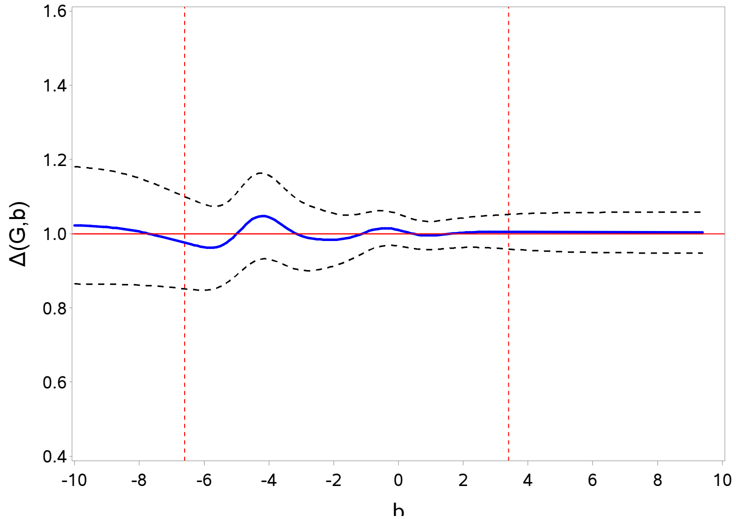
(a) Complete Cases - Assume Normal



(b) Complete Cases - Assume Mixture



(c) Monotone Missing - Assume Normal



(d) Monotone Missing - Assume Mixture

Figure 4.2: Gradient function (solid blue line) and 95% pointwise confidence interval bands (dashed black lines) for fitting a random intercept logistic model to (a) Complete cases assuming a normal distribution for the random intercept, (b) Complete cases assuming a three component mixture of normal distribution for the random intercept, (c) Monotone missing data assuming a normal distribution for the random intercept, and (d) Monotone missing assuming a three component mixture of normal distribution for the random intercept. Red solid horizontal line at gradient function=1 represents the optimal fit. Dashed red vertical lines represent the intervals $[b^*_{min}, b^*_{max}]$ where the data provides information about the support for the random effects distribution.

random intercepts is confirmed by the asymptotic diagnostic test, producing a test statistic of $T = 0.003$ corresponding to a p-value of 0.030 ($U = 1041759$). Fitting the random intercept logistic model to the monotone missing data assuming random intercepts distributed as a three component mixture of normals appeared to improve the model fit (Figure 4.2(d)). The gradient function shows some fluctuation around 1, yet the confidence bands contain 1 within the supportive range of [-6.6, 3.4]. The asymptotic diagnostic test confirms the adequacy of the three component mixture of normals as the distribution of the random intercepts, with a test statistic of $T = 0.0002$ corresponding to a p-value of 0.99 ($U = 430026$). For the monotone missing data, the gradient function exploratory diagnostic tool suggests that is no evidence that further refinement of the random effect distribution can achieve substantial improvement. This is confirmed by the insignificant asymptotic diagnostic test, advocating that the model assuming a three component mixture of normals for the random intercept distribution provides an adequate fit to the data.

4.5 Discussion

The lower residual deviance, optimal gradient function and non-significant asymptotic diagnostic test associated with the models assuming random intercepts as three component mixture of normals, suggest that multimodality of the random intercept is plausible. Assuming a three component mixture distribution also resulted in smaller standard errors of the predicted random intercepts than assuming normality, providing further support for multimodality (Appendix C). Regardless of the missing data scenario, the estimated random intercept distributions as a three component mixture of normal distributions were similar. As the random intercept gives an indication of the underlying propensity for women to be in employment, assuming a three component mixture distribution may represent the underlying mover-stayer scenario. For example, the component corresponding to a large negative random intercept represents those women with very low propensity to be employed; the component with almost mean zero represents women who have a propensity to transition between employment states; and the component corresponding to a large positive random intercept represents women with very high propensity to be employed. This application highlights one example in a panel survey setting where the Gaussian random effects assumption may not be the most appropriate in practice.

There were differences between the observed proportions of movers and stayers in the HILDA dataset and the mixing proportions estimated by assuming the random intercepts were a three component mixture of normals. These differences may be due to the latent mover-stayer scenario, whereby the observed stayers may comprise of individuals who are latent stayers and also comprise of individuals who are latent movers that have not transitioned during the observational period (Lindsey, 1997). Furthermore, the observed and fitted proportions differ as they are, respectively, unconditional and conditional on the explanatory variables included in the random intercept logistic model. Additionally, increased variability in the random effects distribution can be introduced if variability associated with unobserved explanatory variables omitted from the model have subsequently been incorporated into the random effects structure (Verbeke and Molenberghs, 2013). However, quantification of latent stayers was not the primary focus of this analysis. This application demonstrates that the assumption of a three component mixture of normals for the random intercept distribution may adequately capture the heterogeneity of a potential underlying mover-stayer scenario.

The random intercept logistic model considered in this application is too simple to address questions about employment transitions in Australian working aged women. More appropriate analyses would consider more than two employment states by distinguishing between part-time and full-time employment, and also distinguishing between unemployed and not in the labour force. Furthermore, additional explanatory variables could be considered, such as a lagged employment status term to account for state dependence. We have also only focused on two-level GLMMs, however in panel survey settings, three-level models may be necessary to take into account higher order clustering at the household level. The models considered in this chapter have focused on the commonly used logit link, though the models could also be formulated in terms of the probit or the complementary log-log link. However, marginal differences are expected for the alternative link functions, as Neuhaus et al. (2013) reported negligible differences for the impact of misspecification of logistic mixed models with either the logit or the complementary log-log link. Future work should consider more complex models such as additionally including individual-specific slopes, or multi-process models whereby random effects are shared between multiple processes. However, this application serves as an example to demonstrate the underlying mover-stayer scenario and potential multimodal random effects in simple logistic mixed models applied to panel data.

The likelihood reformulation method (Liu and Yu, 2008) used to estimate parameters in the logistic mixed model assuming mixture of normal distributed random intercepts was sensitive to the starting values and number of adaptive quadrature points used in the estimation. As a sensitivity analysis the likelihood reformulation method applied to the motivating example was re-fitted with the number of quadrature points varying from 10 to 80 for both missing data scenarios (Appendix B). The estimates and standard errors for the complete case data appeared to stabilise when the adaptive quadrature points exceeded 54, therefore the results in Table 4.3 for the complete cases are based on 54 adaptive quadrature points. Conversely, for the monotone missing data, the estimates and the standard errors were more variable. The estimates appeared to stabilise when the quadrature points exceeded 51, though the standard errors still exhibited variability. The results in Table 4.3 for monotone missing data was based on 61 adaptive quadrature points as the standard errors for all parameters were consistently small and parameter estimates were similar to neighbouring models (i.e. quadrature points 57, 59, 64, 69). The practical use of the likelihood reformulation method in the literature is not well documented. As such, there is limited information regarding the selection of starting values for the random effects distribution, number of adaptive quadrature points, model choice selection or restrictions necessary to obtain optimal model fitting. The models in this chapter used estimated fixed effect coefficients from a logistic model assuming normality as starting values for the fixed effect parameters. The starting values for the parameters corresponding to the random intercept distribution were based on the estimated three component mixture of normals fitted to the predicted random intercepts from the logistic model assuming normality. The analyses used the same restrictions implemented by Verbeke and Molenberghs (2013) to model random intercepts as a three component mixture of normals. Further work investigating the practical use of the likelihood reformulation method is required, including suggestions or guidelines for selection of starting values, number of quadrature points and model selection.

The gradient function diagnostic tool has been used as an exploratory tool to identify potential misspecification of the random effects distribution. The limitation of the graphical tool is that the confidence bands used to identify potential random effect misspecification are based on pointwise estimates, and are simply an informal tool to quantify the strength of departure from a gradient function of one (Verbeke and Molenberghs, 2013). Further, the graphical tool only provides information within a supportive region based on individuals with non-constant response profiles, and hence potentially only provides information about random effect misspecification for the latent movers. The asymptotic diagnostic test of Drikvandi et al. (2016) addresses these issues. The formal diagnostic tool appropriately tests for misspecification along the whole real line, by testing whether the departure of the gradient function from one is due to distributional misspecification of the random effects and not random variability. Thus, the diagnostic test can provide evidence of misspecification for all individuals within the mover-stayer context. However, as both of the diagnostic tools are based on the idea of the gradient function, they explicitly assume that the conditional distribution of the GLMM is correctly specified. This assumption may be restrictive, and as such, development of approaches to relax this assumption is an area of ongoing research (Drikvandi et al., 2016).

Comparing the adequacy of the model fits for models with alternative random effect distributions is not-straight forward, as standard asymptotic theory does not apply (Litiére et al., 2008). Furthermore, model comparisons of different numbers of components in the heterogeneity model are complicated by boundary problems, such that formal comparison of the log-likelihood using a likelihood ratio test will not follow standard rules (Molenberghs and Verbeke, 2005). Therefore, the residual deviance, calculated as the negative of twice the log-likelihood, has been used as a measure of the model fit and for model comparisons. Although comparison based on the estimated log-likelihood can be limited by reflecting the quality of the technique to obtain an approximation of the model likelihood (Molenberghs and Verbeke, 2005), model comparisons based on information criteria are not easily derived for generalised linear mixed models (Steele, 2013) and comparisons are not straight forward (Molenberghs and Verbeke, 2005). Therefore, as utilised by Molenberghs and Verbeke (2005), McCulloch and Neuhaus (2011a) and Neuhaus et al. (2013) among others, the residual deviance has been used as an indicative measure of the model fit.

This case study highlights that potential random effect distributional misspecification and attrition can often occur simultaneously in panel survey settings. Maximum likelihood estimation is known to be valid under the assumption of MAR, however it is unknown if estimates are robust to missingness when the random effects distribution is misspecified. This naturally leads to the following questions. In such a mover-stayer scenario, what additional impact does attrition have on violations of the assumed random effects distribution in panel survey settings? Furthermore, how extreme does the random effect distribution have to be from normality before violating the normality assumption impacts inference? To address these questions, Chapters 5 and 6 consider simulation studies based on this HILDA case study to assess the impact of attrition and misspecified random intercept distribution within potential mover-stayer scenarios. Using the results from the case study considered in this chapter, Chapter 5 considers the specific departure from normality arising from an asymmetrical three component mixture of Gaussians. The asymmetric mixture distributions are simulated to represent the potential mover-stayer scenario observed in this chapter, and considers a range of random intercept variances motivated by the HILDA case study. Following on from the initial simulation study, Chapter 6 assesses the robustness of inferences in random intercept logistic models to violations of the normality assumption by considering a range of true symmetric three component mixture of normal distributions. The simulated random intercept distributions vary in severity of departures from normality, identifying scenarios whereby random effect misspecification can impact maximum likelihood inference.

This case study also highlights that flexibly modelling the random effect distribution can help guard against the impact of distributional misspecification. By modelling the random intercepts as a three component mixture of normal distributions, increasing the flexibility of the assumed random effects provided a better model fit than the conventional normal distribution. To further explore the benefits of flexibly modelling the random effect distribution within a potential mover-stayer scenario and to determine an optimal modelling strategy, Chapter 7 proposes the use of the Vertex Exchange Method (VEM) to non-parametrically estimate the random effects distribution in logistic mixed models. Within a sensitivity analysis framework, Chapter 7 compares the performance and practicality of implementing VEM and a selection of existing flexible random effect methods implementable in standard software. In practice, the application of GLMMs is often not restricted to random intercept models. Therefore, Chapter 7 considers the practicality of implementing flexible random effect distributions in GLMMs with univariate and bivariate random effects (i.e. random intercepts and random slopes) when applied to the HILDA case study.

5 Investigating the impact of incorrectly assuming normality in an underlying mover-stayer scenario: Applications to panel surveys

5.1 Introduction

In applications of random intercept logistic models it is standard practice to assume the random effects are normally distributed[1]. However, as demonstrated in the HILDA case study in Chapter 4, the assumed normal distribution may not be adequate to capture underlying heterogeneity of the random effects in a mover-stayer scenario. In this scenario, inducing a more flexible distribution for the random effects may be required to accommodate the latent stayers in the social process under investigation (see Section 2.5 for more details). In a random intercept logistic model, it may be more appropriate to supplement the assumed normal distribution with endpoints at positive and negative infinity as considered by Davies et al. (1992), or by modelling the random intercepts as a three component mixture of normal distributions as demonstrated in the HILDA case study (Chapter 4). Albeit the development and availability of methods to flexibly model the random effects distribution, the normality assumption is typically taken for granted in practice (Verbeke and Molenberghs, 2013).

Incorrect assumptions of the underlying random effects distribution can impact the parameter estimates and standard errors of the model parameters in a GLMM, and thus, result in incorrect interpretation and inference. Theoretical results for the random intercept logistic model show that misspecifying the random effects distribution can produce biased estimates of parameters directly related to the random intercept (Neuhaus et al., 1992), such as the intercept constant and the random intercept variance estimate. Similar theoretical results were reported by Neuhaus et al. (2013) for the more general class of GLMMs with random intercepts and random slopes, with bias typically restricted to parameters directly related to the misspecified random effects. However, estimation of the joint density in GLMMs is typically not of closed form expression[2], thus limiting the derivation of theoretical properties to the restricted scenario when explanatory variables are unrelated to the response.

[1] As highlighted by McCulloch and Neuhaus (2011b), most statistical software packages only allow the assumed distribution for the random effects to be normal.

[2] See Section 3.1 for cases of GLMMs with closed-form likelihoods.

To establish results in more general settings, simulation studies have been utilised to explore the robustness of inferences in GLMMs to misspecification of the random effects distribution (e.g. Heagerty and Kurland 2001; Litiére et al. 2011; Neuhaus et al. 2013). Studies investigating the impact of misspecification on inference for the model parameters in random intercept logistic models generally report biased estimates of the intercept constant and the variance component (Neuhaus et al., 1992; Heagerty and Kurland, 2001; Agresti et al., 2004; Litiére et al., 2008; McCulloch and Neuhaus, 2011a). However, there is no general consensus about the impact on estimating the fixed effects parameters, particularly parameters capturing the effects of time-invariant explanatory variables. McCulloch and Neuhaus (2011a) suggest that most aspects of statistical inference are generally robust to distributional violations from the normality assumption, including the time-varying explanatory variables. Negligible bias in the point estimators of the time-invariant fixed effects parameters has been reported (Neuhaus et al., 1992; McCulloch and Neuhaus, 2011a), with some bias and loss of efficiency reported for true distributions far from normality. However, another body of research has suggested sensitivity to the assumed random intercept distribution, reporting severely biased estimates of time-invariant parameters (Heagerty and Kurland, 2001; Agresti et al., 2004; Litiére et al., 2008). Particularly for true skewed random effects distributions that are incorrectly assumed to be normally distributed (Litiére et al., 2008), and for true random effects that differ from the shape of the assumed normal distribution with large variability of the random effects (Heagerty and Kurland, 2001).

The ambiguity about the impact of misspecifying the random effects distribution has been further exacerbated by the lack of investigation of issues common in panel survey settings. For instance, previous literature has not considered the impact of misspecification on estimating the effects of categorical explanatory variables. Further research is required as data in panel surveys are typically collected from self-reported questionnaires, and thus, analysis of categorical variables is prominent. Furthermore, limited research has considered the impact of misspecifying the random effects distribution on estimating model parameters in the presence of missing data due to attrition. Hartford and Davidian (2000) showed that intermittent missingness is more susceptible to the impact of misspecified random effects in non-linear mixed effects models. In the context of GLMMs, only one study has shown that intermittent missingness and attrition can affect the power to detect variance components when the true random effects distribution is positively skewed, or positively skewed and leptokurtic yet assumed to be normal (Wang, 2010b). Under the assumption of MCAR or MAR missingness, maximum likelihood estimation of GLMMs can provide consistent estimation. However, this implicitly assumes that the other aspects of the model are correctly specified, including distributional assumptions for the random effects. As demonstrated in the HILDA case study in Chapter 4, this is an area requiring further research as misspecification of the random effects distribution and missing data can simultaneously occur in practice.

Simulation studies investigating the impact of distributional misspecification on the inference of parameter estimates in GLMMs can be assessed by simulating a variety of true distributions for the random effects and examining the performance under the assumption of normality (e.g. Litiére et al. 2008; Neuhaus et al. 2013). As highlighted in Table 2.1, no research has investigated the impact of assuming normality when the true random intercepts are multimodal with three modes (trimodal). This is an area requiring research, as the random intercepts distribution may be better represented as an asymmetric multimodal distribution if an underlying sub-population structure exists (as demonstrated in Chapter 4). However, limited studies have considered simulating true asymmetric multimodal distributions. Litiére et al. (2008) simulated random intercepts as an asymmetric two component mixture distribution, reporting inconsistent and biased estimates when incorrectly assuming normality for the random intercepts[3]. In addition to the biased estimation produced for the intercept constant and the variance component estimate, biased estimation of the time-invariant explanatory variable was reported with relative bias ranging between -15 to -9% for the larger true random effect variances ($\sigma_b^2 = 16, 32$). Furthermore, literature investigating the mover-stayer scenario has not considered simulation studies to examine the robustness of inferences in random intercept logistic models to the normality assumption of the random effects in potential mover-stayer scenarios. Typically researchers attempting to accommodate and quantify the mover-stayer scenario in a social process of interest will compare the performance of fitting a random intercept logistic model assuming a more flexible random effects distribution with the fit of the model assuming normality (e.g. Davies et al. 1992). To my knowledge, no research has aimed to quantify the impact of assuming normality when the random intercepts are multimodal with three modes due to an underlying mover-stayer scenario.

These gaps in the literature naturally lead to the following questions. How robust is the random intercept logistic model to violations of the normality assumption characterised by trimodal distributions due to an underlying mover-stayer scenario? Furthermore, does missing data due to attrition have an additional impact on violations of the assumed random effects distribution in panel survey applications?

To address these questions, the research presented in this chapter utilised a simulation study to evaluate the robustness of inferences in random intercept logistic models applied to panel survey settings, focusing on misspecifying the random effects distribution and the presence of missing data. The simulation study considers the specific departure from normality characterised by an asymmetric three-component Gaussian mixture model, motivated by the HILDA case study presented in Chapter 4 to represent the mover-stayer scenario. The simulation study investigates the impact of incorrectly assuming normally distributed random intercepts when the true distribution is multimodal due to a potential underlying mover-stayer scenario, for random intercept logistic models applied to complete data and missing data due to attrition.

[3]Inconsistency and bias were also reported for other skewed true random effects distributions, including the exponential, log-normal and power function.

5.2 Simulation study design

In this study data is simulated from a random intercept logistic model with a range of probability distributions specified for the random intercepts. The random intercepts are generated from six asymmetric mixture distributions with random intercept variance motivated by the HILDA case study, and two missing data scenarios. The process used to simulate data from the random intercept logistic model and to simulate the missingness due to drop-out is detailed in Section 5.2.1 and 5.2.2, respectively.

5.2.1 Data generating model

The parameters and design matrix for this simulation study are derived from the analysis of the women with complete employment histories from 2001 in the HILDA case study (Chapter 4). To maintain the correlation structure of the explanatory variables in the HILDA survey at baseline and over time, resampling techniques were utilised to generate the explanatory variables. For each iteration of the simulation, 1000 women were randomly selected without replacement from the 1359 women with complete employment history data (complete cases). The explanatory variables of the 1000 randomly resampled women were used to generate clustered binary responses to represent employment status using the random intercept logistic model in Equation 4.1. Further details of the data generating model used in this simulation study are outlined in Section 3.4.1. The fixed effect parameter values for the simulations were selected to be similar to the estimates for the complete case in Table 4.3: $\beta_0 = 1.2, \beta_1 = 0.1, \beta_2 = -0.3, \beta_3 = -0.1, \beta_4 = -1.5, \beta_5 = -2.8, \beta_6 = -2.3$ and $\beta_7 = -0.4$.

The random intercept (b_i) was simulated from an asymmetric three component mixture of normal distributions to represent the mover-stayer scenario,

$$b_i \sim \pi_1 N(\mu_1, \sigma_1^2) + \pi_2 N(\mu_2, \sigma_2^2) + \pi_3 N(\mu_3, \sigma_3^2) \tag{5.1}$$

where π_k, μ_k and σ_k^2 are the mixing proportions, component means and component variances for the $k = 1, 2, 3$ components, respectively. The parameter values for the mixture distribution were inspired by the component estimates assuming the random effects were distributed as a three component normal mixture for the complete case analysis sub-group in the HILDA case study (Table 4.3). To ensure the expected value of the simulated random intercept was zero, the values for the mixing proportions were fixed at $\pi_1 = 0.12, \pi_2 = 0.55$ and $\pi_3 = 0.33$, and the component means were fixed at $\mu_1 = -5.5, \mu_2 = -0.75$ and $\mu_3 = 3.25$. Six different combinations for the component variances $(\sigma_1^2, \sigma_2^2, \sigma_3^2)$ were considered, such that the overall random effect variability was similar to the range of estimated variances in the motivating example (range ± 1). The true total random effect variances considered were $\sigma_b^2 = 8, 9, 10, 11, 12, 13$. The values of component variances were selected in an iterative process such that the ratio between component standard deviations were similar to the estimated mixture distribution for the complete case analysis

Table 5.1: Component standard deviations ($\sigma_1, \sigma_2, \sigma_3$) used to generate the random intercepts in the simulation study for true random effect variances ranging from 8 to 13.

True Variance	σ_1	σ_2	σ_3
$\sigma_b^2 = 8$	0.675	0.9	0.495
$\sigma_b^2 = 9$	1.11	1.48	0.814
$\sigma_b^2 = 10$	1.425	1.9	1.045
$\sigma_b^2 = 11$	1.6725	2.23	1.2265
$\sigma_b^2 = 12$	1.8938	2.525	1.3888
$\sigma_b^2 = 13$	2.0925	2.79	1.5345

(Table 4.3), and were fixed at $\sigma_1 = 0.75 \times \sigma_2$ and $\sigma_3 = 0.55 \times \sigma_2$. For the six true random variance scenarios considered in the simulation, the selected component standard deviations are shown in Table 5.1 and the density distributions of the simulated random intercepts are shown in Figure 5.1.

Simulations were performed under two missing data scenarios, complete data (the full dataset with no missing data) and incomplete data due to attrition. Attrition was imposed using a missing at random (MAR) mechanism, where the probability of drop-out was a Markov process depending on employment status of the previous wave ($y_{i,j-1}$) (Bonate, 2011). The MAR drop-out was simulated using a probabilistic approach (Bonate, 2011) as detailed in Section 5.2.2. The simulated wave-to-wave attrition rates were similar to the 29.5% rate observed in the HILDA case study (Table 4.1).

For each of the six random effects and two missing data scenarios, 1000 datasets were generated. A random intercept logistic model assuming Gaussian random effects was fitted to each simulated dataset. To assess the sensitivity of the normality assumption on estimating model parameters under misspecification of the random effects distribution, the performance measures of percentage bias, coverage of the 95% confidence intervals and ratio of the mean standard error to the empirical standard error were used as described in Section 3.5. Criteria for acceptable performance were percentage bias within −10% and 10% (Marshall et al., 2010), coverage rates within 93.6% and 96.4% (Burton et al., 2006) and standard error ratios within 0.9 and 1.1 (Neuhaus et al., 2013). For the performance measures relating to the random intercept distribution, the variance estimate of the random intercept was compared to the overall variance of the three component mixture distribution ($\hat{\sigma}_b^2 = \sum_{k=1}^{3} \hat{\pi}_k(\hat{\sigma}_k^2 + \hat{\mu}_k^2) - (\sum_{k=1}^{3} \hat{\pi}_k \hat{\mu}_k)^2$, for $k = 1, 2, 3$.)

Simulations and analyses were conducted in SAS (Version 9.4, SAS Institute, Cary NC). All random intercept logistic models were fitted using the SAS procedure NLMIXED with adaptive Gaussian Quadrature using 20 quadrature points.

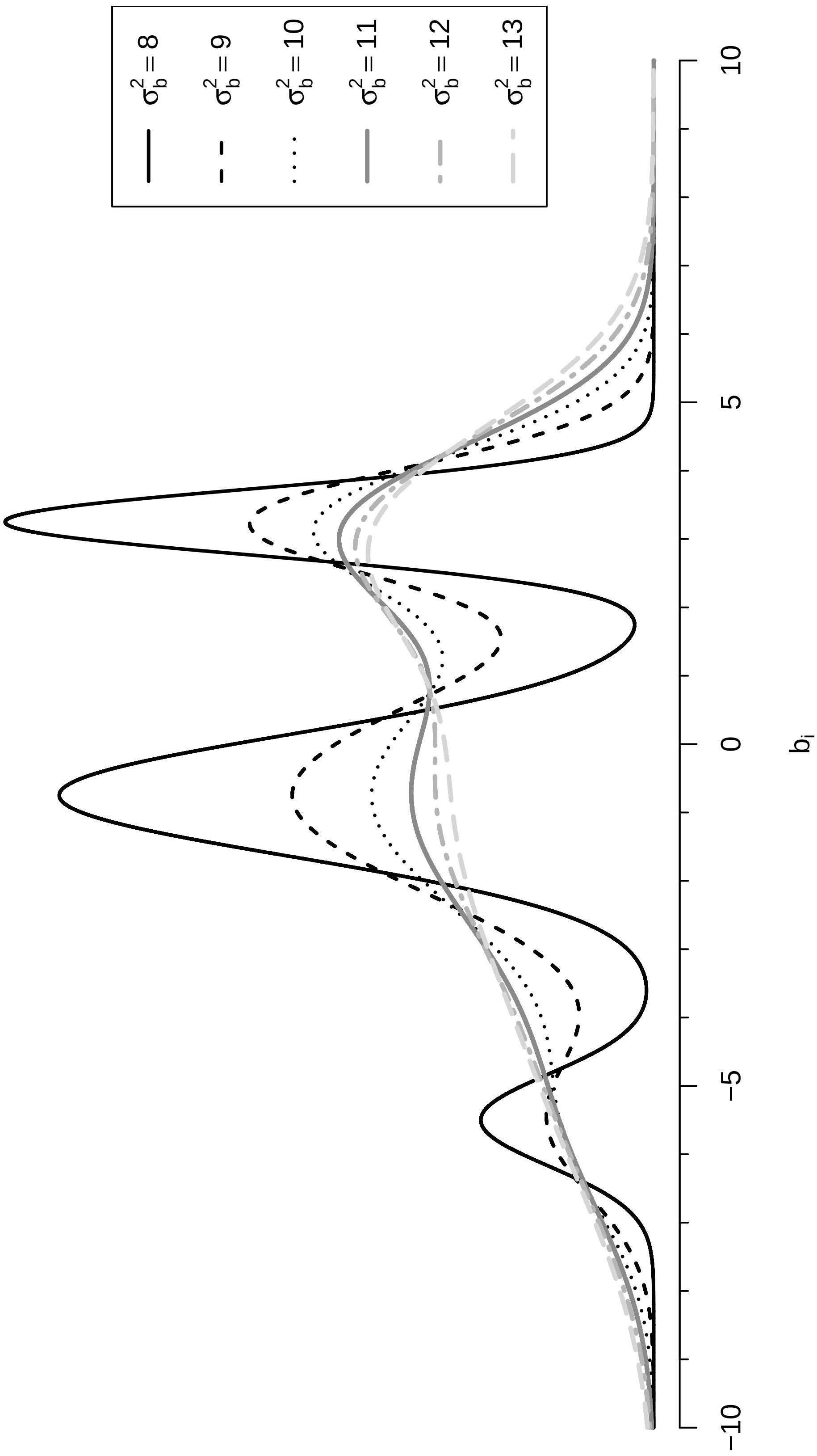


Figure 5.1: Density of the six true random intercept distributions considered in the simulation study: $\sigma_b^2 = 8, 9, 10, 11, 12$ and 13.

5.2.2 Drop-out generating model

For each of the 1000 simulated complete datasets generated for the six random effect variance scenarios ($\sigma_b^2 = 8, 9, 10, 11, 12, 13$), the following drop-out model was used to generate missingness to represent attrition. The missingness was generated according to a MAR missingness mechanism, in which the missingness is dependent on explanatory variables and the response at the previous wave ($y_{i,j-1}$). In order to simulate MAR drop-out similar to the attrition rates observed in the HILDA case study, the coefficients used in the drop-out generating model was derived by the HILDA case study. As described in Section 3.4.2 a logistic model, termed the drop-out model, can be used to model the conditional probability of drop-out,

$$p_{ij}(\alpha) = \Pr(R_{ij} = 0 | R_{i(j-1)} = 1, y_{i(j-1)}, \boldsymbol{x_{ij}}; \boldsymbol{\alpha}) \tag{5.2}$$

where R_{ij} is an indicator variable for whether individual i is observed at time j.

The drop-out model captured wave-specific effects by including linear and quadratic wave terms (w_{1j} and w_{1j}^2), and included the following explanatory variables: employment status at the previous wave $y_{i,j-1}$, age at first wave w_{2i}, highest level of education attained at first wave (w_{3i}, w_{4i}, w_{5i}), and dependent children at the previous wave ($w_{6i,j-1}$ and $w_{7i,j-1}$). The coefficients of the drop-out model to simulate MAR attrition were based on fitting a logistic regression model (Equation 5.3) to the 1927 women with monotone missing data in the HILDA case study (results presented in Appendix D). The coefficients used in the drop-out generating model were: $\alpha_0 = 0.28$, $\alpha_1 = -0.69$, $\alpha_2 = 0.04$, $\alpha_3 = -0.03$, $\alpha_4 = 0.33$, $\alpha_5 = 0.45$, $\alpha_6 = 0.55$, $\alpha_7 = -0.59$, $\alpha_8 = -0.15$ and $\alpha_9 = -0.33$.

$$\begin{aligned}\text{logit}(p_{ij}(\alpha)) = \alpha_0 + \alpha_1 w_{1j} + \alpha_2 w_{1j}^2 + \alpha_3 w_{2i} + \alpha_4 w_{3i} + \alpha_5 w_{4i} + \\ \alpha_6 w_{5i} + \alpha_7 w_{6i,j-1} + \alpha_8 w_{7i,j-1} + \alpha_9 y_{i,j-1}\end{aligned} \tag{5.3}$$

The conditional probability of individual i ($i = 1, ..., 1000$) missing at wave j ($j = 2, ..., 11$) was estimated based on Equation 5.3. As detailed in Section 3.4.2, for each individual at each wave, if the random draw ($u_{ij} \sim U[0, 1]$) was less than the conditional probability of drop-out (i.e. $u_{ij} < p_{ij}(\alpha)$) the individual was dropped for that wave and subsequent waves, otherwise the individual remained in the study. Further details are given in Appendix D.

5.3 Simulation study results

Figure 5.2 presents the percentage bias of the parameter estimates from the random intercept logistic model simulations across the true random effect variance (σ_b^2) for the two missing data scenarios. With increasing true random effect variance, Figure 5.2 shows that misspecification produced larger biased estimators of the intercept constant (β_0) and the random intercept standard deviation (σ_b). Assuming normality produced biased estimates of the parameter, β_3, capturing the effect of women never married or single, with larger true random effect variances

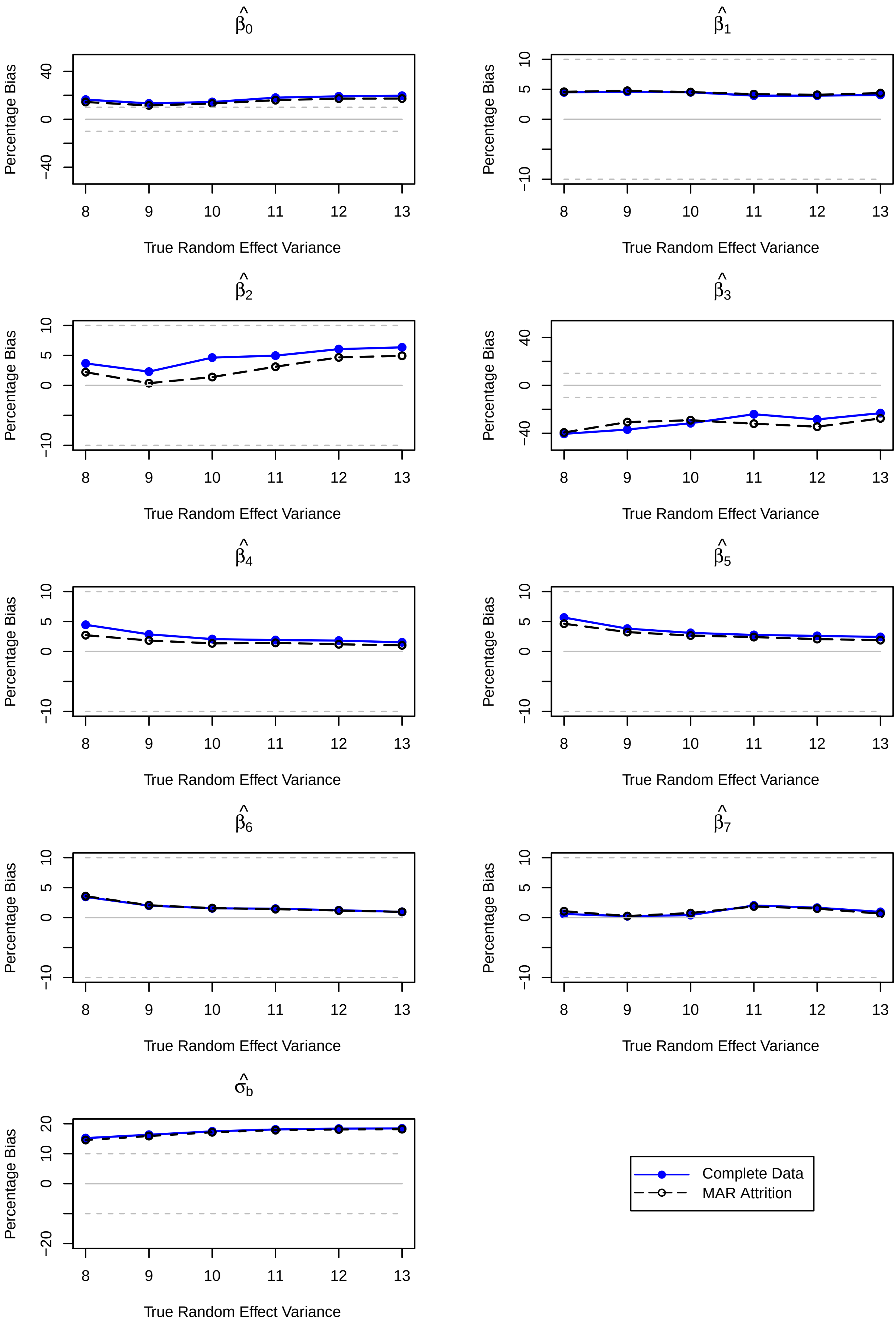


Figure 5.2: Percentage bias for parameter coefficients of random intercept logistic model for increasing true random effect variance (σ_b^2) under two data scenarios, complete data and MAR attrition. Grey horizontal solid line at percentage bias=0 and grey horizontal dashed lines at percentage bias -10% and 10%.

Table 5.2: Average rate (Average Attrition) and range (Range) of attrition in 1000 simulated datasets for each of the six true random effect scenarios for variances ranging from 8 to 13.

True Variance	Average Attrition (%)	Range (min% - max%)
$\sigma_b^2 = 8$	32.0	(27.1 - 36.9)
$\sigma_b^2 = 9$	32.1	(27.2 - 37.0)
$\sigma_b^2 = 10$	32.1	(27.1 - 37.2)
$\sigma_b^2 = 11$	32.2	(27.1 - 37.1)
$\sigma_b^2 = 12$	32.3	(27.3 - 36.9)
$\sigma_b^2 = 13$	32.3	(27.5 - 36.8)

resulting in less biased estimates, yet still outside the allowable ±10% threshold. Misspecification produced little and relatively consistent bias in the effects of the remaining explanatory variables, $\beta_1, \beta_2, \beta_4, \beta_5, \beta_6$ and β_7, with percentage bias within -10% and 10%. Misspecification in the presence of MAR attrition produced similar magnitudes and trends in percentage bias when compared to the complete data, with marginal differences in the trend for β_2 and β_4.

Figure 5.3 presents the coverage rates of 95% confidence intervals for the logistic model simulations across the true random effect variance for the two missing data scenarios. The coverage rates were typically close to the nominal rate of 95% for all parameter estimates with the exception of the estimate for the random effect standard deviation. Extremely poor coverage rates were experienced for σ_b, with nominal coverage rates less than 30% for all true random effect variances and both missing data scenarios. Misspecification of the random effects in the presence of attrition resulted in similar coverage rates and trends to the complete data scenario, with generally less extreme deviations from the nominal rate than for the complete data. There were larger differences in the coverage rate for σ_b between the missing data scenarios, with the differences decreasing as the true random effect variance increased.

Figure 5.4 presents the ratio of the mean of model-based standard errors to the empirical standard error of the parameter estimates across the true random effect variance for the two missing data scenarios. With the exception of the random effect standard deviation estimate, the ratio was within 0.9 to 1.1 for all model parameters. This indicates that even in the presence of random effect misspecification, the model based standard errors accurately describe the variability of the fixed effect coefficients. The standard error ratio for σ_b exceeded 1.1 for true random effect variances of 12 and 13. Misspecification of the random effects in the presence of attrition resulted in similar magnitudes and trends as for the complete data scenario. In comparison to the complete data, the MAR attrition had smaller standard error ratios for σ_b.

The average attrition rate and range of the 1000 simulated datasets for each of the random effect scenarios are shown in Table 5.2. The actual rate of attrition in the simulated datasets for all scenarios averaged 32.1% (ranging from 27% to 37.2%) and was similar to observed rate of 29.5% in the HILDA subgroup of working aged women (Table 4.1).

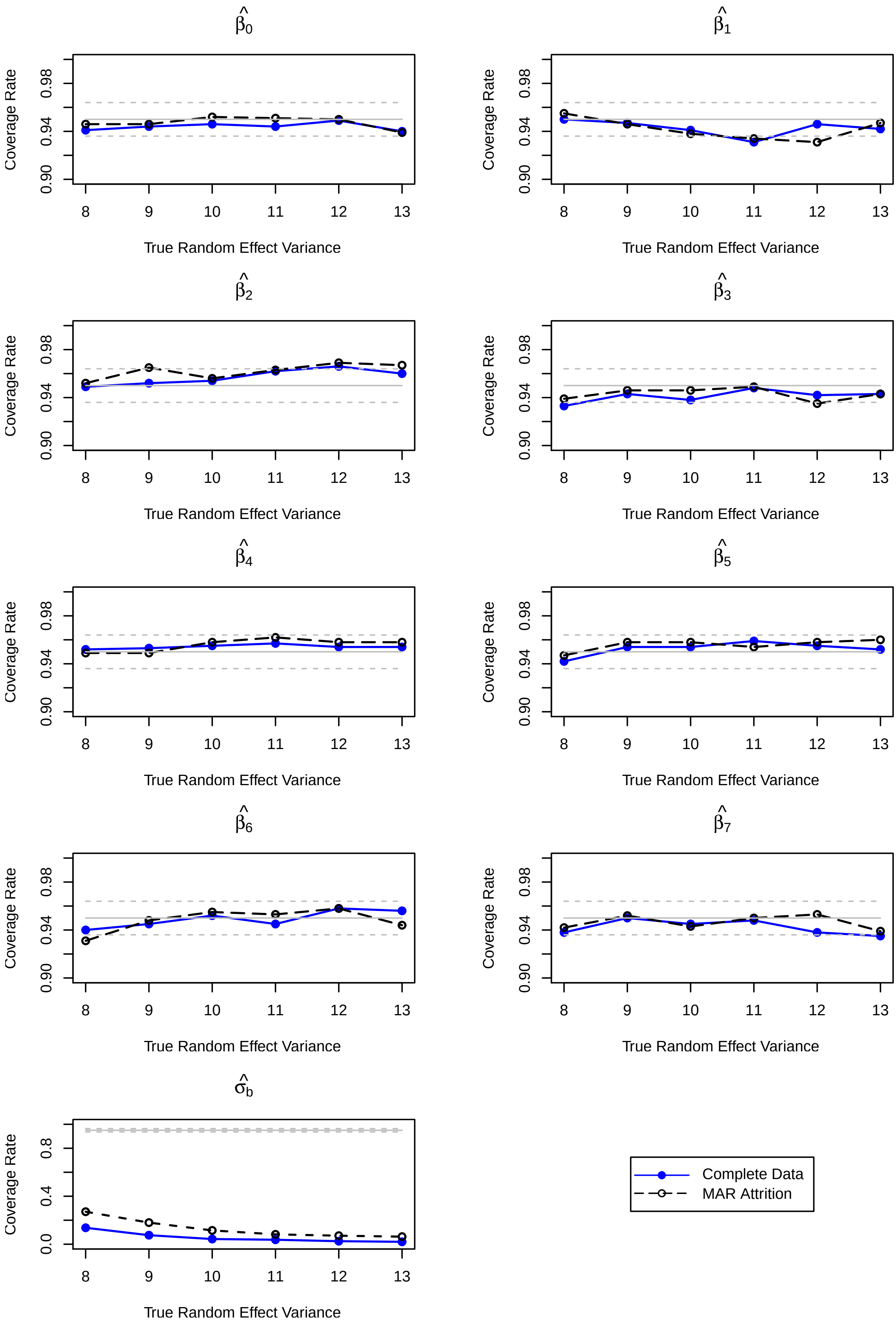


Figure 5.3: Coverage rates for model based 95% confidence intervals for parameter coefficients of random intercept logistic model for increasing true random effect variance (σ_b^2) under two data scenarios, complete data and MAR attrition. Grey horizontal solid line at nominal coverage rate 0.95 and grey horizontal dashed lines at coverage rate 0.936 and 0.964.

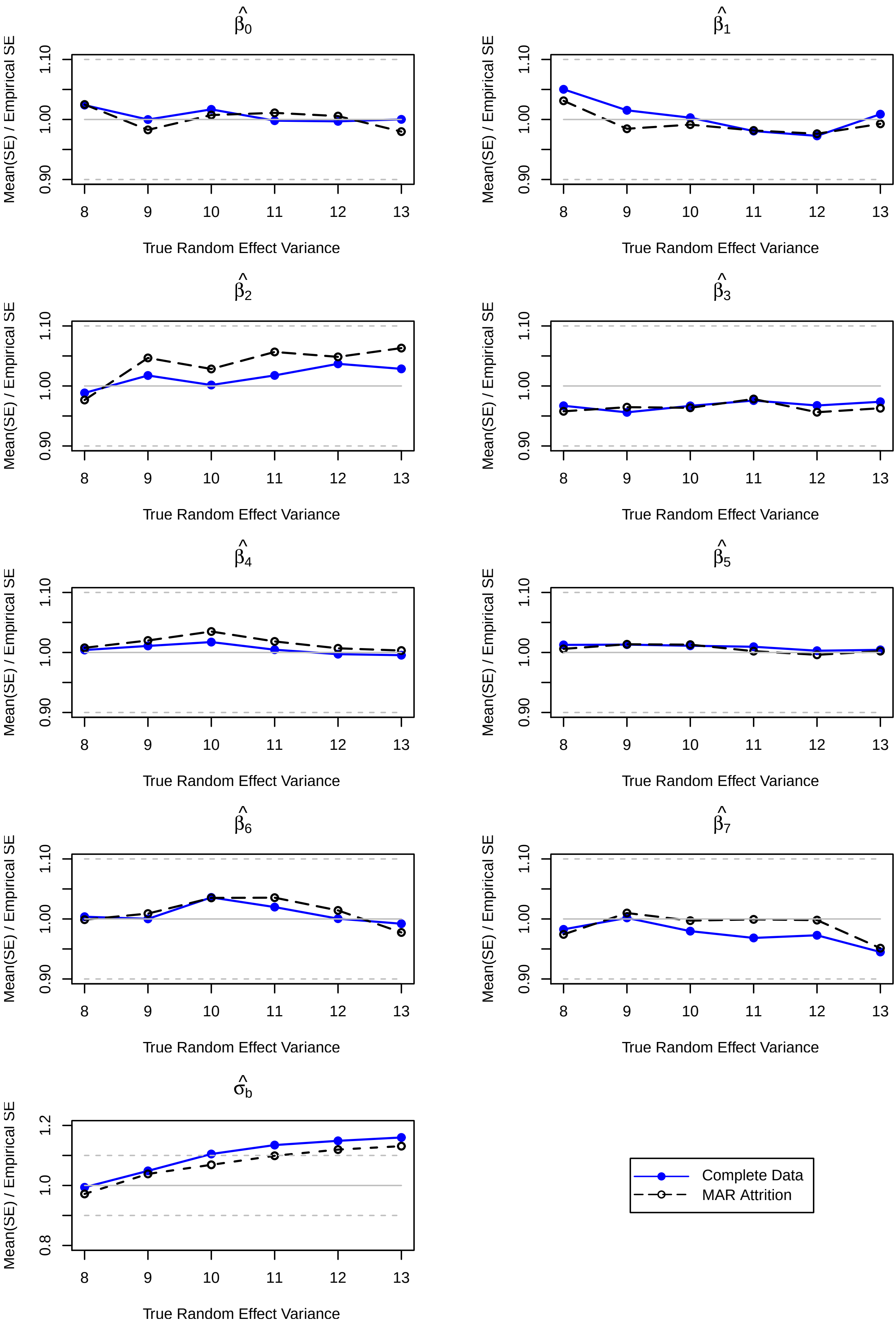


Figure 5.4: Ratio of mean model-based standard error to the empirical standard error for increasing true random effect variance (σ_b^2) under two data scenarios, complete data and MAR attrition. Grey horizontal solid line at ratio=1 and grey horizontal dashed lines at ratio of 0.9 and 1.1.

Table 5.3 presents the average percentage (range) of women within the mover-stayer scenario in the 1000 simulated datasets for the six random intercept scenarios and two data scenarios. The percentage of stayers in the unemployed group was influenced by the true random effect distribution, with a higher proportion continuously unemployed with increasing true random effect variance (Table 5.3). The rate of individuals staying consistently unemployed for the complete data and in the presence of MAR attrition averaged 5.5% (ranging from 2.2% to 10%) and 10% (ranging from 5.8% to 15.4%), respectively. Similarly, the rate of individuals staying consistently employed averaged 53.9% (ranging from 48.8% to 58.7%) and 57.1% (ranging from 52.2% to 62.4%) for the complete data and MAR attrition scenarios respectively (Table 5.3).

Fitting the assumed normal distribution using NLMIXED had excellent convergence rates of 100% for all scenarios.

5.4 Discussion

In practice the random effects of generalised linear mixed models are commonly assumed to be Gaussian distributed. However random effects may be multimodal when categorical fixed effects are omitted or, as shown in Chapter 4, dominated by the latent mover-stayer scenario. By considering the additional impact of attrition within a panel survey application, this simulation study provides a novel insight into the impact of misspecified random intercept distributions on logistic mixed models. This simulation study provides evidence that misspecification due to multimodal random intercepts can impact inference of maximum likelihood estimation of random intercept logistic models within panel survey applications. When the true random intercept distribution is an asymmetric three component mixture of Gaussians, assuming normality generally has minimal impact on the estimation of the fixed effects. However, inferences for parameters associated with the misspecified random intercept distribution were impacted. Misspecification induced large bias estimates of the intercept constant and variance component, and resulted in poor coverage rates and inaccurate model standard errors for the variance component estimate. Misspecification in the presence of MAR attrition resulted in similar bias, coverage and model based standard errors as for the complete data scenario.

The impact of misspecification on estimation of the fixed effect parameters unrelated to the random effect was generally minimal. Consistent with previous literature (Neuhaus et al., 1992, 1994; Heagerty and Kurland, 2001), misspecification of the random intercept distribution generally had minimal impact on estimating time-varying explanatory variables. With the exception of the negative bias observed for β_3, negligible bias was observed for parameters capturing the effects of the time-varying explanatory variables ($\beta_1, \beta_2, \beta_6$ and β_7). Furthermore, misspecification resulted in nominal coverage rates and accurate model based standard errors for the parameters related to the time-varying explanatory variables. The parameter β_3 capturing the effect of being single and never married was consistently underestimated and negatively

Table 5.3: Average percentage and range (in parentheses) of women in the 1000 simulated datasets observed to always be unemployed, transiting between employment states (Mover) and always be employed for the two simulated data scenarios (Complete Data and MAR Attrition) and six random intercept scenarios (True Variance).

True Variance	**Complete Data**			**MAR Attrition**		
	Always Unemployed	Mover	Always Employed	Always Unemployed	Mover	Always Employed
$\sigma_b^2 = 8$	4.2 (2.2 - 6.3)	42.5 (37.6 - 46.3)	53.3 (59.9 - 57.8)	8.4 (5.8 - 11.5)	34.5 (30.7 - 39.0)	57.2 (52.5 - 61.5)
$\sigma_b^2 = 9$	4.6 (2.6 - 6.9)	41.8 (37.2 - 46.0)	53.6 (49.0 - 58.0)	9.0 (6.2 - 11.6)	33.9 (29.9 - 38.6)	57.1 (52.5 - 61.7)
$\sigma_b^2 = 10$	5.2 (3.1 - 7.6)	41.0 (36.6 - 45.1)	53.9 (49.5 - 58.4)	9.7 (6.7 - 12.7)	33.2 (28.3 - 38.1)	57.1 (52.4 - 61.9)
$\sigma_b^2 = 11$	5.8 (3.7 - 8.3)	40.2 (36.0 - 44.5)	54.0 (49.8 - 58.7)	10.4 (7.1 - 14.2)	32.5 (27.3 - 36.5)	57.1 (52.2 - 62.1)
$\sigma_b^2 = 12$	6.4 (4.3 - 9.3)	39.5 (34.4 - 43.6)	54.1 (49.7 - 58.6)	11.0 (7.7 - 14.8)	31.9 (27.3 - 36.5)	57.1 (52.4 - 61.9)
$\sigma_b^2 = 13$	7.1 (4.6 - 10.0)	38.8 (33.9 - 43.5)	54.2 (49.7 - 58.7)	11.7 (8.2 - 15.4)	31.3 (26.5 - 36.4)	57.0 (52.7 - 62.4)

biased. This bias may partly be explained by the large standard error, and hence variability, of the coefficient and the small number of individuals in this marital state. Furthermore, the bias may be partly explained by the relatively stable nature of this time-varying explanatory, exhibiting minimal within-individual variability over time. Estimation of the parameter effects of highest level of education (β_4 and β_5), appeared to be unaffected by the random intercept distributional misspecification. The minimal bias observed for β_4 and β_5 are consistent with previous results of time-invariant covariates for non-extreme departures from normality (Neuhaus et al., 1992; Heagerty and Kurland, 2001). However, biased estimators of time-invariant explanatory variables have previously been reported when the true random effect distribution is highly skewed and has substantial variability (Heagerty and Kurland, 2001; Litiére et al., 2008).

However, inferences about parameters associated with the misspecified random intercept were severely impacted. Misspecification of the random intercept distribution induced large bias in estimation of the intercept constant, with percentage bias exceeding 10% for both missing data scenarios and all six true random effect variances considered (ranging from 11.6% to 19.7%). These results are consistent with theoretical and simulation studies showing bias in the estimate of the intercept constant when the true random effect distribution differs substantially from assumed normality (Neuhaus et al., 1992; Heagerty and Kurland, 2001). Not only is care required when interpreting the intercept constant, biased estimation of the intercept constant may carry over to other aspects of the GLMM and possibly impact the mean estimation of the outcome (McCulloch and Neuhaus, 2011a).

Furthermore, the estimates of the variance component (σ_b) were always severely impacted. Misspecified random intercept distributions resulted in seriously biased estimates of the random effect standard deviation, substantially low corresponding coverage rates and inaccurate model based standard errors. Consistent with the findings of Litiére et al. (2008), larger true random intercept variances resulted in larger bias in the estimation of the variance component. However, unlike the underestimation[4] observed for the asymmetric mixture of two normals considered by Litiére et al. (2008), the asymmetric three component mixture of normals considered in this study resulted in positive bias in the estimation of the variance component. The impact of misspecified random effects on estimating the variance component has received little attention in the literature (McCulloch and Neuhaus, 2011a), and may not be considered of primary inferential interest. However, the variance components are the only available estimate of the true random effect variability (Litiére et al., 2008). Bias can subsequently impact interpretation of the intra-class correlation (ICC)[5], a measure indicating the proportion of unexplained

[4]Perhaps the observed underestimation may be due to the median value of the standard deviation of the random effects being reported in Table II of Litiére et al. (2008), rather than the variance estimate. The thesis reports substantial inconsistency of the parameter estimates for the asymmetric mixture distribution (relative bias for all parameters jointly), reporting that the variance component is the most affected (Litiére, 2007).

[5]As the ICC for the random intercept logistic model has a fixed residual error of $\pi^2/3$, large differences in the ICC may be restricted to scenarios when the magnitude of the random effect variance estimate is not large when compared to residual error

variation at the individual level, often used to express the random intercept variance (Hedeker and Gibbons, 2006). Furthermore, biased estimates of the variance component can make it difficult to evaluate problems in the mean structure (Litiére et al., 2008) and can additionally have an impact on individual-specific interpretations or predictions. McCulloch and Neuhaus (2011b) suggest that severe distributional violations of the assumed random effect distribution can impact the prediction and accuracy of best predicted random effect values. This is an area requiring further research, particularly when the true random effect distribution is multimodal with three or more modes.

In the presence of a misspecified random intercept distribution, MAR attrition has little additional impact on the estimation of parameters in the fixed and random components. In comparison to misspecification in the complete data scenario, there were minimal differences in the bias, coverage rates and accuracy of the model standard errors. These results are consistent with the simulation study by Wang (2010b), whereby ignorable missingness had little impact on bias or coverage when the true random effects distribution was misspecified. As GLMMs are expected to be robust to MAR and MCAR missingness, further research investigating the simultaneous impact of random effect misspecification and missingness will need to consider the MNAR mechanism in addition to other attrition rates.

This simulation study has considered multimodal random intercepts to represent the latent mover-stayer scenario, with total random intercept variances ranging from 8 to 13. The distributions considered in this simulation study were based on the estimated components from the assumed three-component mixture of normal distributions in the case-study presented in Chapter 4. However, predicted random effects have previously been shown to reflect the assumed random effect distribution instead of the true distribution (Verbeke and Lesaffre, 1996; McCulloch and Neuhaus, 2011b). Furthermore, the extreme response patterns of the latent stayers have previously been modelled by incorporating spikes in the random effects at negative and positive infinity (Davies et al., 1992). Therefore, the true random effect variability within a latent mover-stayer scenario may be more extreme than the distributions considered here. However, the true random intercept distributions considered in this simulation study have provided a novel insight into incorrectly assuming normality within a potential mover-stayer scenario.

This simulation study suggests that maximum likelihood estimates of random intercept logistic models can be impacted by random effect distributional misspecification in panel survey applications. Assuming normal random intercepts when the true random intercept is distributed as an asymmetric three component mixture of Gaussians generally results in minimal impact on the estimation of the fixed effects, the parameters often of interest in practice. However, misspecification may produce bias in the estimates of parameters capturing the effects of time-invariant categorical variables with minimal within-subject variability and small cell sizes.

Furthermore, misspecification can induce large bias in the estimates of the intercept constant and variance component. Arguably, these parameters may often not be of primary interest, yet biased estimates may subsequently impact subject-specific interpretation and inferences of the random effect variability, such as the ICC. Therefore, users of random intercept logistic models in panel survey applications should exercise caution when interpreting the intercept constant or variance components when multimodal random effects are suspected. If attrition is assumed to be MAR, there is minimal additional impact on the maximum likelihood estimation in the presence of random effect distribution misspecification.

The asymmetrical random intercept distributions considered in this chapter are motivated by the multimodality identified in the case study in Chapter 4. The random effect variability in this simulation study has been restricted to variances ranging from 8 to 13, however in practice, the variance of the random intercepts may be larger. If the true asymmetric multimodal random intercepts considered in this simulation study can impact inference of maximum likelihood estimates, how robust is the assumed normal distribution to other multimodal distributions? More specifically, when does random intercept misspecification first start to impact interpretation? To address this question, Chapter 6 considers a range of true symmetric three component mixture of normal distributions varying in severity of departures from normality.

6 Investigating the robustness of random intercept logistic models for departures to the normality assumption characterised by symmetric three component normal mixture distributions

6.1 Introduction

Multimodality of the underlying random effects distribution can be a consequence of a latent population substructure, such as a potential mover-stayer scenario as identified in the HILDA case study (Chapter 4). However, multimodality of the random effects distribution can also arise when key time-invariant categorical explanatory variables are omitted from the mean structure of the model (Ghidey et al., 2010). For instance, in an application of a random intercept logistic model to compare the efficiency of a drug in a longitudinal clinical trial, Komarek and Lesaffre (2008) and Verbeke and Molenberghs (2013) demonstrated random intercepts following a multimodal distribution with three modes. The underlying multimodality of the random intercept distribution was postulated to be indicative of an important explanatory variable being omitted from the model (Komarek and Lesaffre, 2008).

Although multimodality of the random effects can occur in practice, limited research has considered the impact of misspecifying the random effects distribution when the true distribution is multimodal. Of the simulation studies that have considered true multimodal distributions, the focus has been on bimodal distributions (e.g. Chen et al. 2002; Komarek and Lesaffre 2008; Litiére et al. 2008; McCulloch and Neuhaus 2011a). Previously Litiére et al. (2008) reported biased estimation of the fixed and random effect parameters in logistic mixed models when true skewed and multimodal (symmetric and asymmetric) distributions were incorrectly assumed to be normally distributed, particularly for random effects with larger variability. However, no studies have considered investigating the impact of incorrectly assuming normality for true trimodal distributions. The results from the simulation study in Chapter 5 highlight a scenario motivated by the potential mover-stayer scenario, whereby incorrectly assuming normality for an underlying asymmetric trimodal random intercept distribution can impact inferences for the intercept constant and the random effects variance component.

However, the impact of multimodality on estimating model parameters in the presence of distributional misspecification of the random effects can differ depending on whether the true distribution is symmetric or asymmetric (Litiére et al., 2008). In comparison to a true asymmetric bimodal distribution for the random intercepts, smaller magnitudes of bias were reported by Litiére et al. (2008) when a true symmetric bimodal distribution was incorrectly assumed to be normal. The sensitivity to skewness has been reported in the context of biased estimation of the random intercept, with larger bias expected to occur when a true asymmetric distribution is incorrectly assumed to be symmetric (Neuhaus et al., 1992). Therefore to remove potential confounding effects of skewness, investigation of the impact of multimodal true random effects may require the consideration of symmetric multimodal distributions.

For true bimodal symmetric random intercept distributions, incorrectly assuming normality can result in modest bias for estimates of the time-invariant explanatory variables (relative bias of up to -14%) and estimates of the random effect variability (relative bias of up to 31%) (Litiére et al., 2008). Similarly, in the context of a random intercept and random slope logistic model, incorrectly assuming normality for a true bimodal symmetric bivariate distribution can produce bias in estimating the effects of both time-invariant and time-varying explanatory variables (Litiére et al., 2008; McCulloch and Neuhaus, 2011a)[1]. Previous research suggests that the impact of misspecification on estimating model parameters can be dependent on how different the true distribution for the random effects is from the assumed distribution, and the variability of the true random effects (Heagerty and Kurland, 2001; Litiére et al., 2008; Vock et al., 2014). Therefore, to further investigate the robustness of random intercept logistic models to multimodality of the random intercept distribution, studies in this chapter consider the impact of incorrectly assuming normality for various symmetric trimodal distributions differing in the severity of departure from a normal distribution.

The primary research objective is to identify scenarios within panel survey settings when random intercept misspecification may first start to impact inferential conclusions. To address this question, the random intercepts are generated from a symmetric three component mixture of normal distributions, initially considering a true normal distribution and increasing in departures from normality. By focusing on the specific case of symmetric multimodal distributions, the component variances and the component mixing proportions can be fixed, allowing the impact of increasing component mean distances to be assessed. Thus, for larger increases in the component mean distances, the more extreme is the departure from the assumed normality distribution, resulting in a distribution with three distinct modes. Additionally, the impact of more extreme multimodality is assessed by considering increasing values for the fixed component variances, as smaller values of the component variance will result in a distribution with more distinct modes. The secondary research objective is to investigate whether missingness due to

[1]McCulloch and Neuhaus (2011a) consider the same simulation scenario as presented in Litiére et al. (2008), however report different magnitudes of bias for the time-varying coefficient. Litiére et al. (2008) (in the correction, Table 1) reported relative bias of -56% however, McCulloch and Neuhaus (2011a) reported bias of -14%. Furthermore, the simulation studies of Litiére et al. (2008) had low convergence rates with up to 30% of analyses failing to converge.

attrition additionally influences the impact of misspecifying the random effect distribution. As in Chapter 5, by considering a simulation based on the HILDA case study in Chapter 4, the impact of misspecification in the presence of complete data and missing data following from MAR attrition is assessed. These studies thoroughly examine the consequences of misspecifying the random effects distribution by assessing estimation bias, confidence interval coverage and the accuracy of model based standard errors.

6.2 Simulation study design

This study simulates data from a random intercept logistic model with random intercepts generated from twenty-one different symmetric mixture distributions increasing in departure from unimodality of the normal distribution, three different component variances and two missing data scenarios. The processes used to simulate data from the random intercept logistic model and to simulate the missingness scenario are detailed below.

The parameters and design matrix of this simulation study are derived from the analysis of women with complete employment data over 11 waves of the HILDA survey as described in the motivating example (Chapter 4). Using the techniques described in Section 3.4.1 and similar to the data generating techniques described in Section 5.2.1, resampling techniques were used to generate the values for the explanatory variables. For each iteration of the simulation, 1000 women were randomly selected without replacement from the 1359 women with complete cases. The explanatory variables of those 1000 women were used to generate clustered binary responses using the random intercept logistic model presented in Equation 4.1. The same fixed effect parameter values utilised in Section 5.2.1 were used to generate the responses ($\beta_0 = 1.2, \beta_1 = 0.1, \beta_2 = -0.3, \beta_3 = -0.1, \beta_4 = -1.5, \beta_5 = -2.8, \beta_6 = -2.3$ and $\beta_7 = -0.4$).

In this simulation study, the random intercept b_i is generated from a symmetric three component mixture of normal distributions. The symmetric random intercept distribution has equal component proportions (i.e. $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$) and equal component variances (i.e. $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$),

$$b_i \sim \frac{1}{3}N(\mu_1, \sigma^2) + \frac{1}{3}N(\mu_2, \sigma^2) + \frac{1}{3}N(\mu_3, \sigma^2) \tag{6.1}$$

where μ_1, μ_2 and μ_3 are the component means. To identify scenarios where misspecifying the random effect distribution impacts model interpretation, a range of random intercept distributions are generated by considering twenty-one combinations of the component means (with increasing departures from normality characterised by distributions with three modes). The random intercept distributions are of increasing component mean distances, with fixed component variances of either $\sigma^2 = 1, 2$ or 4. The different mean combinations for μ_1, μ_2 and μ_3 are selected to have a symmetric distribution with mean zero. The specific case considered here is where $\mu_1 = -\mu_3$ and $\mu_2 = 0$, with μ_3 ranging from 0 to 10, increasing in increments of 0.5. Density plots for selected simulated random intercept distributions of increasing mean component distances (defined as $\mu_3 - \mu_1$) for the three component variance scenarios are shown in Figure 6.1. For the twenty-one mean scenarios and three fixed component variances, Table 6.1 presents the overall true random effect variance for each simulated scenario (defined as $\sigma_b^2 = \sum_{k=1}^{3} \pi_k \left(\sigma^2 + (\mu_k)^2\right) - \left(\sum_{k=1}^{3} \pi_k \mu_k\right)^2 = \sigma^2 + \frac{2}{3}\mu_3^2$, for $k = 1, 2, 3$) .

Table 6.1: True total random intercept variance σ_b^2 of the simulated symmetrical three component mixture of normal distributions for component mean distances $(\mu_3 - \mu_1)$ ranging from 0 to 20, and component variances fixed at either $\sigma^2 = 1, 2$ or 4. The total random intercept variance is defined by $\sigma_b^2 = \sigma^2 + \frac{2}{3}\mu_1^2$.

Component Mean Distance	Component Variances $\sigma^2 = 1$	$\sigma^2 = 2$	$\sigma^2 = 4$
$\mu_3 - \mu_1$=0	1.00	2.00	4.00
$\mu_3 - \mu_1$=1	1.17	2.17	4.17
$\mu_3 - \mu_1$=2	1.67	2.67	4.67
$\mu_3 - \mu_1$=3	2.50	3.50	5.50
$\mu_3 - \mu_1$=4	3.67	4.67	6.67
$\mu_3 - \mu_1$=5	5.17	6.17	8.17
$\mu_3 - \mu_1$=6	7.00	8.00	10.00
$\mu_3 - \mu_1$=7	9.17	10.17	12.17
$\mu_3 - \mu_1$=8	11.67	12.67	14.67
$\mu_3 - \mu_1$=9	14.50	15.50	17.50
$\mu_3 - \mu_1$=10	17.67	18.67	20.67
$\mu_3 - \mu_1$=11	21.17	22.17	24.17
$\mu_3 - \mu_1$=12	25.00	26.00	28.00
$\mu_3 - \mu_1$=13	29.17	30.17	32.17
$\mu_3 - \mu_1$=14	33.67	34.67	36.67
$\mu_3 - \mu_1$=15	38.50	39.50	41.50
$\mu_3 - \mu_1$=16	43.67	44.67	46.67
$\mu_3 - \mu_1$=17	49.17	50.17	52.17
$\mu_3 - \mu_1$=18	55.00	56.00	58.00
$\mu_3 - \mu_1$=19	61.17	62.17	64.17
$\mu_3 - \mu_1$=20	67.67	68.67	70.67

Simulations were performed under two missing data scenarios: complete data (i.e. no missingness imposed) and incomplete data due to attrition. Attrition was assumed to be generated by the missing at random (MAR) mechanism. The same methodology used to impose MAR attrition in Chapter 5 as described in Sections 3.4.2 and 5.2.2 was implemented in this simulation study, resulting in similar wave-to-wave attrition rates observed in the HILDA case study (Table 4.1) and overall attrition rate of 29.5%.

For each of the 21 mean component distances, three variance component settings and two missingness scenarios (126 combinations in total), 1000 datasets consisting of 1000 subjects were generated. A random intercept logistic model assuming Gaussian random effects was fitted to each simulated dataset. When $\mu_1 = \mu_2 = \mu_3 = 0$ the normality assumption is true, however the departure from normality increases as $\mu_3 - \mu_1 > 0$ increases. The simulation study

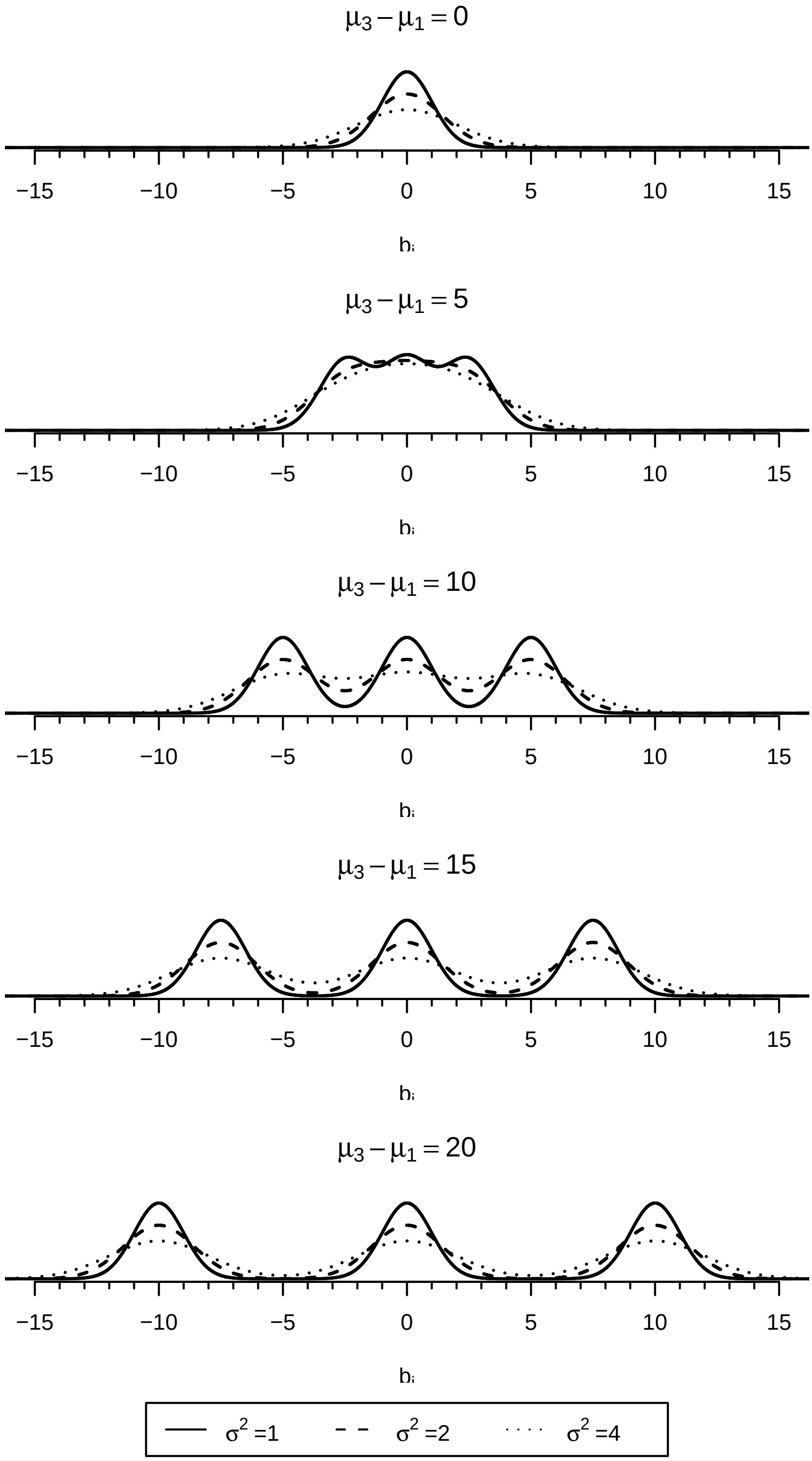


Figure 6.1: Density of the true random intercept distributions for selected component mean distances $\mu_3 - \mu_1 = 0, 5, 10, 15$ and 20 for the three component variance scenarios: $\sigma^2 = 1, 2$ and 4.

examines the robustness of the normality assumption of the random intercept distribution by considering estimation bias, confidence interval coverage and accuracy of model-based standard errors for the estimated model parameters. The performance measures and criteria for acceptable performance are discussed in Section 3.5, and respectively were, percentage bias within -10% and 10%, coverage rates of the 95% confidence intervals within 93.6% and 96.4%, and standard error ratios within 0.9 and 1.1. For the performance measures relating to the random intercept distribution, the variance estimate of the random intercept was compared to the overall variance of the true three component mixture distribution.

Simulations and analyses were conducted in SAS (Version 9.4, SAS Institute, Cary NC). All random intercept logistic models were fitted using the SAS procedure NLMIXED with adaptive Gaussian Quadrature using 20 quadrature points.

6.3 Simulation study results

Due to the large number of scenarios considered in this simulation study (21 mean component distances, 3 component variances and two missingness scenarios) the results are presented over the following four sections. The first two sections correspond to the two research objectives. The results regarding the impact of misspecifying the random intercept distribution with increasing departures from normality are presented in Section 6.3.1 and in Section 6.3.2 the impact of misspecification of the random effects distribution in the presence of MAR attrition is explored. Additionally, Section 6.3.3 presents the results when correctly assuming normality and Section 6.3.4 summarises the attrition rates and the underlying extreme response patterns generated in the simulated datasets.

6.3.1 Severity of departure from normality of the random intercepts

Figure 6.2 presents the percentage bias of the parameter estimates when applying the random intercept logistic model assuming normality to the simulated data for the three component variances scenarios ($\sigma^2 = 1, 2, 4$) in the complete data scenario. With increasing distance between the three component means, defined as $\mu_3 - \mu_1$, Figure 6.2 shows that misspecification generally produced unbiased estimates for all parameters for small to moderate deviations from the assumed normal distribution ($\mu_3 - \mu_1 < 13$). For severe departures from normality characterised by distinct multimodality ($\mu_3 - \mu_1 \geq 13$), misspecification produces biased estimators for parameters associated with the random intercept distribution, the intercept constant (β_0) and random intercept standard deviation (σ_b), and for parameters of the time-invariant explanatory variables capturing education at the first wave, β_4 and β_5. Estimation of the explanatory variable capturing the effect of single women (compared to married or defacto women), β_3, was underestimated for the majority of the scenarios, producing large negatively biased estimates.

Estimation of the intercept constant β_0 was consistently below the threshold of -10 (range:

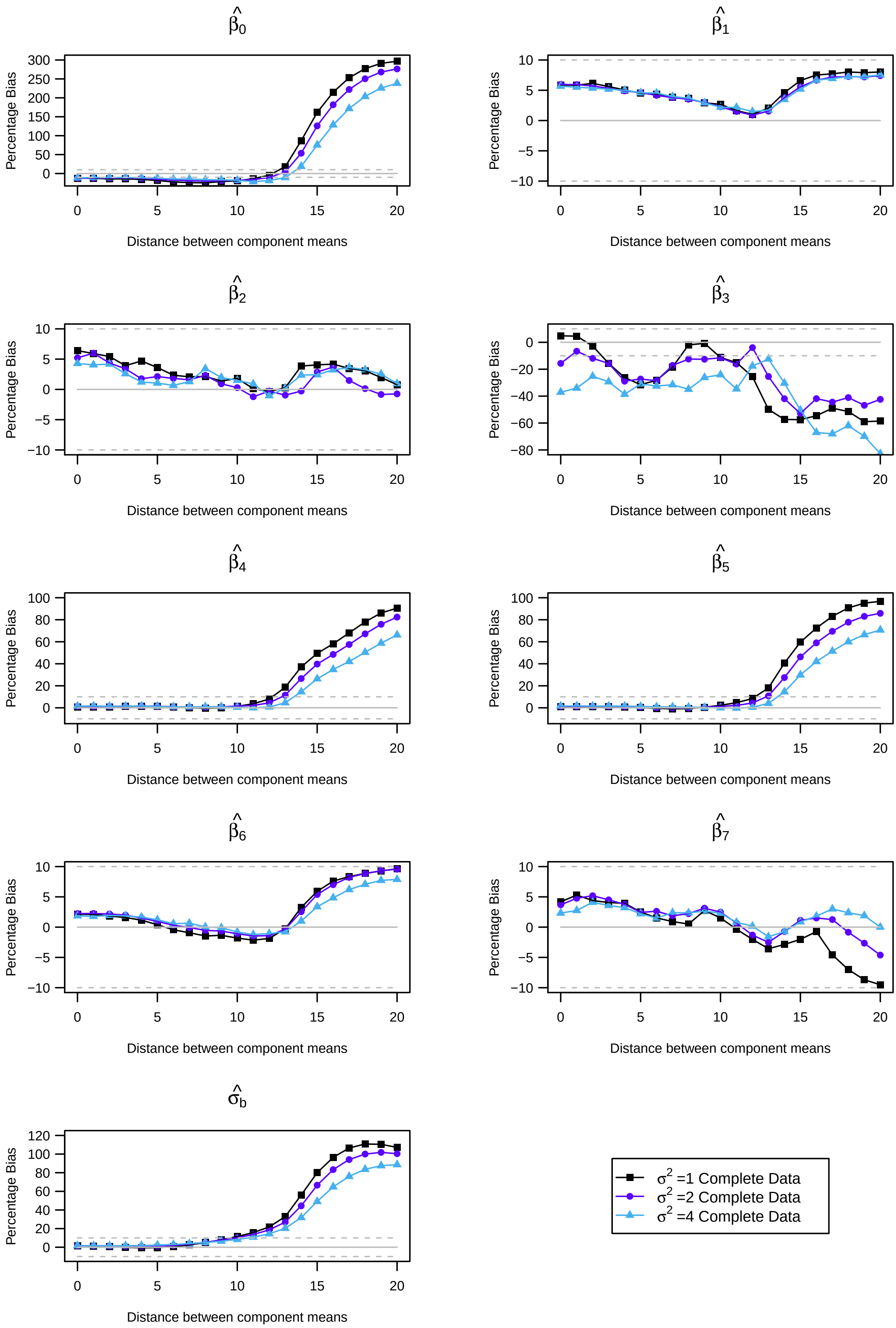


Figure 6.2: Percentage bias for parameter coefficients of the random intercept logistic model applied to the complete data scenario for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under three component variance scenarios ($\sigma^2 = 1, 2$ and 4). Grey horizontal solid line at 0 percentage bias and grey horizontal dashed lines at percentage bias of -10% and 10%.

-10.5 to -23.9) for minor departures from normality ($\mu_3 - \mu_1 < 11$), before increasing non-linearly as component mean distances exceeded 13, reaching percentage bias in excess of 200. A similar trend as for β_0 was observed for the random intercept standard deviation σ_b. Minor distributional misspecification ($\mu_3 - \mu_1 < 11$) produced unbiased estimates for σ_b (range: -0.6 to 11.5), before increasing non-linearly as component mean distances exceeded 10, reaching percentage bias in excess of 100. Similarly, as the departure from normality becomes more severe ($\mu_3 - \mu_1 \geq 14$), misspecifying the random intercept distribution resulted in large biased estimates for the time-invariant explanatory variables relating to the highest education attained at baseline, β_4 and β_5, reaching percentage bias exceeding 60. Negative bias was experienced for the covariate relating to single women, β_3, generally underestimating the true coefficient for majority of scenarios. As the departure from normality became more extreme ($\mu_3 - \mu_1 \geq 16$), the percentage bias for β_3 exceeded -40%. Misspecification produced little and relatively consistent bias for the remaining time-varying explanatory variables, with percentage bias within the $\pm 10\%$ threshold for β_1, β_2, β_6 and β_7 for all component mean distances.

As the variance components of the true random effects distribution increased, misspecification generally produced similar trends as for $\sigma^2 = 1$. For minor deviations from the assumed normality, the three variance components resulted in similar magnitudes of percentage bias. However, for moderate to large deviations from normality, differences in the magnitude of the percentage bias became more apparent. As the component mean distances exceeded 12, the magnitude of the percentage bias for the three component variances started to diverge, particularly when estimating $\beta_0, \beta_4, \beta_5$ and σ_b. Generally as the component variances increased, the magnitude of bias decreased. Thus, the smallest component variance $\sigma^2 = 1$, resulting in the most extreme multimodal distribution, generally resulted in the largest magnitude of bias. One exception was the coefficient capturing the effect of single women (β_3), whereby a similar trend and order of magnitude was observed for all component variance scenarios, with $\sigma^2 = 4$ scenario generally producing the largest negative values of percentage bias.

Figure 6.3 presents the coverage rates of 95% confidence intervals for the parameter estimates when applying the random intercept logistic model assuming normality to the simulated data for the three component variances scenarios ($\sigma^2 = 1, 2, 4$) in the complete data scenario. Figure 6.3 indicates that the coverage rates for $\beta_1, \beta_2, \beta_3, \beta_6$ and β_7 were typically close to the nominal rate of 95% for all component mean distances. As the distance between component means exceeded 14, severe misspecification of the random effects resulted in poor coverage rates for the intercept constant, β_0, and the two coefficients of the time-invariant education explanatory variable, β_4 and β_5. For these parameters, coverage rates were close to the nominal rate for minor and moderate deviations from normality ($\mu_3 - \mu_1 < 14$), before declining and reaching coverage rates of approximately 50%, 80% and 70%, respectively. Extremely poor coverage rates were experienced for estimates of the variance component, with nominal coverage rates for σ_b less than the acceptable 93.6% threshold for component mean distances of 7 or more.

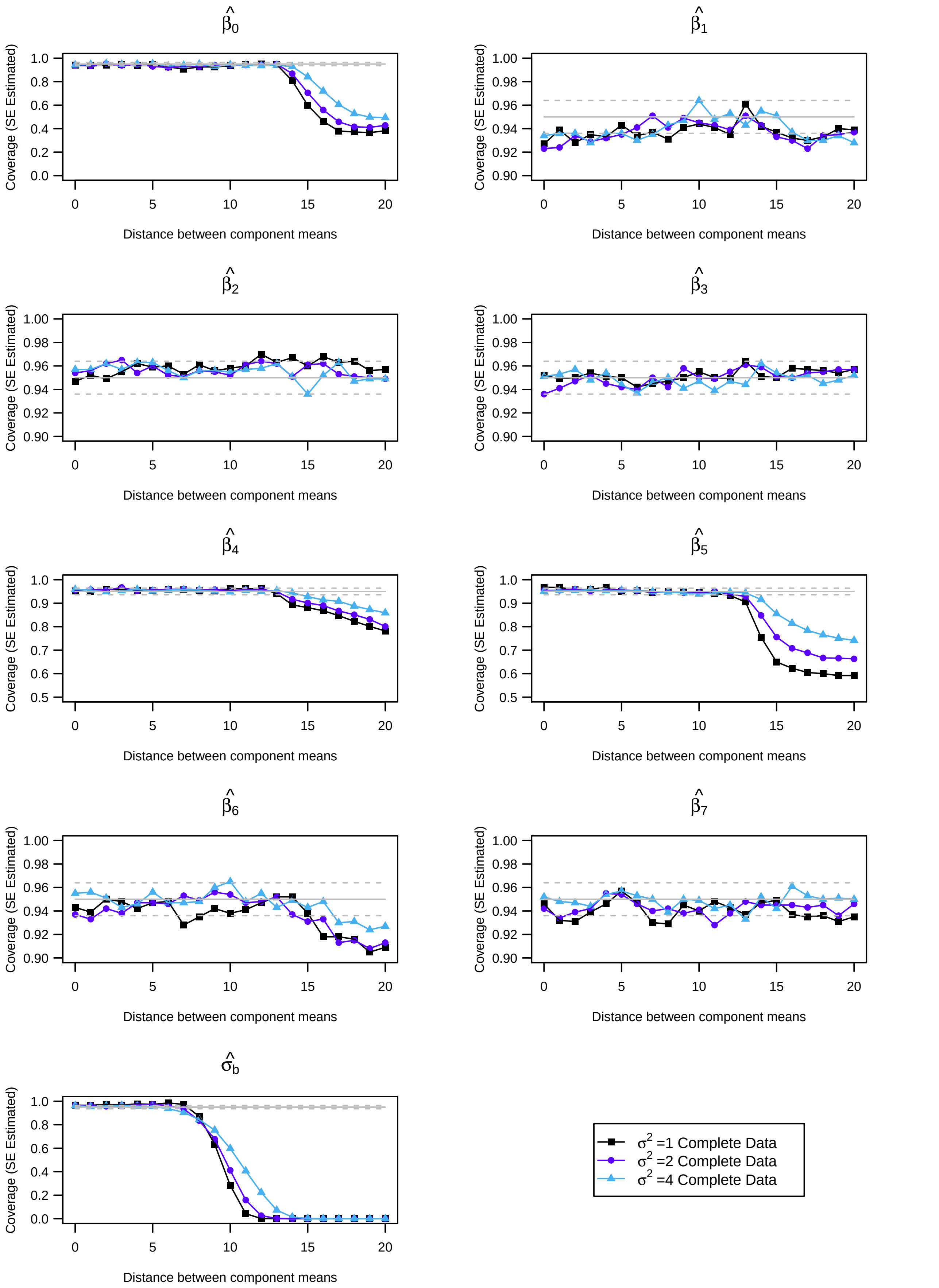


Figure 6.3: Coverage rates for model based 95% confidence intervals for parameter coefficients of the random intercept logistic model applied to the complete data scenario for increasing distances of random intercept component means $(\mu_3 - \mu_1)$ under three component variance scenarios $(\sigma^2 = 1, 2 \text{ and } 4)$. Grey horizontal solid line at nominal coverage rate 0.95 and grey horizontal dashed lines at coverage rate 0.936 and 0.964.

The coverage rates declined rapidly before reaching coverage rates of 0% for severe departures from normality ($\mu_3 - \mu_1 \geq 14$).

Larger component variances generally corresponded with higher coverage rates, with similar trends observed for all three variance scenarios. Generally the three component variance scenarios resulted in similar magnitude of coverage rates, however as the component mean distances increased, divergence in the magnitude of coverage rates was observed. This was particularly true for terms related to the random effects, β_0 and σ_b, and the coefficients capturing the effects of the time-invariant baseline education variable β_4 and β_5. For severe departures from normality ($\mu_3 - \mu_1 > 14$) the coverage rates of β_0, β_4 and β_5 for the three component variances started to decline at differing rates. Coverage rates for σ_b started to diverge and decline rapidly for moderate departures from normality ($\mu_3 - \mu_1 > 8$), before reaching 0% coverage for all variance component scenarios for component mean distances exceeding 14. The coverage rates for the coefficients capturing the effects of the time-varying explanatory variables ($\beta_1, \beta_2, \beta_3, \beta_6$ and β_7) were typically close to the nominal coverage rate, with similar trends and magnitudes observed for the three variance scenarios.

Figure 6.4 presents the ratio of the mean of model-based standard errors to the empirical standard error of the parameter estimates when applying the random intercept logistic model assuming normality to the simulated data for the three component variances scenarios ($\sigma^2 = 1, 2, 4$) in the complete data scenario. With the exception of the random effect standard deviation, the ratio was within the acceptable range of 0.9 and 1.1 for all model parameters. This indicates that even in the presence of random effect misspecification, the model based standard errors accurately describe the variability of the fixed effect coefficients. The ratio for σ_b increased as the distance between the component means increased, with the standard error ratio exceeding 1.1 when the component mean distances ranged between 7 and 11. For component mean distances between 11 and 15, the ratio declined to values below the 0.9 threshold, before increasing again for component mean distances between 15 and 20, reaching ratio values above the 1.1 threshold.

The fluctuations observed for σ_b can potentially be explained by the non-linear increase in empirical standard error as opposed to the linear increase experienced by the other parameter coefficients (results not shown). For instance, the empirical standard error for component variance $\sigma^2 = 1$ increased linearly from 0.06 to 0.25 for component mean distances 0 and 11, respectively, until increasing exponentially from 0.52, 0.94 and 0.96 for distances of 12, 13 and 14. For distances of 15 to 20, the empirical standard error stabilised and slightly decreased from 0.96 to 0.88.

Similar magnitudes and trends of the standard error ratio were observed for all three component variances. As the component variance increased, generally the deviations from the neutral

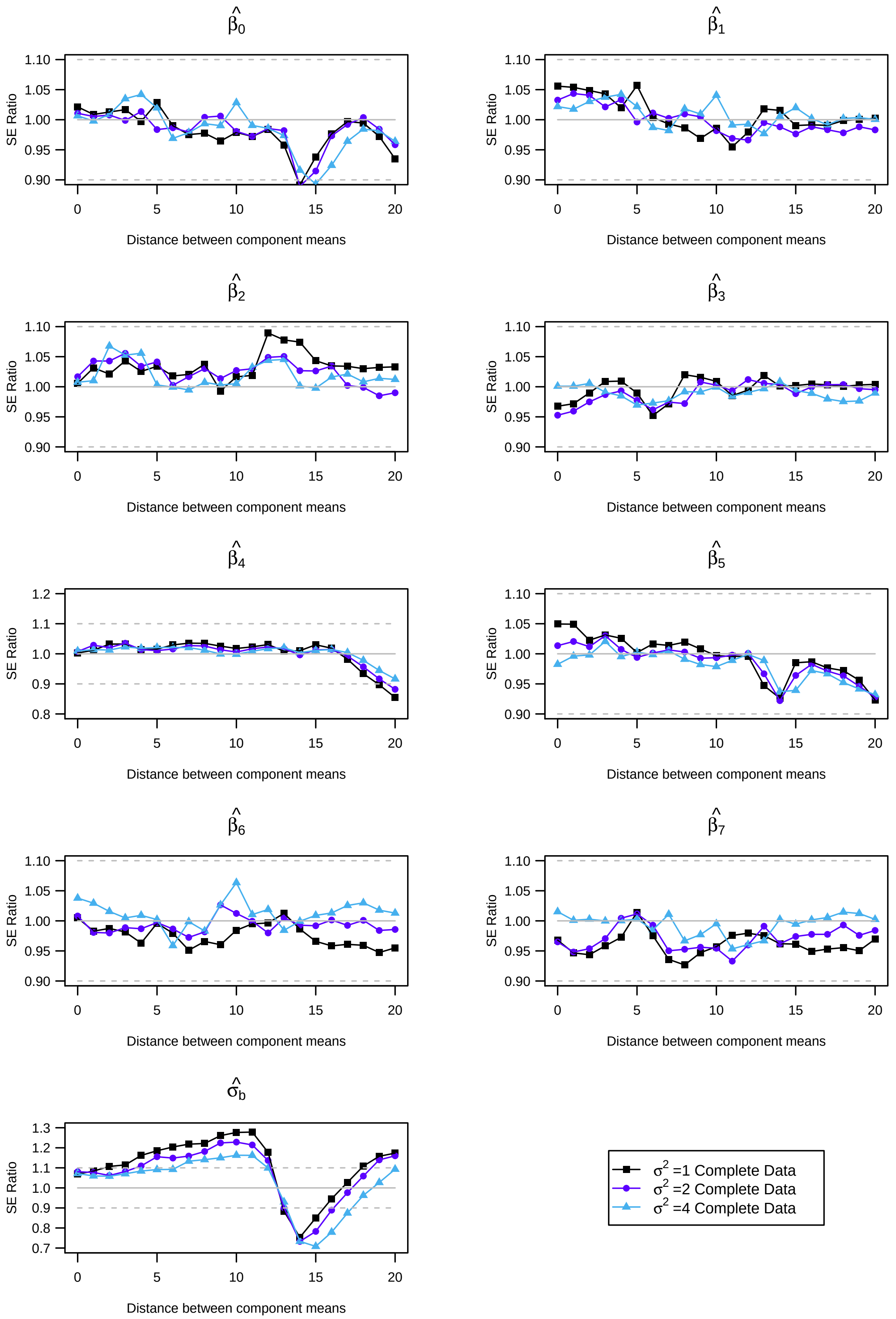


Figure 6.4: Ratio of mean model-based standard error to the empirical standard error for parameter coefficients of the random intercept logistic model applied to the complete data scenario for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under three component variance scenarios ($\sigma^2 = 1, 2$ and 4). Grey horizontal solid line at ratio=1 and grey horizontal dashed lines at ratio of 0.9 and 1.1.

standard error ratio value of 1 decreased. For increasing component mean distances, the magnitudes of the standard error ratio only deviated marginally for β_4, and for increasing trends of σ_b. With the exception of σ_b, for all component mean distances and component variance scenarios, the mean model-based standard errors were generally equivalent to the empirical standard errors.

6.3.2 Additional impact of attrition

In the presence of MAR attrition, incorrectly assuming normality for random intercepts when the true random effects were distributed as a symmetric three component mixture of normals generally resulted in similar magnitudes and trends as observed in the complete data scenario. The percentage bias, coverage rates and standard error ratio for Complete data (C) and MAR attrition (MAR), in Tables 6.2, 6.3 and 6.4 respectively, present the performance measures for the component variance scenario $\sigma^2 = 1$ and selected component mean distances ($\mu_3 - \mu_1 = 0, 5, 10, 15, 20$).

Random effect distributional misspecification in the presence of MAR attrition generally produced similar trends in the percentage bias as for the complete data (Table 6.2). With the exception of β_3, the magnitude of percentage bias for time-varying parameters was similar yet marginally larger than the complete data scenario (mean difference (SD) between MAR attrition and complete data (MAR - C) of 0.7 (0.6), 0.3 (1.1), 0.6 (0.9) and 0.7 (1.2), for $\beta_1, \beta_2, \beta_6$ and β_7). The trend in percentage bias for β_3 over the component mean distance was similar for the two data scenarios, however differences in magnitude between MAR attrition and complete data varied and fluctuated between -8.8 and 18.2. Increasing distance between component means generally resulted in larger differences in the magnitude of bias between the two data scenarios (overall mean difference (SD) of -0.4 (1.8) and -5.3 (13.5) for component mean distance of 0 and 20, Table 6.2). This was particularly true for coefficient estimates of the time-invariant fixed effect parameters and parameters related to the random effect. In the presence of MAR attrition, estimation of the time-invariant parameters, β_4 and β_5, resulted in similar magnitude of bias as the complete data scenario for minor deviations from normality ($\mu_3 - \mu_1 \leq 10$). However, for larger deviations ($\mu_3 - \mu_1 > 10$), misspecification in the presence of MAR attrition resulted in smaller magnitude of percentage bias for β_4 and β_5 than the complete data, with larger differences between the two data scenarios observed with the increasing component mean distances. For parameters related to the random intercept distribution, β_0 and σ_b, the difference in the means between the two data scenarios was negligible for minor deviations from normality ($\mu_3 - \mu_1 \leq 10$). However, differences became larger in magnitude for distances 12 to 15 reaching differences of 15.4 and 13.4 for β_0 and σ_b respectively, before declining to differences of 3.6 and 1.7 at the most extreme mean distance of 20. As highlighted by the bold text in Table 6.2, when the complete data resulted in bias outside the acceptable performance range of $\pm 10\%$, the corresponding bias in the presence of MAR attrition would

Table 6.2: Percentage bias for parameter estimates of the random intercept logistic model with random effects simulated as a symmetric three component mixture distribution with $\sigma^2 = 1$ and selected component mean distances ($\mu_3 - \mu_1 = 0, 5, 10, 15, 20$) for the two missing data scenarios: complete data (C) and MAR attrition (MAR). The values in bold indicate the percentage bias not within the acceptable performance range of -10% and 10%.

	Distance between component means									
Parameter	0		5		10		15		20	
	C	MAR	C	MAR	C	MAR	C	MAR	C	MAR
β_0	**-12.4**	**-14.1**	**-18.4**	**-19.6**	**-19.4**	**-23.9**	**161.5**	**176.9**	**296.6**	**300.2**
β_1	5.8	6.3	4.5	5.0	2.7	3.4	6.6	8.2	8.0	7.5
β_2	6.4	7.6	3.6	3.4	1.8	1.5	4.1	6.8	0.8	0.1
β_3	4.8	0.2	**-31.5**	**-37.2**	**-11.3**	**-17.5**	**-57.5**	**-39.2**	**-58.5**	**-60.7**
β_4	0.6	0.6	1.4	1.9	1.4	2.0	**49.4**	**35.0**	**90.5**	**53.3**
β_5	0.9	0.9	0.3	0.5	2.3	1.0	**59.7**	**52.8**	**96.7**	**79.7**
β_6	2.1	2.2	0.4	0.5	-1.8	-2.0	5.9	7.8	9.7	**11.4**
β_7	4.2	5.2	2.5	2.7	1.5	0.8	-2.0	0.0	-9.5	-6.6
σ_b	1.5	1.4	-0.4	-0.1	**11.5**	**12.2**	**80.0**	**93.4**	**107.1**	**108.8**
Mean Difference (SD)	-0.4 (1.8)		-0.6 (2.0)		-1.2 (2.5)		3.8 (10.6)		-5.3 (13.5)	
Minimum Difference	-4.6		-5.7		-6.2		-14.4		-37.2	
Maximum Difference	1.2		0.5		0.8		18.2		3.6	

also be deemed unacceptable.

Misspecification of the random effects in the presence of MAR attrition generally resulted in similar trends and magnitude of coverage rates to those for complete data (Table 6.3). For all the component mean distances, the mean difference of the coverage rates between the two data scenarios (MAR - C) for all parameters were within 0 and 0.04 (results not shown). The coverage rate observed for the two data scenarios was similar for component mean distances less than 10, with mean difference (SD) of 0 (0.01) for $\mu_3 - \mu_1 = 0, 5$ and 10 (Table 6.3). As the component mean distances increased, differences between the coverage rates became larger (mean difference (SD) of 0.02 (0.04) and 0.03 (0.05) for component mean distances of 15 and 20). Coverage rates for coefficients relating to the time-varying parameter differences were similar across all component mean distances (mean differences (SD) of 0 (0.01) for $\beta_1, \beta_2, \beta_3, \beta_6$ and β_7, results not shown), however, for coefficients capturing the time-invariant parameters, β_4 and β_5, increasing distance between component means resulted in larger differences in coverage rates between the two data scenarios. For larger departures from normality ($\mu_3 - \mu_1 > 14$), the difference in the coverage rates increased from 0 to reach 0.09 for β_4 and from 0.02 to reach 0.13 for β_5. Similarly, the difference in coverage rates for β_0 marginally increased from 0.03 for component mean distance of 15 to 0.05 for component mean distance of 20. Coverage rates for the variance component estimate, σ_b differed between the two data scenarios for component mean distances ranging from 10 to 14, before converging to coverage rates of 0 for component mean distances exceeding 15. With some minor exceptions, unacceptable coverage rates identified for the complete data were also outside the acceptable limits for the MAR attrition data scenario (Table 6.3).

The impact of misspecification in the presence of attrition produced similar accuracy of model based standard errors as for the complete data (Table 6.4). In comparison to complete data scenario, standard error ratios of the fixed effect parameters were of similar magnitudes for the MAR missing data scenario. With the exception of σ_b, all parameters had negligible difference of the standard error ratio between the MAR missing and complete data scenarios for component mean distances less than 10 (range of differences:-0.02 to 0.04, results not shown). For component mean distances of 15 and 20, the mean difference (SD) of the standard error ratio between the MAR missing and complete data scenarios increased from 0.009 (0.019) to 0.032 (0.017). As the component mean distances increased, larger differences in the standard error ratios between the data scenarios were noticed for β_0, β_3 and β_4, reaching respective differences between MAR missing and complete data scenarios of 0.05, 0.04 and 0.04 at the most extreme component mean distance ($\mu_3 - \mu_1 = 20$). In the presence of attrition, the standard error ratio for σ_b was smaller than the complete data analysis for component mean distances less than 15 (range of differences: -0.16 to -0.03). However, for mean distances exceeding 15 the MAR missing data scenario produced more extreme standard error ratio for σ_b than the complete data (range of differences: 0.01 to 0.07). Albeit some differences in the magnitude

Table 6.3: Coverage rates for model based 95% confidence intervals for parameter coefficients of random intercept logistic model with random effects simulated as a symmetric three component mixture distribution with $\sigma^2 = 1$ and selected component mean distances ($\mu_3 - \mu_1 = 0, 5, 10, 15, 20$) for the two missing data scenarios: complete data (C) and MAR attrition (MAR). The values in bold indicate the coverage rate is not within the acceptable performance range of 0.936 and 0.964.

Parameter	**Distance between component means**									
	0		5		10		15		20	
	C	MAR	C	MAR	C	MAR	C	MAR	C	MAR
$\hat{\beta}_0$	0.942	**0.933**	0.943	0.945	**0.934**	**0.927**	**0.599**	**0.625**	**0.382**	**0.429**
$\hat{\beta}_1$	**0.927**	**0.928**	0.943	0.954	0.944	0.943	0.937	**0.929**	0.939	0.955
$\hat{\beta}_2$	0.947	0.949	0.959	0.955	0.958	0.954	0.960	0.956	0.957	**0.966**
$\hat{\beta}_3$	0.952	0.951	0.950	0.941	0.955	0.947	0.950	0.949	0.957	0.962
$\hat{\beta}_4$	0.952	0.956	0.955	0.953	0.962	0.960	**0.881**	**0.932**	**0.782**	**0.882**
$\hat{\beta}_5$	**0.968**	**0.968**	0.951	0.955	0.945	0.948	**0.650**	**0.763**	**0.592**	**0.711**
$\hat{\beta}_6$	0.943	0.940	0.947	0.944	0.938	0.949	0.938	**0.925**	**0.909**	**0.920**
$\hat{\beta}_7$	0.946	0.943	0.957	0.961	0.940	0.944	0.949	0.953	**0.935**	0.951
$\hat{\sigma}_b$	**0.967**	0.953	**0.973**	**0.972**	**0.284**	**0.352**	**0**	**0**	**0**	**0**
Mean Difference (SD)	0.00 (0.01)		0.00 (0.01)		0.00 (0.01)		0.02 (0.04)		0.03 (0.05)	
Minimum Difference	-0.014		-0.009		-0.008		-0.013		0	
Maximum Difference	0.004		0.011		0.068		0.113		0.119	

Table 6.4: Ratio of the mean model-based standard error to the empirical standard error of the random intercept logistic model with random effects simulated as a symmetric three component mixture distribution with $\sigma^2 = 1$ and selected component mean distances ($\mu_3 - \mu_1 = 0, 5, 10, 15, 20$) for the two missing data scenarios: complete data (C) and MAR attrition (MAR). The values in bold indicate the standard error ratio is not within the acceptable performance range of 0.9 and 1.1.

	Distance between component means									
Parameter	0		5		10		15		20	
	C	MAR	C	MAR	C	MAR	C	MAR	C	MAR
β_0	1.021	1.012	1.029	1.041	0.979	0.989	0.938	0.959	0.935	0.990
β_1	1.056	1.046	1.057	1.060	0.986	0.990	0.990	1.005	1.002	1.009
β_2	1.006	1.019	1.034	1.029	1.017	1.012	1.043	1.035	1.033	1.045
β_3	0.968	0.984	0.990	0.992	1.009	1.001	1.002	1.015	1.004	1.046
β_4	1.003	1.004	1.016	1.021	1.018	1.013	1.030	1.070	**0.855**	0.900
β_5	1.050	1.052	1.002	0.988	0.997	0.989	0.985	0.990	0.923	0.937
β_6	1.005	0.992	0.996	1.006	0.984	1.001	0.966	0.986	0.955	0.987
β_7	0.968	0.970	1.014	1.034	0.957	0.966	0.961	0.970	0.970	1.002
σ_b	1.069	1.018	**1.185**	**1.129**	**1.277**	**1.234**	**0.850**	**0.823**	**1.173**	**1.219**
Mean Difference (SD)	-0.005 (0.020)		-0.003 (0.023)		-0.003 (0.017)		0.009 (0.019)		0.032 (0.017)	
Minimum Difference	-0.016		-0.020		-0.017		-0.040		-0.055	
Maximum Difference	0.051		0.057		0.042		0.027		0.007	

of the standard error ratio, similar conclusions about the accuracy of the standard errors were made for the two data scenarios.

Similar small differences in the performance measures estimated under the two data scenarios were noted for the other component mean distances and component variances $\sigma^2 = 2$ and $\sigma^2 = 4$ (results not shown).

6.3.3 Correctly assuming normality

Assessing the performance of the random intercept logistic model when the assumed and true random intercepts are normally distributed, not only provides evidence to gauge the performance of the simulation study but assesses inferential impact of correctly assuming normality in a panel data setting. For component mean distances of zero, correctly assuming normally distributed random intercepts generally produced consistent estimates, with minimal bias, close to nominal coverage rates and accurate model based standard errors.

Some minor exceptions include bias in the estimation of β_0, with percentage bias ranging from -11.5% to -14.1% for the three variance components and both data scenarios. This observed bias in β_0 may be due to variability of the mean for the simulated true random effect distribution, that has subsequently been captured by the intercept constant. For the variance components $\sigma^2 = 2$ and 4, estimation of β_3 resulted in biased estimates of -15.7% (-20.7%) and -37.1% (-36%) for the complete (MAR missing) data scenario (results not shown). The observed variability in the coefficient estimates may be a consequence of the small magnitude of the true coefficient value (-0.1) and the relatively large standard error estimated for the explanatory variable in the HILDA case study as shown in Table 4.3[2]. In terms of coverage rates, there were marginal deviations from the acceptable performance limits of 93.6% and 96.4% for two parameters, β_1 and σ_b. For variance components $\sigma^2 = 1, 2$ and 4, coverage rates for β_1 were 92.7% (92.8%), 92.3% (92.8%) and 93.4% (94%) for the complete (MAR missing) data scenario. Coverage rates for estimating σ_b were larger than the acceptable upper limit in the complete data scenario for variance component scenario $\sigma^2 = 1$ and 2, however were within the acceptable limits for all variance component scenarios for MAR missing data scenario. Coverage rates of β_5 were marginally above the nominal rate for the $\sigma^2 = 1$ variance component, however were within acceptable limits for $\sigma^2 = 2$ and 4 (results not shown). The standard error ratio was within the acceptable limits of 1 ± 0.1 for all parameters for both data scenarios and the three variance scenarios.

[2]These two contributing factors may indicate high bias in regards to percentage bias, however in absolute terms the bias for β_3 for $\sigma^2 = 2$ and 4 is only 0.016 (0.021) and 0.037 (0.036) for the complete (MAR missing) data scenario.

6.3.4 Simulated attrition rates and mover-stayer scenario

The actual rate of attrition in the 1000 simulated datasets for the 126 scenarios averaged 32.6% (range: 26.4% to 38.3%) and was similar to the observed rate of 29.5% for the HILDA subgroup of working aged women (Table 4.1). The simulated attrition rates were influenced by the distance between the component means, with attrition rates averaging 31.4% (range: 26.4% to 36.2%) for component mean distances of zero and increasing to an average of 33.4% (range: 28.5% to 38.3%) as component mean distance increased to twenty. The attrition rates for the three variance component scenarios were similar (results not shown).

In the simulation study, the rate of subjects observed to remain unemployed during the 11 waves for the complete and MAR missing data scenarios averaged 13.3% (ranging from 0% to 36.8%) and 16.7% (ranging from 1.1% to 38.2%), respectively. The number of observed stayers in the unemployed group was influenced by the true random effect distribution, with a higher proportion continuously unemployed for more extreme distributions. For the complete data scenario there was a proportion of simulated datasets with no subjects observed to experience continuous unemployment for component mean distances of 5 or smaller (32%, 28%, 17%, 6%, 0.5%, 0.03% of the 1000 simulated datasets for component mean distances 0, 1, 2, 3, 4 and 5). The rate of subjects staying employed during the 11 waves averaged 51.6% (ranging from 46.2% to 58.1%) for the complete data scenario and 54.5% (ranging from 47.8% to 63.4%) for the MAR scenario.

Fitting the assumed normal distribution using NLMIXED had excellent convergence rates of 100% with the exception of one scenario ($\mu_3 - \mu_1$=12, σ_b=4 and missingness=MAR) with 99.9% convergence.

6.4 Additional simulation studies

Simulation studies may be restricted by the parameters and design of a particular dataset, or by specific conditions imposed. To assess the reproducibility of the results presented in the simulation study described above, two additional simulation studies were undertaken to assess sensitivity to the imposed conditions and the specified random intercept distributions. The first additional study assesses the impact of misspecified random effects distribution in a random intercept logistic model applied to simulated data from a randomised clinical trial. Longitudinal data arising from a clinical trial may differ to longitudinal data arising from panel surveys in numerous ways, including the number and type of explanatory variables, number of participants and time-points, and spacing between time-points. By considering the same true symmetric random intercept distributions, component variance scenarios and missing data scenarios as in the primary simulation study, the first additional study assesses whether the results observed in the primary simulation study motivated by the HILDA case study are also observed within a clinical trial setting. A second additional study is based on the same HILDA motivating case study as presented in this chapter, however, instead of fixing the component variances used to generate the random intercept mixture distribution, the total random effect variance is fixed at $\sigma_b^2 = 20, 25, 30$ and 35. Thus, the second simulation study aims to assess whether the fixed component variances used to generate the random intercepts in the primary study, that consequently varies the total random effect variance, potentially impacts the results presented in the primary study. More details and the results of the two additional simulation studies and are presented in Appendix E and Appendix F, and are briefly described in Sections 6.4.1 and 6.4.2.

6.4.1 Simulation study within clinical trial setting

This secondary simulation study generates the same random intercepts as detailed in the primary study, however data for the random intercept logistic model were generated to represent repeated binary response data and a treatment effect in a randomised clinical trial. The simulation study design is similar to the simulation considered previously by Litiére et al. (2008), and is based on a case study comprising of patient data from a randomised clinical trial comparing two treatments for chronic schizophrenia over an eight week period.

Details regarding the simulation design are presented in Appendix E.1, and are briefly described here. Using a similar simulation study set-up as detailed by Litiére et al. (2008) repeated binary response data for a total of 1000 patients over six time-points were generated following initial administration of treatment. The binary data represents the severity of schizophrenia assessed at six fixed time-points at 0, 1, 2, 4, 6 and 8 weeks, and was generated using a random intercept logistic model with two explanatory variables: a time-invariant binary variable representing treatment (randomly allocated with equal probability) and a covariate capturing time as a continuous variable. The random intercepts were simulated from the same symmetric three component mixture of normal distributions considered in the primary simulation study, with mean components ranging from $\mu_3 = 0$ to 10 increasing in increments of 0.5, and three variance component scenarios $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1, 2$ or 4. As considered in the primary simulation study, the additional impact of attrition within a misspecified random intercept logistic model was assessed by considering two missingness scenarios: complete data and attrition following the missing at random mechanism (assuming a 30% overall attrition rate).

The results are presented in Appendix E.2, though briefly, similar conclusions as in the primary study are described. The impact of incorrectly assuming normal random intercepts when the true distribution was a symmetric three component mixture of normals was predominately restricted to parameters associated with the random effect. The assumed normal distribution was robust to minor departures from normality, however moderate to severe departures ($\mu_3 - \mu_1 > 10$) resulted in biased estimates and poor coverage rates for the parameters estimating the intercept constant and random effect standard deviation. Severe misspecification (distances > 15) also resulted in biased estimates of the treatment effect and poor coverage rates of the time effect. The impact of misspecification in regards to bias, coverage rates and accuracy of model based standard errors were similar in terms of magnitude and trend for the three variance component scenarios. Similarly, misspecification in the presence of missing data due to MAR attrition was similar to that reported for the complete data. The results from this secondary simulation study are consistent with the results identified in the primary simulation study. This suggests that the impact of misspecification within a clinical trial context is similar to the results within a panel survey setting.

6.4.2 Simulation study within panel survey setting: considering fixed total random effect variance

One potential limitation of the primary simulation study is that the trends observed for the increasing variance components may be confounded by the consequential increasing total random effect variance. To investigate the potential impact, an additional simulation study considered the same panel survey setting motivated by the HILDA case study and the same random intercept distribution as considered by the primary study. However, by altering the component variances, this secondary simulation study fixes the total random effect variance at $\sigma_b^2 = 20, 25, 30$ and 35. Details regarding the simulation design are presented in Appendix F, though briefly, by using the same methodology described in this chapter, random intercepts were simulated as a symmetric three component mixture of normal distributions. However, the mean components were restricted to range from $\mu_3 = 0$ to 5 increasing in increments of 0.5. This restriction was imposed as higher values of μ_3 would generate higher σ_b^2 values than the 20 to 35 considered in this study. Simulations were performed under two missing data scenarios: complete data and incomplete data due to attrition. As in the primary simulation, attrition was assumed to be generated by the MAR mechanism with similar wave-to-wave attrition rates observed in the HILDA case study.

The results are described in further detail in Appendix F, though briefly, the results of the secondary simulation complement the results presented in the primary study (for comparable mean component distances). The simulated random intercepts considered in the secondary simulation were not as extreme as the multimodal distributions considered in the primary simulation (as shown in Figure F.1). Consequently, assuming normality had a minimal impact on the interpretation of the fitted model for small departures from normality. Generally the impact of misspecification was restricted to the terms associated with the misspecified random intercept distribution (i.e. β_0 and σ_b). With the exception of the intercept constant and the covariate relating to single women (β_3), misspecification resulted in unbiased estimation of the fixed effect parameters and σ_b. Excellent coverage rates and accurate model based standard errors were produced for all parameters, with the exception of σ_b. The magnitude and trends in the performance measures were similar for the four total random effect variances, with some deviations in the trend for coverage and standard error ratio for smaller total random effect variances, corresponding to the largest departures from normality. The impact of misspecification in the presence of missing data due to MAR attrition was similar to that in the complete data. Therefore, the negligible impact of incorrectly assuming normality identified in this additional simulation study is similar to minor to moderate deviations from the assumed normality considered in the primary simulation study. Thus, these results suggest that the varying total random intercept variance does not subsequently interact with the impact of misspecified random intercept distributions.

6.5 Discussion

The primary simulation study investigates the impact of incorrectly assuming normally distributed random effects in random intercept logistic models applied to panel data, when the true distribution is a symmetric three component mixture of normals. In practice the random effects of generalised linear mixed models are commonly assumed to be Gaussian distributed, however as identified in the HILDA case study (Chapter 4) multimodality may exist. By considering a range of finite mixture distributions with increasing component mean distances, this simulation study provides a novel insight into the impact of misspecifying symmetric multimodal distributions and identifying scenarios when misspecification impacts maximum likelihood estimation. This simulation study also investigates the additional impact of attrition within a panel survey application, by considering a similar attrition rate of 29.5% observed in the HILDA case study (Chapter 4).

The assumed normal distribution was robust to minor and moderate deviations of the true distribution from normality. However, as the true random intercept distribution departed substantially from normality (i.e. small component variances and larger mean distances), misspecification impacted maximum likelihood estimation of fixed and random effects. Misspecification resulted in biased estimates and poor coverage rates for parameters relating to the random effect distribution and time-invariant fixed effects, and produced inaccurate standard errors when estimating the random effect variability. Misspecification in the presence of MAR attrition resulted in similar magnitude and trends of bias, coverage and standard errors as for the complete data scenario.

The impact of incorrectly assuming normal random intercepts when the true random intercept is a symmetric three component mixture of normal distributions generally resulted in minimal impact on the fixed effect coefficients. For larger departures from the assumed normal distribution (component mean distances exceeding 12), misspecifying the random intercept distribution resulted in biased estimates and poor coverage rates for the intercept constant (β_0) and for parameters of the time-invariant explanatory variables β_4 and β_5. These results are consistent with previous literature, whereby estimation of the intercept constant and time-invariant covariate effects is vulnerable to distributional misspecification of the random intercept (Neuhaus et al., 1992; McCulloch and Neuhaus, 2011a). Modest bias and loss of efficiency of time-invariant covariates have been reported for true distributions far from the assumed distribution (Agresti et al., 2004; Litiére et al., 2007, 2008) and large random effect variability (Litiére et al., 2008). With the exception of β_3, estimation of the time-varying explanatory variables ($\beta_1, \beta_2, \beta_3, \beta_6$ and β_7) was generally robust to random effect misspecification, as reported elsewhere (Neuhaus et al., 1992; Heagerty and Kurland, 2001; McCulloch and Neuhaus, 2011a). The parameter β_3 was consistently underestimated and negatively biased, however the bias may partly be explained by the large standard error of the true coefficient and the small magnitude of the true coefficient. Additionally, the bias may partly be explained by the stable nature of the explanatory variable, such that it is almost time-invariant. Of the 1396 women with Complete Case data, only 71 (5.1%) of women transitioned at least once from being single into a different marital status category (of whom 52 only transition once). Thus, even if the explanatory variable capturing women being single or never married is time-varying, the within-subject variability in the explanatory variable may influence the impact of incorrectly assuming normality.

Estimation and inference of the variability of the random intercept (σ_b) were severely impacted by misspecifying the assumed random effect distribution. For moderate to severe departures from normality characterised by distinct multimodality (component mean distances exceeding 8), assuming normal random intercepts resulted in seriously biased estimates of the variance component σ_b with extremely poor coverage rates and inaccurate model based standard errors. These findings are consistent with previous literature, suggesting that estimation of the variance component is sensitive to misspecification (i.e. Litiére et al. 2008; McCulloch and Neuhaus 2011a). Furthermore, as reported by Litiére et al. (2008), larger true random intercept variances (corresponding to the larger component mean distances) resulted in larger bias in the estimation of the variance component. The accuracy of the model based standard errors for σ_b fluctuated above and below the acceptable limits. These fluctuations coincide with the varying and skewed sampling distributions of the model based standard errors and the exponentially increasing empirical standard error estimates (results not shown). Arguably estimates of the variance component are often not of primary inferential interest (McCulloch and Neuhaus, 2011a). However it is the only estimate of the true random effect variability (Litiére et al., 2008) and is commonly used to estimate alternative summary measures of the unobserved between-subject variability such as the intra-class correlation[3]. Biased estimation of the variance components can subsequently impact the prediction and accuracy of best predicted random effect values (McCulloch and Neuhaus, 2011b). These results suggest that if the random intercept distribution is suspected to be multimodal with moderate to extreme departures from normality, caution should be exercised in interpreting estimates of the random effect variability or complementary summary measures.

[3]However, as detailed in Section 3.1.1, the variance of the residual error in a random intercept logistic model is fixed to $\frac{\pi^2}{3}$. Therefore, the bias in the ICC is expected to be minimal unless the true random effect variability is small and the bias of the variance component is large. For the scenarios considered here, the bias of the ICC is between 0% and 5% (i.e. for $\sigma^2 = 1$: ICC bias= 2%, 0%, 3%, 5% and 4% for component mean distances of 0, 5, 10, 15 and 20).

In the presence of MAR attrition, the impact of misspecifying the random intercept distribution resulted in similar magnitudes and trends in the performance measures as observed in the Complete Case data scenario. The minimal additional impact of MAR attrition is similar to the results observed for the asymmetrical random effects considered in Chapter 5. These results are consistent with the findings of Wang (2010), whereby there was no interaction between misspecified bivariate random effect distribution and attrition on the bias or mean-squared error of the estimated parameters and standard errors. Unlike Wang (2010) who considered two missingness patterns (intermittent missing and attrition), two missing mechanisms (MCAR and MAR) and three missingness rates (10%, 20% and 30%), this simulation study has only considered one attrition rate, one missingness pattern and one missing mechanism. Further research considering other attrition rates, types and mechanisms of missingness is required, particularly in the potentially realistic MNAR scenario.

Previous simulation studies investigating the impact of misspecification have considered a relatively small number of true random effect distributions, predominately focused on unimodal or bimodal distributions. Furthermore, no previous literature has considered a variety of the same family of distributions ranging in departures from normality. The simulation study presented in this chapter has focused on symmetrical three component mixture of normal distributions, considering 21 different mean components and three component variances. Some of the scenarios considered were extreme, with simulated random intercepts having total random effect variances that may be unrealistic in practice (McCulloch and Neuhaus, 2011a). Furthermore, the restriction of equal mixing proportions and component variances used to generate the symmetrical random intercept distributions are potentially unrealistic representation of underlying heterogeneity. However, the aim was to identify scenarios where misspecification of the random intercept distribution impacted interpretation. Maximum likelihood estimates were impacted for component mean distances exceeding 10, which corresponds to one of the first scenarios where the true random effect is distinctly multimodal. Consistent with previous findings, as the departure from the assumed normal distribution becomes more severe, the greater the impact on inference. This was particularly true for parameters related to the misspecified random effect and the time-invariant explanatory variables.

As for the asymmetrical multimodal distribution simulated in Chapter 5, incorrectly assuming normality when the random intercepts were a symmetrical three component mixture of normals predominately impacted parameters associated with the random effect distribution. For random intercepts with similar total variances to the asymmetrical random intercept distribution identified in the motivating example in Chapter 4 and considered in Chapter 5, symmetric distribution with component means distances of 7 to 9, resulted in similar impact as the asymmetrical mixture of normals simulated in Chapter 5. That is, marginally biased estimation of β_0 and β_3, and poor coverage rates and inaccurate standard errors when estimating σ_b. Therefore, regardless of the symmetric or asymmetric multimodal distributions, incorrectly assuming normal random intercepts when the random intercept was trimodal with total random effect variances ranging between 9 to 18 (see Table 6.1 corresponding to $\mu_3 - \mu_1$ in the range 7 to 9) resulted in similar inferential impact as observed in Chapter 5.

As detailed in Chapter 2, literature assessing the impact of misspecified random effect distributions has predominately been within applications to clinical trials and biomedical settings (e.g. Litiére et al. 2008; Neuhaus and McCulloch 2011; Heagerty and Kurland 2001). Within the biomedical settings, the simulation studies generally consider a small number of explanatory variables, primarily focusing on the intercept, a time effect and a treatment effect (e.g. Litiére et al. 2008). More explanatory variables including within- and between-cluster covariates (Neuhaus and McCulloch, 2011; Heagerty and Kurland, 2001) and interactions (Heagerty and Kurland, 2001) have been considered. However there has been limited literature considering situations within panel survey data, whereby adjustments for numerous explanatory variables would often be included in GLMMs. The simulation studies presented in Chapter 5 and in this chapter provide a novel insight into the impact of misspecified random intercept distribution in these more realistic scenarios of including time-varying and time-constant explanatory variables within a panel survey setting.

The results of simulation studies may be restricted to the particular setting imposed. To ensure that the results observed in the primary simulation study were not restricted to the HILDA case study, an additional simulation study within a clinical trial setting was considered as detailed in Appendix E. Albeit the smaller number of explanatory variables, and hence, variability of the random intercepts, the impact of misspecification within a clinical trial context was similar to the results within a panel survey setting (Appendix E). Furthermore, the primary simulation study has a large number of varying factors, including the differing total random effect variance for all scenarios (as highlighted in Table 6.1). The differing total random effect variance may subsequently impact the results, as comparisons may be confounded by the increasing total random effect variance. However, the additional secondary simulation study presented in Appendix F with fixed total variances ($\sigma_b^2 = 20, 25, 30$ and 35) suggests that the varying total random intercept variance in the primary study does not interact with the impact of incorrectly assuming normality.

In summary, the results from the primary simulation study suggest that the assumed normality in random intercept logistic models is robust to minor deviations from the shape of normality. However, when the true distribution is a symmetrical three component mixture of normals, severe departures from the assumed normal random intercept distribution can impact inference of maximum likelihood. Particularly when multimodality of the true random intercept can be observed in the form of distinct modes, distributional misspecification resulted in bias, lower coverage rates and inaccurate model based standard errors. The impact of misspecification was generally restricted to parameters associated with the random effect distribution and time-invariant explanatory variables. Unlike the bias and poor coverage rates produced for the effects of time-invariant explanatory variables, estimation of the effects of time-varying explanatory variables was generally robust to misspecification. However, even if the explanatory variable is time-varying, it appears that the degree of within-subject variability of the explanatory variable can also influence the impact of misspecification (i.e. bias observed for the parameter capturing the effect of single women compared to married women, β_3). The primary simulation study considered a variety of true distributions, with varying component mean distances and component variances. The different component variances ($\sigma^2 = 1, 2, 4$) produced similar results, with the most severe impact of misspecification corresponding to the most extreme distributions (i.e. large component mean distances and smaller component variances). As observed in Chapter 5, the presence of MAR attrition resulted in minimal additional impact on maximum likelihood estimation of the model parameters. Estimates of the fixed effects parameters, typically the parameters of interest, were generally robust to small or moderate departures from the assumed normal distribution. However, for time-invariant explanatory variables and parameters related to the misspecified random effect, caution in regards to inferential conclusions should be exercised when the random intercept distribution is suspected to be trimodal with distinct modes.

The impact of potential misspecification of the random effects distribution can be minimised by increasing the flexibility of modelling the random effects. Flexible methods do not directly assess misspecification, though they can be used as a form of sensitivity analyses to help identify potential misspecification (Agresti et al., 2004). Approaches relaxing the parametric assumption of the random effects in GLMMs include using finite mixtures of normal distributions (Magder and Zeger, 1996; Verbeke and Lesaffre, 1996) or penalised Gaussian mixtures (Komarek and Lesaffre, 2008). Alternatively, the random effects can be approximated using non-parametric (Laird, 1978; Heckman and Singer, 1984; Aitkin, 1999; Lesperance et al., 2014) or semi-parametric techniques (Chen et al., 2002; Vock et al., 2014). In the next Chapter, a novel application of the Vertex Exchange Method (VEM) is implemented to flexibly model the random effects distribution. The proposed method utilises non-parametric maximum likelihood to simultaneously estimate the random effects and the coefficients of the fixed effects in logistic mixed models. The performance of the proposed method is assessed by applying the methodology to the HILDA case study considered in Chapter 4, and will be compared to a selection of existing flexible random effect methods within a sensitivity analysis framework. The sensitivity analyses presented in Chapter 7 will provide a novel comparison and assess the practicality of flexibly modelling the random effect distribution in logistic mixed effect models applied to panel survey data, including in the presence of attrition.

7 Vertex Exchange Method to flexibly model random effect distributions in logistic mixed models

7.1 Introduction

In many practical applications the assumption of Gaussian distributed random effects in GLMMs may be too restrictive and may not appropriately capture the latent heterogeneity (Vock et al., 2014). The application of the random intercept logistic model in Chapter 4 provides an example whereby the assumed normal distribution may not adequately capture the heterogeneity in a potential underlying mover-stayer scenario. As shown in Chapters 5 and 6, maximum likelihood estimation of GLMMs may not be consistent if the random effect density is substantially misspecified. Estimation of intercept constant and variance components was consistently biased, had poor coverage rates and, for the variance components, had inaccurate standard errors. This can subsequently impact inference of the magnitude of the between-individual variability of not only the individuals in the HILDA analysis, but also in the population of working aged women in Australia (McCulloch et al., 2008). Furthermore, it can impact the efficiency of individual-specific predictions of response profiles (McCulloch and Neuhaus, 2011b). For large departures from normality, inference of parameters capturing the effects of time-invariant explanatory variables was impacted by biased estimation and poor coverage rates. To help guard against the impact of misspecifying the random effect distribution, the parametric normality assumption can be relaxed by increasing flexibility of the assumed distribution. In the application considered in Chapter 4, modelling the random intercepts as a three component mixture of normal distributions provided a better fit than the assumed normal distribution. However, it is unclear whether three components sufficiently captures the potential latent mover-stayer scenario. Finite mixtures of normal distributions are extremely flexible, however, other methods may more adequately capture the potentially extreme underlying heterogeneity of the latent mover-stayer scenario.

As reviewed in Section 2.7.1, a suite of methodology is available to relax the assumption of Gaussian distributed random effects in GLMMs. Efficient methods have been developed to fit GLMMs with flexible, parametric classes of random effect densities including, the class of t-distributions (Lee and Thompson, 2008) and skew extensions of the t-distribution or normal distributions (Ho and Lin, 2010). However, the parametric classes are generally not flexible enough to capture multimodal densities, and the potential sub-population structure of the mover-stayer scenario. Alternatively, flexibility may be achieved by assuming that the random effects belong to a smooth class of densities represented by the semi-non-parametric (SNP) formulation of Gallant and Nychka (1987). Approaches that utilise the SNP class of densities to fit random effects in GLMMs (Chen et al., 2002; Vock et al., 2014) are sufficiently flexible to capture a range of densities including skewed, multimodal, and thick- or thin-tailed densities, and hence, may be appropriate to capture underlying sub-populations. An approach that allows a large degree of flexibility is to leave the random effects distribution completely unspecified by modelling the random effects non-parametrically. Computational approaches to obtain the non-parametric maximum likelihood (NPML) estimator of the random effects distribution have been proposed (Laird, 1978; Aitkin, 1999; Follmann and Lambert, 1989; Lesperance and Kalbfleisch, 1992; Heckman and Singer, 1984; Rabe-Hesketh et al., 2003; Wang, 2010a), and result in a discrete distribution on a finite number of support points (Lindsay, 1983). To overcome the discreteness of NPML estimators, smooth non-parametric maximum likelihood (SNPML) estimators resulting in a continuous density have been proposed, whereby the smoothing is obtained using finite mixtures of Gaussian distributions (Magder and Zeger, 1996), kernel methods (Knott and Tzamourani, 2007) or methods using a predictive recursive algorithm (Tao et al., 1999), however, the degree of smoothness is often arbitrary.

As highlighted above, many methods are available to increase the flexibility of the assumed random effect distribution. However the performance of these methods within the context of capturing the latent heterogeneity of potential mover-stayer scenarios is limited. The choice of the appropriate approach may be dependent on the goal of analysis (Vock et al., 2014). For example, a commonly stated disadvantage of non-parametric approaches is that the discrete approximation with a finite number of support points may not provide adequate insight into the true data generating mechanism (Vock et al., 2014). However, estimation of the random effect density may not be of primary interest, in which case non-parametric approaches may be appropriate (Vock et al., 2014). The primary aim of this study is to induce sufficient flexibility of the assumed random effects distribution to help protect against potential distributional misspecification of the random effects, rather than estimate the random effects distribution. This chapter investigates NPML estimation in GLMMs as an appropriate modelling strategy for potential mover-stayer scenarios in panel survey data, considering both univariate and bivariate random effects.

The aim of NPML estimation in GLMMs is to simultaneously find maximum likelihood estimates of the parameter coefficients and the random effect distribution, i.e., the number of support points, and their corresponding location and probability weights. To obtain the NPML estimator of the random effects distribution, the optimal number of support points needs to be determined. One approach is to start with a large grid of support points and within an iterative procedure either merge or omit support points as determined by directional derivative-based algorithms[1] (Butler and Louis, 1992; Lesperance and Kalbfleisch, 1992; Tsonaka et al., 2009; Baghfalaki and Ganjali, 2014; Lesperance et al., 2014). Alternatively, approaches can start with a single support point and subsequently incorporate additional support points one by one as determined by the directional derivative (Heckman and Singer, 1984; Follmann and Lambert, 1989; Rabe-Hesketh et al., 2003). The EM algorithm is often used to maximise the likelihood for a given number of support points, either within the iterative procedure (Follmann and Lambert, 1989), or considering the number of support points as unknown but fixed, and fitting a number of models with varying numbers of support points (Aitkin, 1999).

A limited number of non-parametric computational approaches are available to fit GLMMs with multiple random effects. Lesperance et al. (2014) propose an algorithm to compute the NPML estimate of logistic mixed models with bivariate random effects. Their proposed Direct Search Directional Derivative (DSDD) approach uses a direct search method (Torczon, 1991) to identify maxima of the gradient function to include as support points, and estimates the mixing proportions using the constrained Newton method proposed by Wang (2007). Within a similar context to GLMMs, Tsonaka et al. (2009) applied the Vertex Exchange Method (VEM) of Böhning (1985) to estimate the unspecified distribution of the multivariate random effects shared between the response and missingness sub-models in a shared parameter model. The VEM is a directional derivative based-algorithm that, for a fixed grid of support points, exchanges probability weight from a 'bad' support to a 'good' support in an iterative process. In the context of linear mixed models with bivariate random effects, Baghfalaki and Ganjali (2014) recently proposed an algorithm that is faster than the traditional VEM by updating the probability weights for all grid points in each iteration. An advantage of the VEM approach is the simplicity and applicability in the specification of the random effects distribution (Baghfalaki and Ganjali, 2014). However, application of the traditional VEM to flexibly model multiple random effects in GLMMs has yet to be utilised.

In this chapter, we apply the VEM (Böhning, 1985) to fit univariate and bivariate random effects in binary logistic mixed effects models. We consider the performance of flexibly modelling the random effects non-parametrically when applied to panel survey data. The VEM is applied to the HILDA case study considered in Chapter 4 to analyse women's employment participation over eleven waves of the HILDA survey. We extend the random intercept logistic model applied in Chapter 4 to the bivariate random effect scenario, considering both random intercepts and random slopes.

To assess the performance of the VEM, we compare the fit of the VEM approach to five alternative flexible approaches currently available and implemented in standard software. These comparisons within a sensitivity framework provide evidence of the practicality of available methods when applied to panel survey data, in addition to potential mover-stayer scenarios.

[1] The directional derivative has also been referred to as the Gateaux derivative by Heckman and Singer (1984), and subsequently by Rabe-Hesketh et al. (2003).

As highlighted in Chapter 4, the focus is not to quantify or address the potential latent mover-stayer scenario. The focus is to assess the appropriateness of flexibly modelling the assumed random effects distribution in panel data to relax the parametric assumption and help protect against the impact of misspecifying the random effect density.

7.2 Non-parametric estimation of the logistic mixed effects model

This section focuses on non-parametric estimation of the logistic mixed model, whereby the distribution of the random effects is left completely unspecified. Specifically, non-parametric estimation of this model assumes that the random effects distribution is discrete with unknown number of support points. This section details an approach using the Vertex Exchange Method of Böhning (1985) to maximise the log-likelihood with respect to the random effects distribution. It begins by introducing the statistical framework (Section 7.2.1) and describing non-parametric estimation of the random effects distribution in GLMMs (Section 7.2.2). This is followed by Section 7.2.3, where the optimisation procedure to estimate the random effects distribution using the Vertex Exchange Method is outlined.

7.2.1 Statistical framework

Following the statistical notation presented in Section 3.1, let y_{ij} denote the response for individual $i(i = 1, ..., N)$ at time $j(j = 1, ...n_i)$. We restrict our attention to the two-level logistic mixed effects model and consider an alternative parameterisation to the model considered in Section 3.1, such that,

$$\text{logit}(\Pr(y_{ij} = 1)) = \boldsymbol{x_{ij}}'\boldsymbol{\beta^{(1)}} + \boldsymbol{z_{ij}}'(\boldsymbol{\beta^{(2)}} + \boldsymbol{b_i}) \tag{7.1}$$

where $\boldsymbol{x_{ij}}$ is j^{th} row of the design matrix $\boldsymbol{x_i}$ for the f covariates with no random effect component, $\boldsymbol{z_{ij}}$ is the j^{th} row of the design matrix $\boldsymbol{z_i}$ for the q covariates with both fixed and random components, and $\boldsymbol{b_i}$ is the q-dimensional vector of random effects. In this parameterisation of the model, the $\boldsymbol{\beta^{(1)}}$ and $\boldsymbol{\beta^{(2)}}$ denote the $f-$ and q-dimensional vector of regression coefficients corresponding to $\boldsymbol{x_i}$ and $\boldsymbol{z_i}$ respectively. This alternative parameterisation enables the random effects distribution not to be restricted to have zero mean, that will subsequently be utilised in the estimation of $\boldsymbol{\beta^{(2)}}$.

7.2.2 Non-parametric estimation

Non-parametric estimation of a GLMM implies no parametric distributional assumptions are made for the random effects. Hence, the $\boldsymbol{b_i}$ are completely unspecified and it is assumed that $\boldsymbol{b_i} \sim G$ with $G \in \Omega_M$, where Ω_M is the set of all distributional functions on the parameter space M of $\boldsymbol{b_i}$. The marginal density of $\boldsymbol{y_i}$ is given by

$$f(\boldsymbol{y_i}|G, \boldsymbol{\theta}) = \int_{\Omega_M} f(\boldsymbol{y_i}|\boldsymbol{b_i}, \boldsymbol{\theta}) dG(\boldsymbol{b_i}) \tag{7.2}$$

where $\boldsymbol{\theta} = (\boldsymbol{\beta^{(1)}}, \boldsymbol{\beta^{(2)}})$ is the parameter vector. The random effect distribution G can be either a continuous or discrete distribution. However, it has been shown that the non-parametric maximum likelihood estimate (NPMLE) of the unspecified G is a discrete distribution with finite support (i.e. at most $C \leq N$ support points) (Laird, 1978; Lindsay, 1983). Therefore, G is a discrete distribution and thus, Ω_M is reduced to a set of discrete distributions such that the marginal density is given by

$$f(\boldsymbol{y_i}|G, \boldsymbol{\theta}) = \sum_{c=1}^{C} \pi_c f(\boldsymbol{y_i}|\boldsymbol{\mu_c}, \boldsymbol{\theta}) \text{ for } C \leq N \tag{7.3}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu_1}, ..., \boldsymbol{\mu_C})$ are the support points in the $q-$dimensions and $\boldsymbol{\pi} = (\pi_1, ..., \pi_C)$ are the corresponding weights of the discrete distribution, G.

7.2.3 Non-parametric estimation using the Vertex Exchange Method to estimate the random effects distribution

The following optimisation procedure uses the Vertex Exchange Method of Böhning (1985) to non-parametrically estimate the unspecified random effects distribution in a GLMM. The optimisation procedure was originally described by Tsonaka et al. (2009) to estimate shared parameter models with unspecified random effects, and as detailed further below, this study adapts the procedure to estimate the random effects distribution in a logistic mixed model. The optimisation requires an iterative two step procedure. In the first step, for $\boldsymbol{\theta}$ fixed at its current estimate ($\hat{\boldsymbol{\theta}}$), G is estimated using the Vertex Exchange Method. In the second step, $\hat{\boldsymbol{\theta}}$ is updated by maximising the profile likelihood at the estimated $\hat{G}$ from the first step by using the Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm (Nocedal and Wright, 2006). These two steps are repeated in an iterative process until convergence is reached.

The Vertex Exchange Method (VEM) algorithm is detailed in Section 3.2.3.2, and will be briefly described here. Consider the random effects to be located within a fixed q-dimensional grid of support points, $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_C$ (for $C \leq N$), such that each of the q dimensions of the random effects consists of K equally spaced and equally weighted support points (i.e $C = K^q$). The VEM algorithm is based on the directional derivative to iteratively exchange probability weight from a 'bad' support point to a 'good' support point. The grid of support points, and thus the locations $\boldsymbol{\mu}_1, ..., \boldsymbol{\mu}_C$, are kept fixed throughout the estimation procedure, such that only the corresponding weights $\pi_1, ..., \pi_C$ are updated. Originally, the VEM was utilised within a two phase procedure to obtain the NPMLE (Böhning, 1999), whereby the locations of the VEM estimate (Phase 1) were updated using the EM algorithm (Phase 2). However, as the EM algorithm has minimal additional improvement to the model fit (Böhning, 1999), the random effects distribution can be estimated using only the VEM step (Tsonaka et al., 2009). By considering a very dense grid of support points for $\boldsymbol{\mu}$, the resulting estimated distribution derived by the VEM will provide an approximate NPMLE of G (Tsonaka et al., 2009).

Therefore, the choice of the range of the grid and the number of support points may impact the performance of the VEM algorithm to adequately approximate the NPMLE. Following the methodology of Tsonaka et al. (2009), the range of the grid is based on the scaled random effect, $\boldsymbol{b_i}^* = \hat{\boldsymbol{S}}_b^{-1}\boldsymbol{b_i}$, where $\hat{\boldsymbol{S}}_b$ is the estimated Cholesky decomposition of the random effects covariance matrix corresponding to the GLMM assuming q-dimensional multivariate normally distributed random effects. The q-dimensional grid for $\boldsymbol{b_i}^*$ is defined by $[-v, v]^q$, where v is selected to be large enough not to be affected by boundary issues (i.e. such that the resulting solution does not have support points located at $-v$ and/or v in any of the q-dimensions). Tsonaka et al. (2009) suggest that v values of 4 or 5 should be sufficient, however, if the resulting solution has support at the boundaries then the models should be refit with larger values of v (i.e. a wider range). Empirical Bayes predictions of the random effects were calculated as the grid-point corresponding to the maximum posterior probability, i.e. as the posterior mode.

Implementation of the estimation procedure requires initial values for the parameters ($\boldsymbol{\theta^0}$) and the weights of the support points ($\boldsymbol{\pi}^0$) to be specified. As suggested by Tsonaka et al. (2009), the initial values for the model coefficients, $\boldsymbol{\theta^0} = (\boldsymbol{\beta}^{0(1)}, \boldsymbol{\beta}^{0(2)})$ are set to be the estimated coefficients corresponding to a logistic mixed effects model assuming $q-$dimensional multivariate normal random effects. As previously mentioned, the initial grid for the random effects is fixed and assumes that the support points are equally spaced and equally weighted. Therefore, the weights for each of the C support points, π_c^0, are set to $1/C$. In regards to the second step of the optimisation procedure, to avoid identifiability issues, estimation of $\boldsymbol{\theta}$ in the second step is restricted to parameters with no random effect component, i.e. $\boldsymbol{\beta}^{(1)}$. By utilising the constraint that the mean of the random effects is zero (i.e. $\sum_{c=1}^{C} \pi_c \boldsymbol{\mu}_c = \boldsymbol{0}$), the coefficients of $\boldsymbol{\beta}^{(2)}$ are fixed at the initial values ($\boldsymbol{\beta}^{0(2)}$) during the optimisation procedure and after model convergence are updated by $\hat{\boldsymbol{\beta}}^{(2)} = \boldsymbol{\beta}^{0(2)} + \sum_{c=1}^{C} \hat{\pi}_c \boldsymbol{\mu}'_c$.

7.3 Statistical models and flexible random effect methodology applied to the HILDA case study

The VEM approach to estimate random effect logistic models with an unspecified random effect distribution is illustrated in an application to analyse employment participation in the HILDA panel survey. To demonstrate the applicability of the approach, the VEM algorithm is applied to estimate univariate and bivariate random effect logistic models. The performance of the VEM approach is compared to other approaches to flexibly model the random effects. Details of the application and the statistical methodology of alternative approaches applied to logistic models with univariate and bivariate random effects are presented in Sections 7.3.1 and 7.3.2, respectively.

7.3.1 Random intercept logistic model

The VEM approach was applied to the case study considered in Chapter 4. The same random intercept logistic model as detailed in Equation 4.1 was used to analyse employment participation of working aged women over 11 waves of the HILDA survey.

The model in Equation 4.1 was fitted assuming an unspecified random intercept distribution and estimated using the VEM approach. Initial values for the coefficient estimates (both $\beta^{(1)}$ and $\beta^{(2)}$) were obtained from a random intercept logistic model assuming normally distributed random effects. The initial grid of support points consisted of 301 grid-points with the range based on the Cholesky decomposition of ± 5 standard deviations of the assumed normal density (i.e. $b^*_{0i} \in [-5, 5]$). The VEM procedure reaches convergence when the stopping criteria for the log-likelihood function (Section 3.2.3.2) between two consecutive iterations less than 10^{-7}, and when the maximum directional derivative over the grid of support points is less than 10^{-3} (Tsonaka et al., 2009). The VEM algorithm was implemented in R (R Version 3.0.2) using syntax developed by Dr. Tsonaka (see Section 3.2.3.2 for details).

As part of assessing the performance of the proposed VEM approach, inference for the VEM model was compared to the following methods: (i) logistic model assuming normal random intercepts, (ii) logistic model assuming three component mixture of normal distributions, (iii) the semi-non-parametric non-linear mixed model (SNP-NLMM) of Vock et al. (2014), (iv) the logistic model with endpoints approach of Berridge and Crouchley (2011b) and (v) non-parametric maximum likelihood estimation (NPMLE) derived using the Gateaux method (NPMLE-Gateaux) as described by Rabe-Hesketh et al. (2003). The same random intercept logistic model in Equation 4.1 was estimated by the aforementioned approaches.

The methods (ii) to (v) relax the Gaussian assumption into more general and flexible distributions, or in the case of the VEM and the NPMLE-Gateaux leave the random effect distribution completely unspecified. The methodologies of the flexible random effect approaches are detailed in Section 3.2. Briefly, the SNP-NLMM approach assumes the random effects follow a smooth density that can be represented by the semi-non-parametric method by Gallant and Nychka (1987). The SNP representation does not cover all continuous densities, however it is flexible enough to capture a variety of distributions including multimodal distributions (Vock et al., 2014). The logistic model with endpoints has been developed by Berridge and Crouchley (2011b) to identify stayers in a latent mover-stayer model. The logistic model assumes that the latent stayers can be represented by random intercept values of negative and positive infinity, whilst the movers are assumed to have normal distributed random intercepts. The latent mover-stayer model is not restricted to the two types of stayers (i.e. always staying in $y = 0$ or always staying in $y = 1$), and can also account for only one type of stayer by incorporating one endpoint (at either $-\infty$ or $+\infty$ respectively). The NPML estimate derived by the Gateaux method, is a non-parametric approach that utilises the directional derivative to increase the number of support points one at a time until the largest maximized likelihood is achieved (Rabe-Hesketh et al., 2003).

The assumed normal distribution and three component mixture of normal distributions for the random intercepts were implemented in SAS (Version 9.4, SAS Institute, Cary NC) using adaptive quadrature. As detailed in Section 4.3, random intercept logistic models assuming normality were estimated using 20 quadrature points, and the model assuming three component mixture of normals was estimated using 54 or 61 quadrature points for the women with complete case or monotone missing data, respectively. The SNP NLMM method was implemented in SAS using the syntax developed by Vock et al. (2014). The SNP NLMM was fit using adaptive Gaussian quadrature and 51 quadrature points, with the random effect density estimated based on 20 grid points. As suggested by Vock et al. (2014) and detailed in Section 3.2.2, a maximum smoothness parameter of four knots ($K_{max} = 4$) was selected for the univariate random effects. The number of knots corresponding to the optimal SNP model (i.e. either $K = 0, 1, 2, 3$ or 4), was selected using the Akaike information criterion (AIC). The logistic models with one or two endpoints at negative and/or positive infinity were implemented in the R-package sabreR (Crouchley, 2007) (R Version 2.08). The logistic model with one endpoint was estimated based on non-adaptive quadrature and 24 quadrature points, and the logistic model with two endpoints was based on adaptive quadrature and 6 quadrature points. The NPMLE-Gateaux was implemented in STATA (Version 13.1, StataCorp, College Station TX) using the GLLAMM procedure, with the initial model having one support point. At the completion of each model fit with M support points, the Gateaux method assessed for additional support points within a grid of 30 points ranging between ± 5 standard deviations of the assumed normal random intercept distribution. If the directional derivative exceeded 10^{-5} at any location within the specified grid, a new support point was introduced (Rabe-Hesketh et al., 2003). This was estimated using the previous model estimates as initial starting values and the additional support point at the location corresponding to the greatest increase in likelihood. All analyses were performed for both subgroups of women: the 1359 women with complete case data (Complete Cases) and the 1927 women with monotone missing data (Monotone Missing).

7.3.2 Random intercept and random slope logistic model

In many practical applications, GLMMs with only a random intercept may be too simple to appropriately capture the heterogeneity. For longitudinal data it is often appropriate to consider a GLMM with random intercepts and random slopes. To investigate the practicality of using the VEM approach to estimate bivariate random effects, a logistic mixed model with random intercepts and a random time effect is considered to model employment participation in the HILDA case study.

Extending Equation 4.1 to incorporate a random intercept and a random slope requires the time-varying age covariate in Equation 4.1 to be separated into two components: the time-invariant term, age at first wave, and a time-varying term, wave. Therefore, the time-effect considered in this model is the wave of the HILDA panel survey. To aid interpretation of initial time, the two age components were transformed to begin at zero. The transformations were obtained by subtracting 1 from the wave of the HILDA survey and by subtracting 30 (the minimum age at baseline) from the reported age at first wave, such that $Wave_{ij} = 0, 1, 2, ..., 10$ and $AgeBaseline_i$ has values between 0 and 15. Furthermore, to avoid numerical and convergence issues, and to ensure the coefficients of the time components were of similar magnitude to the other model coefficients, the time components were again transformed by dividing each term by 10 such that $W_{ij} = Wave_{ij}/10$ and $A_i = AgeBaseline_i/10$. The following logistic mixed model is used to model employment participation of working aged women over the 11 waves of HILDA:

$$\begin{aligned}\text{logit}[\Pr(y_{ij} = 1|\boldsymbol{b_i})] = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})W_{ij} + \beta_A A_i + \beta_2 x_{2ij} + \\ \beta_3 x_{3ij} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6ij} + \beta_7 x_{7ij}\end{aligned} \tag{7.4}$$

where b_{0i} and β_0 are the random intercept and fixed constant coefficient, b_{1i} and β_1 are the random slope and fixed slope coefficient corresponding to the wave term (W_{ij}), β_A is the coefficient relating to the age at baseline covariate (A_i), and β_2 to β_7 are the corresponding parameter coefficients of the fixed parameters x_{2ij} to x_{7ij} as defined in Section 4.2.

The model in Equation 7.4 was fitted to the HILDA case study assuming the distribution of the random effects $\boldsymbol{b_i} = (b_{0i}, b_{1i})'$ was completely unspecified. Estimation of the model was obtained using the VEM algorithm with the initial grid of support points for the random effects consisting of 31 grid-points in each dimension and the range based on the scaled random effect $\boldsymbol{b_i}^* \in [-5, 5] \times [-5, 5]$ using the Cholesky decomposition of the corresponding model assuming bivariate normal random effects ($\hat{\boldsymbol{S}}_{\boldsymbol{b}}$). The VEM procedure reaches convergence when the stopping criteria for the log-likelihood function (Section 3.2.3.2) between two consecutive iterations less than 10^{-7}, and when the maximum directional derivative over the grid of support points is less than 10^{-3} (Tsonaka et al., 2009). The VEM algorithm was implemented in R (R Version 3.0.2) using syntax that has extended the work from Dr. Tsonaka to estimate bivariate distributions (see Section 3.2.3.2 for details).

The performance of the VEM approach to estimate the logistic mixed model with bivariate random effects was compared to inferences for the parameters in the equivalent logistic mixed model estimated using other approaches to flexibly model the random effects distribution. However, not all the flexible approaches considered in the random intercept logistic model extend to bivariate random effects. The logistic model with endpoints (Berridge and Crouchley, 2011b) is only applicable to the univariate random effect scenario. Therefore the performance of the VEM approach is compared to the following logistic mixed effect models assuming the random effects are distributed as a: (i) bivariate normal, (ii) three component mixture of bivariate Gaussian distributions[2], (iii) bivariate SNP density, and (iv) unspecified bivariate distribution with the NPMLE derived by the Gateaux method. The same random intercept and random slope logistic model in Equation 7.4 is estimated by these four approaches.

The models assuming bivariate normal distributed random effects were estimated using the SAS NLMIXED procedure with adaptive Gaussian quadrature and 20 quadrature points. The logistic mixed models based on the assumption that the random intercepts and random slopes were distributed as a three component mixture of bivariate normals were estimated using the likelihood reformulation method (Liu and Yu, 2008) implemented using SAS procedure NLMIXED. The logistic mixed models assuming that the random effects follow a bivariate SNP distribution were estimated using the SNP NLMM macro (Vock et al., 2014) implemented in SAS. The SNP approach was estimated using adaptive Gaussian quadrature with 21 quadrature points, with the bivariate SNP density estimated based on 15 grid points in each dimension. As suggested by Vock et al. (2014) a maximum smoothness parameter of two knots ($K_{max} = 2$) was selected for the bivariate random effects, and the number of knots, corresponding to the optimal SNP model, was selected using AIC. The NPMLE-Gateaux method used to obtain the NPML estimator of the bivariate random effects distribution was implemented in STATA using the GLLAMM procedure. The initial model started with one support point in the two-dimensional space. Additional support points were assessed using the Gateaux method within a grid of 60 points ranging from -30 to 30 in each dimension (approximately equivalent to the range based on $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$). A new support point was included if the directional derivative exceeded 10^{-5} at any location. The subsequent model was estimated using the previous model estimates as initial starting values and an additional support point at the location corresponding to the greatest increase in the likelihood. All analyses were performed for the two sub-analysis groups of women: the 1359 women with complete case data (Complete Cases) and the 1927 women with monotone missing data (Monotone Missing).

As highlighted in Chapter 4, due to the limited number of explanatory variables considered in this application, the random intercept and random slope logistic model is not appropriate to address questions about employment transitions in Australian working aged women. Additionally, more appropriate analyses would distinguish between full-time and part-time employment, and also unemployment and not in the labour force. This application serves as an example to demonstrate the applicability and practicality of the VEM algorithm and other existing approaches when applied to panel survey data in the presence of a potential underlying mover-stayer scenario. The explanatory variables considered in the random intercept and random slope model are the same as the explanatory variables considered in the random intercept model (Equation 4.1), enabling inferential comparisons between the univariate and bivariate logistic mixed models.

[2]The choice of three components for the mixture of bivariate of normals was motivated by the three components considered in the random intercept logistic model, relevant to the potential underlying mover-stayer scenario.

7.4 Results

The results of the VEM and alternative approaches applied to the HILDA panel survey application are presented in Section 7.4.1 for the random intercept logistic model, and Section 7.4.2 for the random intercept and random slope logistic model.

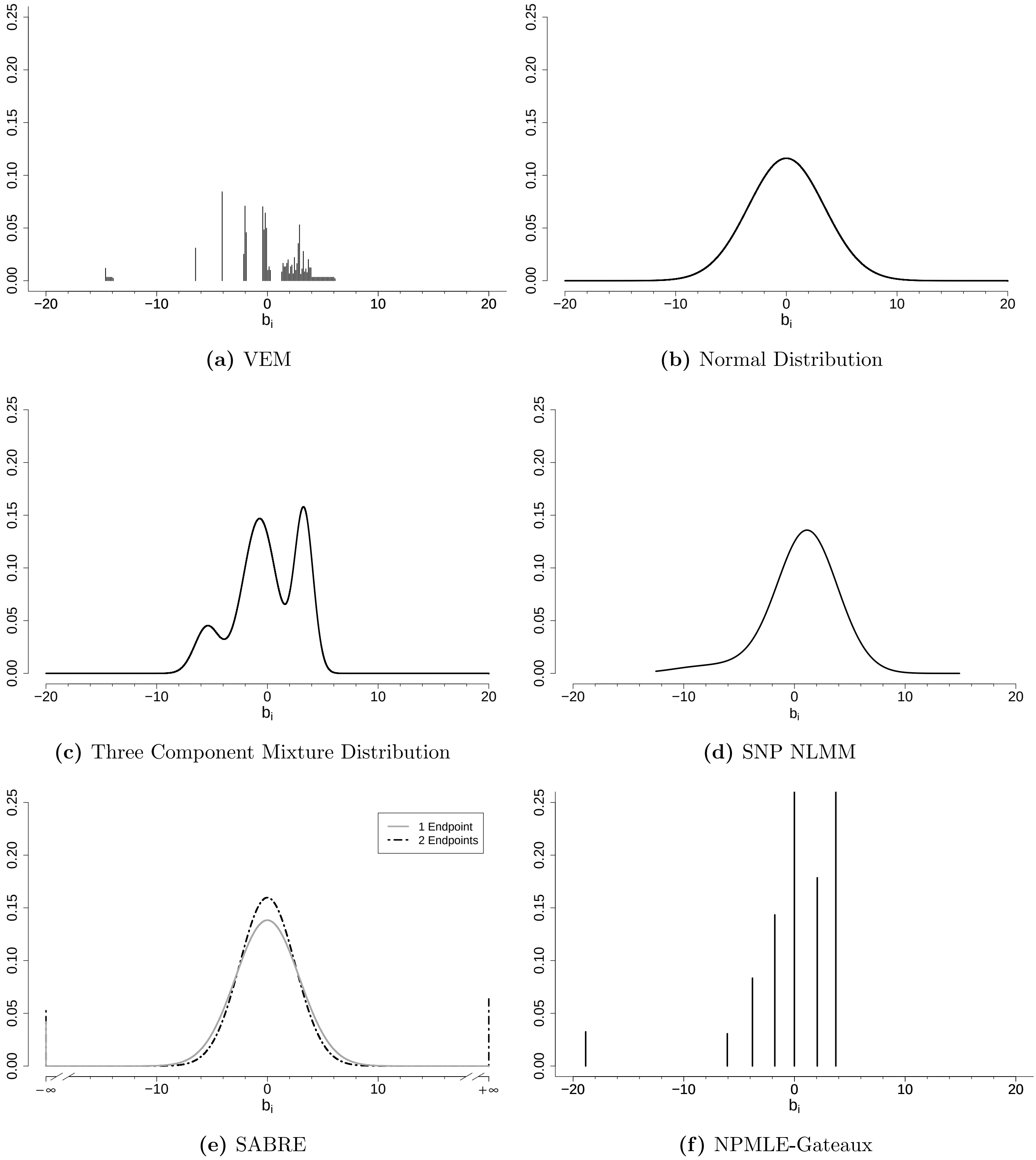
7.4.1 Random intercept logistic models

To assess the performance of VEM to non-parametrically estimate random intercept logistic models, the results of fitting VEM and the six alternative flexible random effects approaches are presented separately for the two analysis sub-groups, the results for women with complete cases are presented in Section 7.4.1.1 and the results for women with monotone missingness are presented in Section 7.4.1.2. The sensitivity of inference to the assumed random effect distribution was assessed by comparing parameter coefficients and standard errors, and the practicality of applying the approaches was assessed in terms of computational efficiency (Section 7.4.1.3).

7.4.1.1 Complete case data

The parameter estimates and corresponding standard errors of the seven approaches applied to model the random intercept logistic model for the 1359 women in the HILDA sample with complete cases are presented in Table 7.1. The estimated random intercept distributions for the seven approaches for women with complete case data are shown in Figure 7.1.

The non-parametric maximum likelihood estimates for the VEM approach are given in the first part of Table 7.1. The VEM approach required 881 iterations to converge to a random effect distribution represented by 62 grid points (Figure 7.1(a)). The resulting specification of the random intercept distribution identifies regions of support for the NPML estimator, suggesting a region with large negative random intercepts (ranging between -14.62 and -13.93, with cumulative probability weight of 0.03) and five other regions of support ranging between -6.49 to 6.11. To investigate whether the random intercept potentially captures the underlying mover-stayer scenario, subject-specific empirical Bayes (EB) predictions of the random intercepts were generated. Unlike the predicted probability that is dependent on the EB estimate and the fixed effects in the model, subjects with similar response profile patterns but have different values for the fixed effect explanatory variables will subsequently have different EB estimates. Hence, caution is required when interpreting the EB estimates in this context. Therefore, the EB estimates will be used merely as an indication for the location of the underlying random effects. The EB random intercept estimates suggest one subject had the most extreme negative random intercept value of -14.62, who was unemployed for all 11 waves and interestingly, at wave 1 reported to be permanently unable to work. Of the remaining women reported to be unemployed for all 11 waves, 60 had an EB estimate of -6.49 and 42 had an EB estimate of -4.09. Generally, the EB estimates were positively related to the number of times women were employed over the 11 waves. Women employed 1 to 10 of the 11 waves had EB



(a) VEM

(b) Normal Distribution

(c) Three Component Mixture Distribution

(d) SNP NLMM

(e) SABRE

(f) NPMLE-Gateaux

Figure 7.1: Random intercept distributions for random intercept logistic models assuming normal and more flexible random effect distributions applied to the HILDA case study for women with complete case data

Table 7.1: Comparison of VEM to other flexible random effect models applied to the 1359 women with complete case data in the HILDA case study considered in Chapter 4. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, where applicable.

	VEM		**Normal**		**3 Component Mixture**		**SNP NLMM (K=2)**		**SABRE 2 Endpoints**		**SABRE 1 Endpoint**		**NPMLE (M=7)**	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-0.895		1.203	0.495	1.068	0.464	0.629	0.461	0.726	0.545	0.828	0.436	-0.133	11.419
Age	0.120	0.003	0.091	0.010	0.090	0.009	0.091	0.009	0.089	0.010	0.096	0.010	0.099	0.010
Marital Status														
Married/Defacto														
Sep/Div/Wid	-0.347	0.140	-0.279	0.141	-0.310	0.140	-0.304	0.140	-0.293	0.141	-0.300	0.138	-0.343	0.144
Single	0.201	0.232	-0.122	0.276	-0.116	0.261	0.010	0.278	0.075	0.288	0.041	0.233	0.112	0.225
Highest Education														
Bachelor or higher														
Year 12/Dip/Cert	-1.092	0.159	-1.532	0.277	-1.521	0.247	-1.262	0.264	-1.039	0.264	-1.066	0.170	-1.145	0.170
Year 11 or less	-2.428	0.168	-2.768	0.291	-2.824	0.267	-2.345	0.288	-2.108	0.298	-2.275	0.201	-2.458	0.199
Dependent Children														
None														
Youngest<5	-2.169	0.116	-2.327	0.156	-2.330	0.150	-2.291	0.151	-2.273	0.155	-2.316	0.145	-2.319	0.145
Youngest 5-24	-0.303	0.103	-0.395	0.123	-0.396	0.119	-0.398	0.122	-0.400	0.124	-0.388	0.112	-0.383	0.111
Random Effect														
Variance	14.197		11.802	0.861	9.074	1.287	11.931	2.130	6.238		8.322		18.835	
μ_1					-5.437	0.662								
σ_1					1.121	0.417								
π_1					0.124	0.044								
μ_2					-0.700	0.342								
σ_2					1.502	0.515								
π_2					0.553	0.145								
μ_3					3.291	0.639								
σ_3					0.839	0.664								
π_3					0.323	0.109								
$\Pr(-\infty)$									0.053		0.042			
$\Pr(+\infty)$									0.064					
$-2ll$	9667		9697		9691		9677		9698		9669		9660	

estimates ranging between -6.49 and 2.90. Of the 631 women who were in employment for all 11 waves, 93 (14.7%) had an EB estimate of -0.42 and 538 (85.3%) had the most positive EB estimate of 2.90. Therefore, by leaving the random intercept distribution unspecified, the VEM induces flexibility to capture extreme random effects.

In comparison to the conventional logistic model assuming normal random intercepts ($-2 \times ll$ =9697), non-parametric estimation based on VEM resulted in smaller residual deviance ($-2 \times ll$=9667) (Table 7.1). Similarly, the residual deviance of the model estimated by VEM was lower (differences ranging from 2 to 31) than the other flexible random effect methods, with the exception of the NPML estimation based on the Gateaux method which had the lowest residual deviance of 9660. As discussed further in Section 7.5, one possible explanation for the differences in the residual deviance is due to the limited and differing support in the extreme random effect values for some methods.

The magnitude and direction of the fixed effect coefficient estimates were similar for the seven estimation approaches, with the exception of changes in the direction of the coefficient relating to single women. The standard errors corresponding to the fixed effect parameters were of similar magnitude for all models, with the exception for standard errors corresponding to the parameters capturing the effects of highest education at baseline (β_4, β_5). The standard error estimate of the VEM method relating to the age term was at least a third the magnitude of that for the other methods. Albeit some differences in the magnitude and direction of the fixed effect parameter estimates, based on the 5% significance level, the same inference would be made for all seven approaches.

Estimation of the parameters relating to the random intercept was sensitive to the statistical approach implemented. The magnitude and direction of the constant coefficient ranged from -0.895 as estimated by VEM to 1.203 when assuming normally distributed random intercepts (Table 7.1). Similarly, the estimate of the random intercept variance differed substantially depending on the flexible random effect method. As the variance of the SABRE models is restricted to the potential mover sub-population, the SABRE random intercept variance can not be directly compared. The variance estimate of the other five approaches ranged from 9.07 (SE=1.29) for the mixture approach to 18.83 for the NPMLE-Gateaux approach. For non-parametric maximum likelihood the random intercept variance estimate is not a model parameter but is derived from the estimated support locations and corresponding probability weights. Therefore, the standard errors for the variance estimate are not computed for the two non-parametric approaches. Furthermore, the VEM approach also derives the constant term from the estimated random intercept distribution, and as such, the corresponding standard error is not computed.

The VEM, three component mixture, SABRE and NPMLE-Gateaux approaches suggest a small proportion of women have extremely negative random intercepts (Figure 7.1). This supports the extreme response patterns, corresponding to the potential latent sub-populations. The SABRE logistic model with endpoints can be used to explicitly account for the latent mover-stayer model. Based on the residual deviance, the logistic model with one support point at negative infinity ($-2ll = 9669$) provides a better fit than the logistic model with support points at negative and positive infinity ($-2ll = 9698$). Both the endpoint models suggest that approximately 5% of the sample are latent stayers in the unemployment category (Figure 7.1(e) and $\Pr(-\infty)$ in Table 7.1). Similarly to the VEM approach, the NPMLE-Gataeux method suggests a support point with weight of 3.25% at -18.9. The SNP NLMM approach was based on an optimal smoothness parameter of two knots ($K = 2$), suggesting that the random intercept distribution is not normally distributed and appears to be skewed (Figure 7.1(d)).

As the VEM is an approximation to the NPMLE, it is useful to directly compare the two methods that estimate the random intercept distribution non-parametrically, the VEM and the NPMLE-Gateaux. The NPML estimate derived by the Gateaux derivative is based on the optimal model with seven support points ($M = 7$), with the estimated random intercept distribution shown in Figure 7.1(f). The regions with support identified by the VEM correspond with the distribution estimated by the NPMLE-Gateaux. Furthermore, these two methods resulted in similar residual deviances and produced similar coefficient estimates and standard errors for the fixed effect parameters, with some differences in the magnitude of the random intercept variance. This provides support that the approximation of the NPMLE obtained by VEM is appropriate.

7.4.1.2 Monotone missing data

The parameter estimates and corresponding standard errors of the seven approaches applied to the 1927 women in the HILDA sample with monotone missing data are presented in Table 7.2. Correspondingly, the estimated random intercept distributions are presented in Figure 7.2. The VEM performed well in the presence of missing data, requiring 892 iterations to converge to the random intercept distribution represented by the resulting 68 grid points (Figure 7.2(a)). Similarly to the complete cases, the resulting distribution of the VEM identified regions of support with large negative random intercept values (ranging between -11.82 and -10.56 with a cumulative probability weight of 0.03) and five other regions of support with values ranging between -6.44 and 6.39. Of the 243 women that were observed to be unemployed for all waves, one woman had an EB predicted value of the most extreme negative random intercept -11.82, 59 with an EB estimate of -6.44 and 183 with an EB estimate of -4.15. Interestingly, the woman with the most negative EB estimate was the same woman with the most extreme negative value in the complete case analysis who had reported as permanently unable to work. As in the complete case analysis, the EB estimates were positively correlated with the proportion of times the women were observed to be employed. Of the 934 women employed for all observed waves, the EB estimates ranged from -2.09 to 3.52 (-2.09 (n=2), -0.14 (n=446), 2.84 (n=481), 3.18 (n=4) and 3.52 (n=1)).

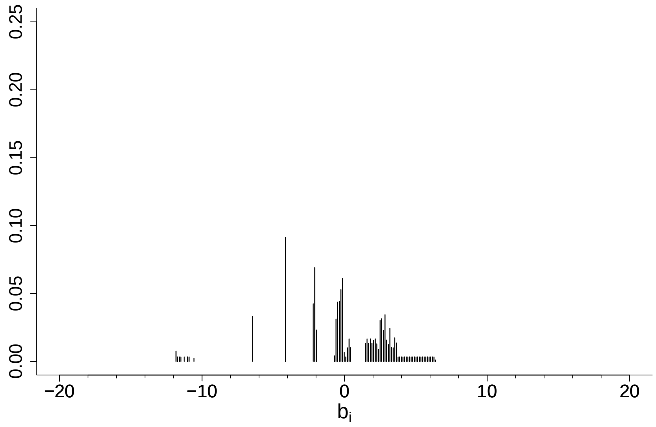
In terms of residual deviance, estimation based on the VEM approach ($2 \times ll = 11518$) resulted in substantially improved model fit over fitting the conventional model assuming normality of the random intercepts ($2 \times ll = 11543$). Similarly, estimation based on VEM had considerably smaller residual deviance than the other flexible random effect models (differences ranging from 7 to 31), with the exception of the NPML-Gateaux method which produced the lowest residual deviance of 11510.

The comparison of the VEM approach to the other flexible random effect approaches applied to the monotone missing data scenario was similar to when the methods were applied to the complete cases. One key difference is noticed in regards to inference. At the 5% significance level, interpretation of the covariate comparing women either separated, divorced or widowed to married women is considered significant for all the models except when assuming normally distributed random intercepts or if the SABRE model is estimated assuming one or two endpoints. Therefore, this application provides an example where the interpretation of fixed effects may differ depending on the implemented methodology and the presence of attrition. As identified in the complete case application, variability in the parameter estimates and standard errors was restricted to the parameters related to the random intercept. The random intercept variance for all models except the SABRE endpoint models, ranged from 9.32 as estimated by the mixture distribution and 17.57 as estimated by the NPMLE-Gateaux method based on 7 support points.

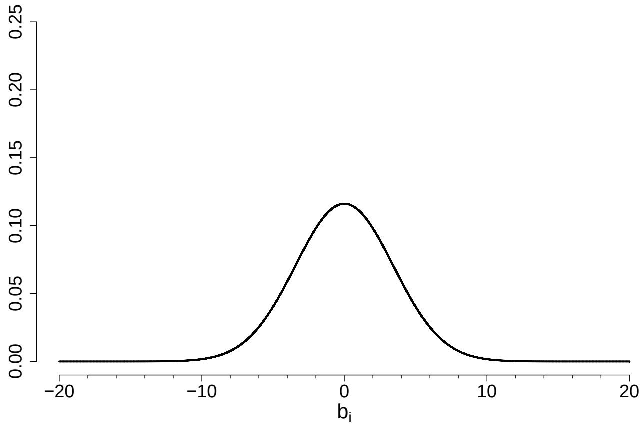
The VEM method approximating the NPMLE, produced similar parameter estimates and standard errors as the NPML estimate derived by the Gateaux method. There were some differences in the magnitude of the standard error of the age term, with the VEM estimate a third the magnitude of the standard error estimated by NPMLE (and the other approaches). The two non-parametric methods had smaller standard errors for the parameters associated with the time-invariant variable, highest education level at baseline, than the other approaches, however it did not alter inferential conclusions. As for the complete case analysis, the supportive regions identified by the VEM approach closely corresponded to those by the NPMLE characterised by the seven support points estimated by NPMLE-Gateaux (Figure 7.2(f)).

7.4.1.3 Computational efficiency

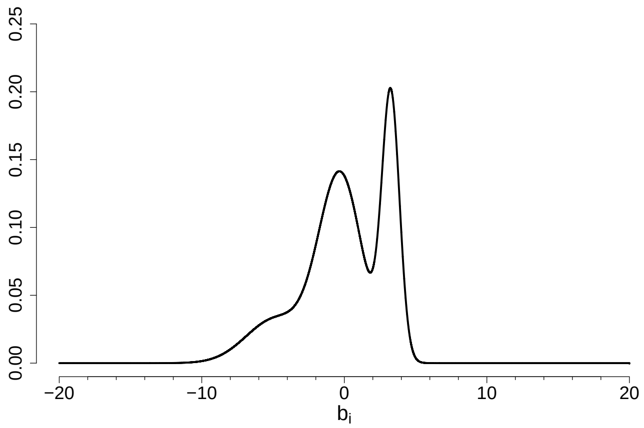
The CPU time required for each of the seven approaches and the two analysis sub-groups are presented in Table 7.3. The VEM approach was the most time consuming, due to the large number of initial grid points and the subsequent grid search required to determine the random effect distribution. By reducing the number of starting grid points from 301 to 101, the computation time for the complete cases reduced from over 6 hours to 44 minutes and resulted in similar parameter estimates and slightly higher residual deviance ($-2ll = 9671$) (Appendix G, Table G.1). The approach that required the least computation time was SABRE, the mod-



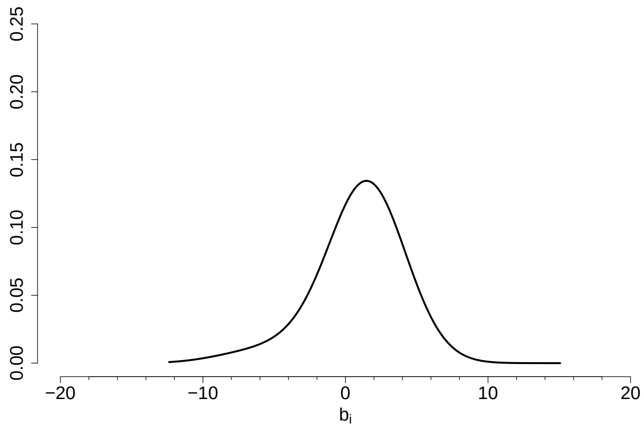
(a) VEM



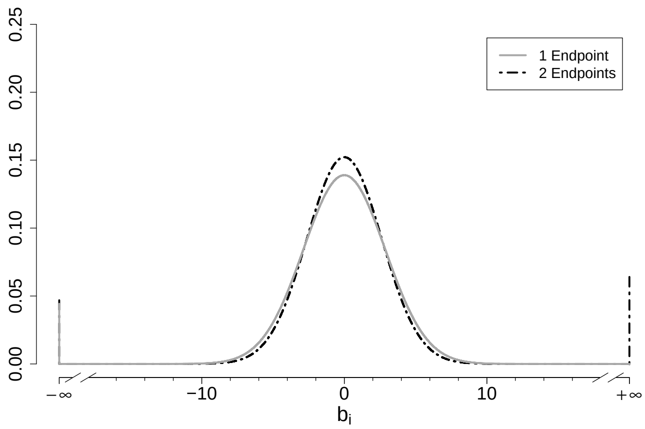
(b) Normal Distribution



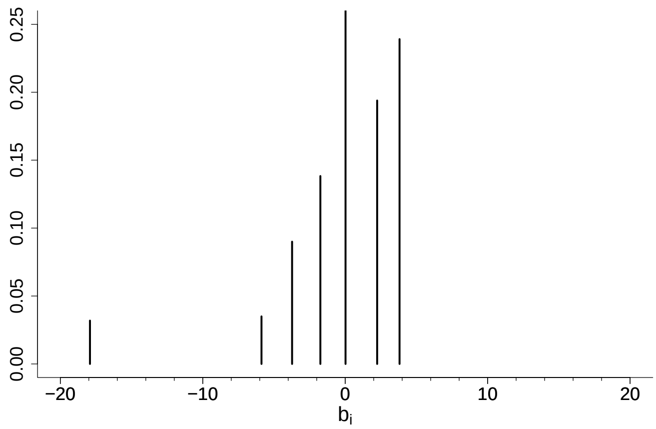
(c) Three Component Mixture Distribution



(d) SNP NLMM



(e) SABRE



(f) NPMLE-Gateaux

Figure 7.2: Random intercept distributions for random intercept logistic models assuming normal and more flexible random effect distributions applied to the HILDA case study for women with monotone missing data

Table 7.2: Comparison of VEM to other flexible random effect models applied to the 1927 women with monotone missing data in the HILDA case study considered in Chapter 4. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, where applicable.

	VEM		**Normal**		**3 Component Mixture**		**SNP NLMM (K=2)**		**SABRE 2 Endpoints**		**SABRE 1 Endpoint**		**NPMLE (M=7)**	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-0.686		1.348	0.459	0.676	0.425	0.893	0.430	0.911	0.509	0.984	0.417	0.071	8.252
Age	0.118	0.003	0.088	0.009	0.095	0.009	0.088	0.009	0.087	0.009	0.092	0.009	0.095	0.009
Marital Status														
Married/Defacto														
Sep/Div/Wid	-0.289	0.135	-0.249	0.133	-0.319	0.137	-0.265	0.133	-0.242	0.135	-0.262	0.134	-0.294	0.139
Single	0.069	0.219	-0.190	0.246	-0.084	0.242	-0.143	0.245	-0.100	0.260	-0.050	0.221	0.003	0.209
Highest Education														
Bachelor or higher														
Year 12/Dip/Cert	-1.166	0.152	-1.642	0.247	-1.555	0.240	-1.450	0.241	-1.281	0.265	-1.170	0.169	-1.232	0.166
Year 11 or less	-2.541	0.159	-2.923	0.257	-2.865	0.265	-2.637	0.257	-2.462	0.281	-2.411	0.199	-2.559	0.186
Dependent Children														
None														
Youngest<5	-2.186	0.111	-2.352	0.146	-2.313	0.141	-2.335	0.142	-2.312	0.146	-2.329	0.136	-2.339	0.137
Youngest 5-24	-0.343	0.100	-0.439	0.117	-0.418	0.116	-0.445	0.116	-0.446	0.119	-0.430	0.109	-0.422	0.107
Random Effect														
Variance	11.937		11.808	0.792	9.323	0.914	11.298	1.536	6.869		8.243		17.572	
μ_1					-4.917	1.707								
σ_1					1.935	0.620								
π_1					0.169	0.121								
μ_2					-0.378	0.348								
σ_2					1.439	0.906								
π_2					0.515	0.260								
μ_3					3.245	0.523								
σ_3					0.671	0.555								
π_3					0.316	0.146								
$\Pr(-\infty)$									0.047		0.044			
$\Pr(+\infty)$									0.068					
-2*ll*	11518		11543		11527		11525		11549		11525		11510	

Table 7.3: Computational CPU time (hours:minutes) required to execute the seven flexible random intercept logistic model approaches when applied to the complete case (Complete Cases) and monotone missing (Monotone Missing) data scenarios.

	Complete Cases	**Monotone Missing**
VEM	6:28	8:10
Normal	0:01	0:02
3 Component Mixture	0:15	0:28
SNP NLMM	1:51	2:58
SABRE - 2 Endpoints	<0:01	<0:01
SABRE - 1 Endpoint	<0:01	<0:01
NPMLE-Gateaux	0:20	0:17

els with either one or two endpoints requiring less than 1 second. Generally the approaches applied to women with monotone missing data required longer CPU time in comparison to the approaches applied to the complete cases, due to the larger sample size.

In regards to the practicality within a panel survey setting, all the existing methodology were easily implementable. Particularly the SNP approach and the NPMLE-Gateaux method. The SNP approach implemented in SAS with the SNP NLMM macro was relatively stable to the choice of quadrature points and the number of grid points used in the random effect density estimation (results not shown). The NPLME-Gateaux method implemented in STATA using GLLAMM was robust to the grid range used in the Gateaux method (results not shown), with minor changes in the estimated random intercept distribution and coefficient estimates for parameters associated with the random intercept.

However as mentioned in Chapter 4 fitting a three component mixture of normal distributions with unequal variances using the likelihood reformulation in SAS was extremely sensitive to starting values (up to the fourth decimal place) and the number of quadrature points. Hence, the final models were obtained after numerous attempts alternating starting values and quadrature points, which is not reflected in the reported CPU time. Similarly, the SABRE models were easily implemented in R, however the model coefficients and the number of endpoints were extremely sensitive to the choice of non-adaptive or adaptive quadrature, in addition to the number of quadrature points and the starting values for the end-point parameters (results not shown). Subsequently, model selection for the SABRE models was not straightforward. The SABRE models with one or two endpoints presented in Tables 7.1 and 7.2 were based on the combination of quadrature points and type of quadrature that resulted in the smallest residual deviance (results not shown). Models with two endpoints were estimated using 6 adaptive quadrature points, and models with one endpoint were estimated using 24 non-adaptive quadrature points. Both one and two endpoint models have been presented to highlight the sensitivity of the SABRE model to user inputs, and to highlight the practicality of SABRE models in panel data settings. Furthermore, it should be noted that differences in the residual deviances for the SABRE models with one or two endpoints may partly be due to the differing user inputs, and thus, caution may be required when interpreting SABRE model comparisons.

7.4.2 Random intercept and random slope logistic model

The results of fitting the VEM and the alternative flexible random effect approaches to the logistic mixed model with bivariate random effects are presented separately for the two analysis sub-groups. The results for the analysis of the complete case and the monotone missing data are presented in Sections 7.4.2.1 and 7.4.2.2, respectively. Fitting the logistic model with random effects distributed as a three component mixture of bivariate normal distributions failed to converge as detailed in Section 7.4.2.3. Therefore, the results presented are restricted to the VEM, Normal, SNP and NPMLE-Gateaux approaches. Furthermore, the NPMLE Gateaux method did not converge to the NPMLE for either of the data analysis sub-groups. For both scenarios, the Gateaux derivative suggested an additional location for a support point, however estimation of the model with the additional support point failed to converge to a resulting solution. Hence, the corresponding results for the NPMLE Gateaux presented in this section are based on the fit of the previous convergent model prior to non-convergence, and hence, cannot be interpreted as the NPMLE solution.

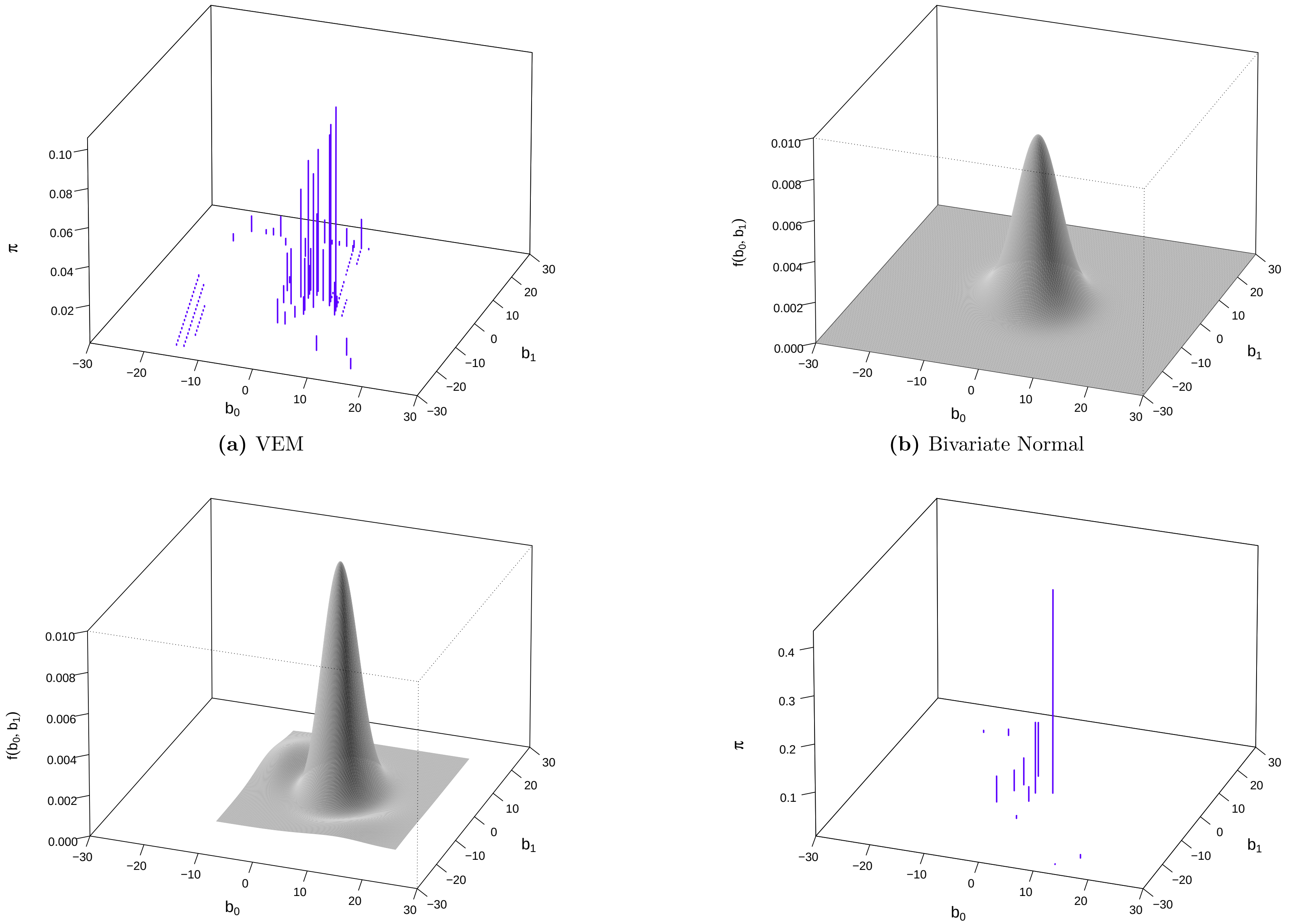
7.4.2.1 Complete case data

The parameter estimates and corresponding standard errors of the four approaches applied to the random intercept and random slope logistic model for the 1359 women in the HILDA sample with complete cases are presented in Table 7.4. The estimated random effect densities for the four approaches are shown in Figure 7.3.

The non-parametric maximum likelihood estimates for the VEM approach are given in the first part of Table 7.4. The VEM approach required 1326 iterations to converge to a random effect density represented by 113 grid points (Figure 7.3(a)). As for the random intercept model, the resulting specification of the random effects distribution identified regions of support in the extreme values, particularly the following combinations: negative random intercepts and negative random slopes, negative random intercepts and positive random slopes, and positive random intercepts and negative random slopes. The EB estimates corresponded with the response profile pattern and the fixed effect explanatory variables. Women observed to be unemployed for all 11 waves had extremely negative random intercepts and random slope values. The most extreme EB estimate for the random intercept b_{0i} and random slope b_{1i} was $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(-17.26, -6.57)', which corresponded to the same woman identified in the random intercept logistic model to have the most extreme negative random intercept. The remaining 103 women employed for all 11 waves had negative random intercepts and negative random slopes ranging between $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(-4.10, -9.75)' and $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(-1.75, -1.81)'. The most negative random intercepts and positive random slopes corresponded to women consecutively employed for a small number of times in the last waves. For instance, two women only employed once at wave 11 had an EB

Table 7.4: Comparison of VEM to other flexible random effect approaches used to estimate the random intercept and random slope logistic models when applied to the 1359 women with complete case data. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance-covariance matrix, where applicable.

	VEM		**Normal**		**SNP NLMM (K=2)**		**NPMLE (M=12)**	
	Est	SE	Est	SE	Est	SE	Est	SE
Constant	2.964		5.001	0.406	3.697	0.317	3.744	0.367
(Wave-1)/10	2.597		2.379	0.294	2.586	0.314	1.344	0.394
(Age Baseline-30)/10	1.252	0.144	0.248	0.316	0.451	0.124	0.650	0.316
Marital Status								
Married/Defacto								
Sep/Div/Wid	-0.445	0.174	-0.250	0.191	-0.305	0.189	-0.257	0.182
Single	0.492	0.272	0.078	0.355	0.281	0.377	0.247	0.242
Highest Education								
Bachelor or higher								
Year 12/Dip/Cert	-1.154	0.172	-1.776	0.331	-1.192	0.297	-1.027	0.215
Year 11 or less	-2.812	0.193	-3.293	0.351	-2.528	0.331	-2.570	0.228
Dependent Children								
None								
Youngest<5	-2.647	0.130	-2.709	0.210	-2.596	0.203	-2.708	0.189
Youngest 5-24	-0.710	0.128	-0.749	0.178	-0.687	0.180	-0.813	0.167
Random Effects								
$\sigma^2_{b_0}$	35.351		17.182	1.691	25.447	3.133	19.253	
$\sigma^2_{b_1}$	78.141		23.869	2.524	32.259	3.587	35.774	
σ_{b_0,b_1}	-22.926		-4.459	1.518	-11.661	2.815	-16.801	
$-2ll$	8995		9105		9063		9031	



(a) VEM

(b) Bivariate Normal

(c) SNP NLMM

(d) NPMLE-Gateaux

Figure 7.3: Random effect density for random intercept and random slope logistic models assuming bivariate normal and more flexible random effect distributions applied to the HILDA case study for women with complete case data

estimate of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(-22.6,17.3)', and 16 women consecutively employed 2 to 4 times in the last 2 to 4 waves had $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(-21.3,23.6)'. Similarly, the most positive random intercepts and negative random slopes corresponded to women consecutively employed for a large number of times in the initial waves. For instance, 13 women consistently employed for the first 8 to 10 waves had an EB value of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(17.1, -24.0)', 9 women consecutively employed for the first 7 to 9 waves had an EB value of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(12.6, -22.5)' and 8 women consecutively employed for the first 5 to 7 waves had an EB value of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$=(8.1, -20.7)'. There was variability around EB values for women with similar response profile patterns due to differences in the values of the explanatory variables (results not shown). Interestingly there were no women with extreme positive random intercepts and slopes. Of the women observed to be employed for all 11 waves, 614 had EB estimate of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(3.8,-1.8)' and 17 with $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(5.5,-3.4)'. Therefore, the VEM approach appears to be a flexible approach to capture the extreme bivariate random effects.

Flexibly modelling the random effects density improved the model fit, with the SNP, NPMLE-Gateaux and the VEM approaches resulting in smaller residual deviances than the conventional model assuming bivariate normally distributed random effects ($-2ll = 9105$). Estimation based on the VEM approach resulted in the lowest residual deviance ($-2ll = 8995$), substantially smaller than that for the SNP ($-2ll = 9063$) and the NPMLE-Gateaux method ($-2ll = 9031$).

The magnitude and direction of the fixed effect coefficients unrelated to the random effects were similar for the four estimation approaches, however the VEM approach generally produced more positive coefficient estimates than the other three approaches. The magnitude of the standard errors corresponding to the fixed effect parameters varied between the four approaches. The VEM approach generally produced the smallest standard errors, particularly for the coefficients relating to baseline education and the number of dependent children. The VEM standard error estimates of the parameters relating to the baseline education was approximately half the magnitude of those for the assumed normal model. At the 5% significance level, inference of the fixed effect parameters differed depending on the estimation approach. Interpretation of the age at baseline is considered significant for all approaches except when assuming bivariate normally distributed random effects. Furthermore, the parameter capturing the effect of women either separated, divorced or widowed comparing to married women was considered significant for the two non-parametric approaches, VEM and NPMLE-Gateaux, however not for the Normal or SNP approach.

Estimation of the parameters relating to the random effects was sensitive to the estimation approach. The magnitude of the constant coefficient ranged from 2.964 for the VEM approach to 5.001 when assuming bivariate normal random effects. Similarly, the magnitude of the wave term coefficient ranged from 1.344 for the NPMLE Gateaux approach to 2.597 for the VEM approach. The corresponding standard errors were similar for the three approaches that estimated standard errors. The estimate of the random intercept variance ($\sigma^2_{b_0}$), random slope variance ($\sigma^2_{b_1}$) and the covariance (σ_{b_0,b_1}) differed substantially depending on the estimation approach. The VEM approach produced the largest estimates (in terms of magnitude) for all components of the variance-covariance matrix, with an estimate of the random slope variance ($\sigma^2_{b_1} = 78.1$) at least double the magnitude of that for the other methods.

The two non-parametric approaches suggest that a proportion of women have extreme random effects. The inclusion of random slopes continues to support the extreme response patterns that correspond with the latent sub-population. The women with extremely negative random intercepts and negative slopes may represent the latent stayers in unemployment (resulting in predictive probabilities of almost zero for any time-point), and similarly, women with extremely positive random intercepts and positive slopes may represent latent stayers in employment (predictive probabilities of almost one for any time-point). Therefore, as observed in the random intercept logistic model, this is suggestive of latent stayers in the unemployment state. The SNP approach, based on an optimal smoothness parameter of two knots in the two-dimensional space (K=2), suggests that the random effect density is not bivariate normal (i.e K=0) but a skewed, unimodal density (Figure 7.3(c)).

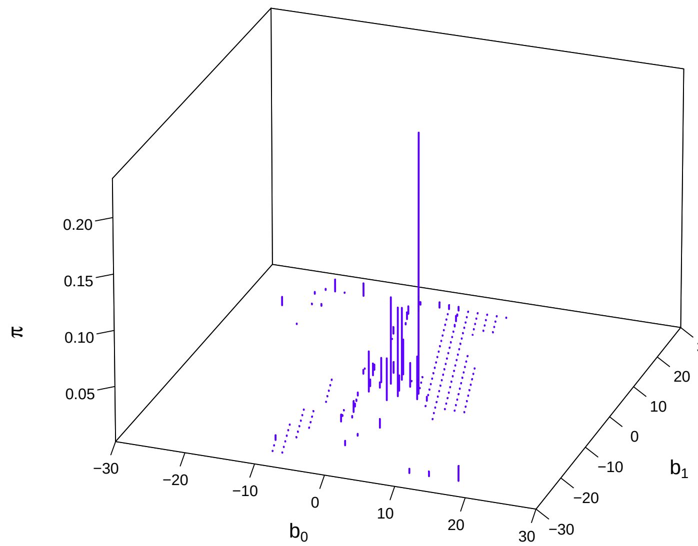
As the NPMLE-Gateaux method did not converge to the NPML estimator, the model fit and estimated random effect distribution (Figure 7.3(d)) can only be interpreted as an approximation to the NPML estimator. Therefore the performance of the VEM to approximate the NPML estimator can not be directly assessed. However, it is reassuring that the resulting discrete random effect distribution of the NPMLE-Gateaux approach with 12 support points was located in similar regions as identified by VEM (Figure 7.3(a)). However, unlike the VEM approach, the NPMLE-Gateaux did not have any support points in regions with large negative random intercepts and random slopes. However this may be due to the NPMLE-Gataeux not converging to the NPML estimate.

7.4.2.2 Monotone missing data

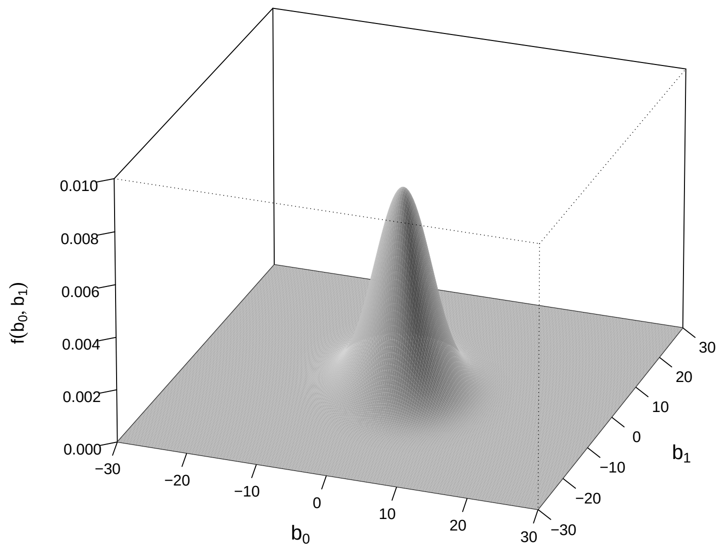
The results of the VEM approach and the other flexible random effect models applied to the 1927 women with monotone missing data are shown in Table 7.5 with the resulting random effect distributions presented in Figure 7.4. The VEM performed well in the presence of missing data, requiring 1162 iterations to converge to a bivariate random effect distribution represented by 169 grid points (Figure 7.4(a)). The resulting distribution was similar to the distribution for the complete cases, yet with more variability. Similar regions of support were identified as for the complete case analysis, however the supportive region in the extreme negative random intercept and negative random slope was not as extreme. The EB predicted random effects appeared to be correlated with the number of waves a woman was in the HILDA study, the response profile pattern and the values of the explanatory variables. For example, the extreme negative random intercept and random slope of $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(-9.9, -22.2)' corresponded to 10 women

Table 7.5: Comparison of VEM to other flexible random effect approaches used to estimate the random intercept and random slope logistic models when applied to the 1927 women with monotone missing data. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance-covariance matrix, where applicable.

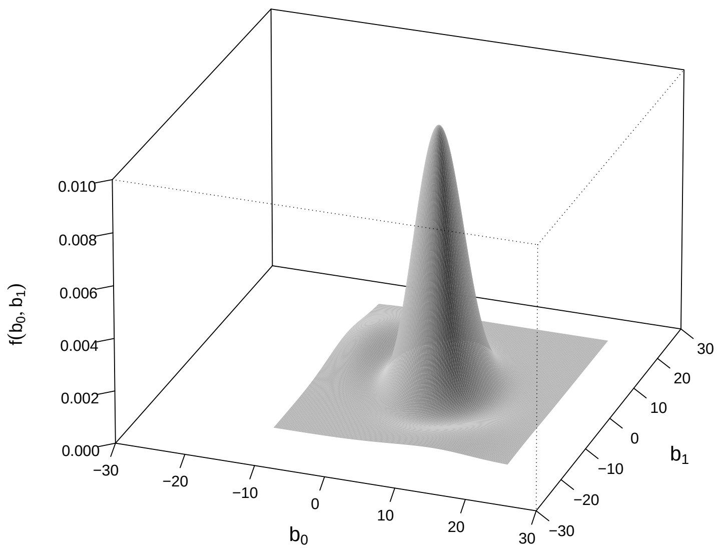
	VEM		**Normal**		**SNP NLMM (K=2)**		**NPMLE (M=9)**	
	Est	SE	Est	SE	Est	SE	Est	SE
Constant	3.599		5.060	0.365	3.920	0.284	4.183	0.272
(Wave-1)/10	2.820		2.389	0.277	2.619	0.445	1.230	0.284
(Age Baseline-30)/10	0.705	0.146	0.291	0.274	0.441	0.222	0.334	0.159
Marital Status								
Married/Defacto								
Sep/Div/Wid	-0.466	0.162	-0.242	0.180	-0.259	0.193	-0.298	0.165
Single	0.260	0.228	-0.057	0.316	-0.024	0.514	-0.074	0.215
Highest Education								
Bachelor or higher								
Year 12/Dip/Cert	-1.316	0.162	-1.892	0.297	-1.406	0.274	-1.440	0.168
Year 11 or less	-2.750	0.171	-3.420	0.312	-2.841	0.308	-2.557	0.195
Dependent Children								
None								
Youngest<5	-2.746	0.123	-2.782	0.196	-2.699	0.180	-2.768	0.168
Youngest 5-24	-0.812	0.123	-0.869	0.168	-0.812	0.167	-0.981	0.148
Random Effects								
$\sigma^2_{b_0}$	32.438		17.569	1.563	26.765	4.304	11.907	
$\sigma^2_{b_1}$	98.404		26.757	2.697	37.555	4.424	19.799	
σ_{b_0,b_1}	-25.498		-5.204	1.489	-13.090	3.471	-7.369	
$-2ll$	10740		10867		10824		10843	



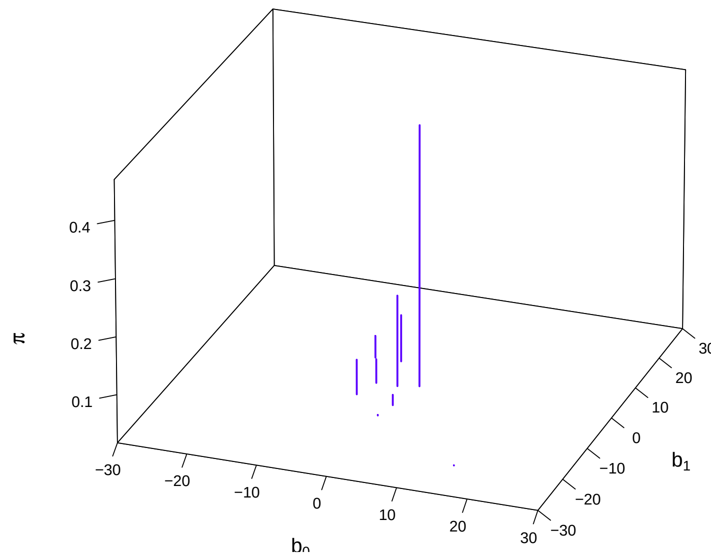
(a) VEM



(b) Bivariate Normal



(c) SNP NLMM



(d) NPMLE-Gateaux

Figure 7.4: Random effect density for random intercept and random slope logistic models assuming bivariate normal and more flexible random effect distributions applied to the HILDA case study for women with monotone missing data

unemployed for all 11 waves. For the remaining women always observed in unemployment, the EB estimates varied from $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(-24.0, 18.0)' for women observed for 9 to 11 waves, to $\hat{\boldsymbol{b}}_{\boldsymbol{i}}$ =(-1.3, 0.4)' for women observed for 1 to 5 waves.

In comparison to the conventional logistic mixed model assuming bivariate normal random effects ($-2ll$ = 10867), estimation based on the VEM substantially improved the model fit ($-2ll$ = 10740). Furthermore, the VEM had a smaller residual deviance than the other two flexible random effect models, SNP and the NPMLE-Gateaux (differences of 84 and 103, respectively).

In regards to inference, the statistical significance (at the 5% level) differed depending on the estimation approach. As in the complete case data scenario, interpretation of the age at baseline was considered significant for all approaches except when assuming bivariate normally distributed random effects. Furthermore, the covariate comparing women separated, divorced or widowed to married or defacto women was only considered significant for VEM. As identified in the complete cases, variability was observed for all parameter estimates, particularly for parameters related to the random effects. The VEM produced the largest estimates (in terms of magnitude) for all components of the variance-covariance matrix. There were differences in the magnitude of the standard errors, with the VEM consistently producing the smallest standard errors (for all parameters).

The NPML estimation using the Gateaux failed to converge to the NPMLE, so the model fit and the estimated random effect distribution is based on the convergent model with 9 support points (Figure 7.4(d)). As for the complete case scenario, it is reassuring that the approximate distribution of the NPMLE-Gateaux approach has support points in similar locations identified by VEM.

7.4.2.3 Computational efficiency

The CPU time required for each of the four approaches to model the bivariate random effects distribution when applied the two analysis sub-groups is presented in Table 7.6. The VEM approach was the most time consuming, due to the large number of initial grid points and the subsequent grid search required to estimate the bivariate random effects distribution. Due to the complexity estimating the bivariate distribution, all four approaches required substantially more computational time than required to estimate the random intercept logistic model (Table 7.3). The standard logistic mixed model assuming bivariate normality required almost 2 hours for the complete case analysis, and almost 3 hours for the analysis with monotone missing data.

In contrast to the random intercept logistic model, implementation of the flexible random effects and estimation were not straightforward. Additionally, not all of the flexible approaches converged. Convergence of the SNP approach for the bivariate random effects was sensitive

Table 7.6: Computational CPU time (hours:minutes) required to execute the four flexible random intercept and random slope logistic model approaches when applied to the women with complete case data (Complete Cases) and the women with monotone missing data (Monotone Missing).

	Complete Cases	**Monotone Missing**
VEM	37:44	50:52
Normal	1:45	2:42
SNP NLMM	5:26	6:29
NPMLE-Gateaux	2:18	0:45

to the number of quadrature points. For example, for the complete case data scenario, the model fit with a smoothness parameter of two ($K = 2$) failed to converge when estimated with 20, 25 or 31 quadrature points. Similarly, for the monotone missing data scenario the model with $K = 2$ failed to converge when estimated with 21 quadrature points, yet it did converge when estimated with 25 or 31 quadrature points. Therefore, the results in Table 7.5 for the monotone missing data are based on 31 quadrature points and 20 grid points (the model corresponding to the minimum AIC). However, even though the SNP approach converged, not all of the standard errors of the model parameters were estimated. Both of the models presented in Tables 7.4 and 7.5 failed to estimate standard errors for one of the smoothness parameters used to estimate the random effect density. Furthermore, the SNP approach was also sensitive to the parameterisation of the model covariates. For instance, when the time components (age at baseline and the wave term) were not transformed to be interpreted as a per 10 year increase, the SNP approach with two knots failed to converge (results not shown).

The NPMLE-Gateaux method implemented in STATA using the GLLAMM procedure failed to converge to the NPMLE. The results presented in Tables 7.4 and 7.5 are based on models with 12 and 9 support points, corresponding to the last model that successfully converged. Therefore, the resulting model fit from the NPMLE-Gateaux method can not be defined as the NPML estimator, and the reported CPU time refers to the last converging model (Table 7.6). Furthermore, the NPMLE-Gateaux method was sensitive to the number of initial support points, resulting in unstable coefficient estimates and estimated random effects distribution (results not shown).

Fitting the random effects as a three component mixture of bivariate normal distributions failed to converge. Similar convergence problems have previously been reported by Ghidey et al. (2010) in the context of linear mixed models, whereby estimation of the heterogeneity model with more than two components of bivariate normals was not feasible. By reducing the number of parameters considered in the model and restricting the components of the finite mixture to have the same variance-covariance matrix, the model still failed to converge. As for the random intercept application, the likelihood reformulation method implemented in SAS was sensitive to the starting values and the number of quadrature points. Limited literature is available regarding the selection of starting values. In this application starting values for the fixed effect coefficients were set as the estimated coefficients from the corresponding model assuming multivariate normal random effects. Starting values for the parameters of the three component mixture, the mixing proportions and the component means for the random intercept and random slopes, were based on the results of a three component cluster (K-means) of the Empirical Bayes estimates of the corresponding model assuming bivariate normal distributed random effects. A range of starting values based on the results for three component mixture parameters were assessed iteratively. However, none of the starting values considered resulted in a convergent model.

In contrast to the other flexible modelling approaches, the VEM model was easily implementable to the bivariate logistic mixed model application. The VEM approach was relatively robust to the choice of initial starting values for the coefficients and the initial grid, including the number of grid points and the range of grid points (Appendix H). The sensitivity was restricted to parameters relating to the random effects and the estimated variance-covariance matrix, predominately due to the solutions having supportive regions on the boundary of the two-dimensional grid. The resulting solution of the VEM approach had support at the boundaries, particularly at the minimum and maximum random slope values (as shown in Figures 7.3(a) and 7.4(a)). Even when the initial grid for b_i^* was defined as $[-7,7]\times[-7,7]$ the resulting solution had support at the boundaries (Appendix H). The boundary issue still remained when different parameterisations of the wave and age at baseline term were considered (results not shown). Therefore, the boundary issue is a direct consequence of the extreme response patterns observed in the application considered in this chapter.

7.5 Discussion

In this chapter the Vertex Exchange Method (Böhning, 1985) was applied to estimate logistic mixed models with unspecified random effect distributions. The performance of the VEM was assessed in an application to the HILDA panel survey. By extending the random intercept model considered in Chapter 4 to random intercepts and random slopes, this chapter presents a novel application of the VEM to flexibly model univariate and bivariate random effects in GLMMs. The VEM approach performed well to estimate logistic models with univariate and bivariate random effects, inducing sufficient flexibility to capture the underlying heterogeneity. Although computationally intensive, the simplicity of the VEM approach (Baghfalaki and Ganjali, 2014) was consequently easily implementable. The resulting specification of the random effects distribution is an approximation to the non-parametric maximum likelihood estimator, identifying regions of support in the $q-$dimensional random effects density. The specified random effects distribution for both the univariate and bivariate models suggested regions of support in the extremes corresponding to potential latent stayers in the unemployment state. Thus, the results from this application suggest that non-parametric estimation of GLMMs using VEM can provide an appropriate modelling strategy to capture the potentially extreme underlying heterogeneity of latent mover-stayer scenarios.

To assess the performance of VEM, the model fit of the VEM approach was compared to alternative random effect approaches currently implementable in standard software. In addition to the conventional logistic model with normal random effects, four flexible random effect approaches were considered: a finite mixture of normal distributions (Verbeke and Lesaffre, 1996), SNP density (Vock et al., 2014), normal distribution with endpoints at positive and/or negative infinity (Berridge and Crouchley, 2011b), or non-parametric estimation using the Gateaux derivative (Rabe-Hesketh et al., 2003). Multivariate extensions of the aforementioned methods were applied to flexibly model the bivariate random effects in a logistic mixed model, except for the end-point model of Berridge and Crouchley (2011b). For both logistic models with univariate and bivariate random effects, the conventional logistic model assuming normality of the random effects had the largest residual deviance. This suggests that the normality assumption may not be the most appropriate distribution to capture the extreme random effect distribution resulting from potential mover-stayer scenario. Increasing the flexibility of the assumed random effect distribution resulted in an improved model fit (in terms of residual deviance). For both the random intercept logistic model and the more complex random intercept and random slope logistic model, non-parametric estimation of the random effects, either by VEM or the Gateaux approach, substantially improved the model fit compared to the equivalent model assuming normality. Furthermore, the non-parametric approaches had lower residual deviance than any other flexible modelling approach considered. This application suggests that non-parametric estimation may provide an efficient and suitable computational approach in panel survey applications.

Not only are non-parametric estimation approaches useful to guard against possible implications of misspecified random effect distributional assumptions (Agresti, 2013), they are also efficient (Butler and Louis, 1992; Agresti, 2013), particularly when the random effects distribution is not of direct interest (Agresti, 2013). However, non-parametric approaches can be susceptible to estimated mass points located on the boundary (boundary solutions), particularly for GLMMs with categorical responses (Skrondal and Rabe-Hesketh, 2004). This was evident for VEM estimation of the random intercept and random slope logistic model. Not only did the resulting distribution of the VEM approach have supportive regions at the minimum and maximum bounds of the initial grid range, the boundary issue was reflected in the large estimates for all components of the variance-covariance matrix. To overcome and further investigate the boundary issue, VEM bivariate models were refit with wider grids (Appendix H), as suggested by Tsonaka et al. (2009). Due to the nature of the extreme observed response profiles the boundary issue remained. However even as the boundaries were made more extreme, final inference of the fixed effects did not change (Appendix H). Furthermore, non-parametric estimation of the random intercept logistic model by NPLME-Gateaux may have resulted in a boundary solution. The resulting random intercept distributions had small probability mass at extreme negative values (corresponding to predictive probability of almost zero). The impact of the boundary solution was also reflected in the large estimated random intercept variance estimate and the large standard error of the constant coefficient of the NPMLE-Gateaux method. As observed in this application, within the context of an underlying mover-stayer scenario, there will always be support at the boundaries. Therefore, as observed in Appendix H, regardless of the range of the initial grid, the likelihood for approaches where random effect distributions put appreciable probability on large magnitude values will always be further improved by giving people with constant response profiles probability weight at the extremes. When boundary solutions (or mover-stayer scenario) are suspected, caution is required when interpreting parameters related to the random effects from these approaches. However, as primary interest is often on the fixed effects, any boundary issues are expected to have minimal inferential impact on the fixed effect parameters unrelated to the random effects.

Sensitivity of model parameters and inferential conclusions to the assumed random effects distribution as assessed within a sensitivity analysis framework may indicate misspecification (Litiére et al., 2008; Neuhaus et al., 2011). In the random intercept logistic model considered here, estimation of the fixed effect parameters was moderately robust to the assumed random intercept distribution. Consistent with the findings of McCulloch and Neuhaus (2011a), more variability in the coefficient estimates and standard errors was observed for parameters capturing the effects of time-invariant explanatory variables than for the time-varying explanatory variables. However, variability was predominately restricted to the parameters relating to the random effect distribution. The sensitivity of estimating the parameters relating to the random intercept is consistent with previous literature (Litiére et al., 2008; McCulloch and Neuhaus, 2011a). Albeit some differences in the estimation of the fixed effect parameters, inferential conclusions were similar for all parameters, with the exception of one time-varying explanatory variable.

In contrast to the random intercept model, the estimation of the logistic model with random intercepts and random slopes appeared to be sensitive to the assumed bivariate distribution, including the fixed effect parameters. Consistent with the limited literature examining misspecification of the joint random intercept and random slope distribution (Litiére et al., 2008; McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013), estimation of time-varying parameters were more robust to the assumed random effect density than time-invariant parameters. The substantial variability in the coefficients relating to the random effects, including the variance-covariance matrix, corroborate with previous results in the literature (Litiére et al., 2008; McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013), and suggest inconsistent estimation of these parameters. It has previously been shown that only minimal bias of fixed effects is expected, unless there is severe distributional misspecification (Neuhaus et al., 2013). As model estimation appears to be sensitive to the assumed bivariate density, this suggests that in this application caution is required when interpreting the model parameters. Therefore, the difference in inferential conclusions identified within this sensitivity analysis framework highlights the importance of exploring more flexible random effect distributions to assess the robustness of distributional assumptions.

As panel survey data are susceptible to missing data, the performance of the approaches was assessed in a complete case analysis scenario and in the presence of monotone drop-out missing data. For the respective flexible approaches, there were minimal differences in the magnitude of the fixed effect parameter estimates for the two data scenarios. This may be related to the nature of missingness. Previously it has been suggested that employment status is related to the likelihood of responding to the HILDA survey (Watson and Wooden, 2009). However, the similarity of the parameter coefficients for the complete case and the monotone missing data scenarios suggests that missingness may not be related to the outcome in the study population considered in this application. Hence, the potential consistency of the maximum likelihood estimation in the presence of ignorable missingness may partly explain the similarities of the parameter estimates.

No studies could be identified which compare the performance of flexibly modelling the joint random intercept and random slope distribution in GLMMs. By investigating the performance of existing methodology to flexibly model univariate and bivariate random effect logistic models in a panel survey application, we provide a novel insight into the practicality of flexibly modelling random effects. The flexible approaches considered to estimate the random intercept logistic model were easily implementable and performed relatively well, regardless of the missing data scenario. However, the SABRE approach was sensitive to the number of quadrature points and the use of non-adaptive or adaptive quadrature. This may be due to the extreme assumed random intercept distribution. By placing weight at negative and positive infinity, variability of the subject-specific random effects is not permissible. As the logistic model considered in the case study adjusts for explanatory variables, variability of the subject-specific random effects is expected, even at the extremities. By placing endpoints at the negative and/or positive infinity, potential stayer sub-populations can be identified, however in the presence of explanatory variables it may be too restrictive. Furthermore, the default specification of the initial grid in the SNP NLMM macro may not have been wide enough to capture the extreme negative random intercepts in the case study. The considered approaches performed well in the panel survey application, particularly the SABRE and two non-parametric approaches, providing sufficient flexibility to capture the extreme random intercept distribution.

However, the additional complexity of estimating random slopes subsequently complicated the implementation and performance of flexibly fitting logistic models with bivariate random effects. The commonly implemented logistic model assuming bivariate normal random effects was easily implemented, yet failed to adequately capture the potential underlying heterogeneity. Increasing the flexibility of the assumed random effect distribution, either by semi-non-parametric techniques or non-parametric techniques, was not as straightforward as for the random intercept model. With the exception of the VEM approach, the remaining flexible approaches had issues in regards to model convergence. Regardless of the choice of starting values, the three component mixture of bivariate normal distributions failed to converge. Even in the context of univariate random effects, estimation and convergence of GLMMs assuming finite mixtures of normal distributions are known to be sensitive to initial parameter values (Litiére et al., 2008). Furthermore, the NPMLE-Gateaux method failed to converge to the non-parametric maximum likelihood estimator, and convergence of SNP NLMM was sensitive to starting values and estimation settings. Even when the SNP approach converged, not all standard errors were estimable. Furthermore, as in the random intercept scenario, the default range of the grid used to estimate the SNP random effect density may not have been wide enough to capture the extreme random effects. In contrast, the VEM approach was easily implementable and had no convergence issues. Furthermore, the VEM approach was relatively stable to the choice of initial starting values and initial grid size. However, parameters relating to the random effects were sensitive to the initial grid range due to boundary issues. This highlights the importance of sensitivity analyses to identify parameters susceptible to boundary solutions for flexible approaches that place probability on large magnitude values of the random effects distribution.

One of the practical limitations of the VEM approach is the computation time required for model convergence. For both univariate and bivariate random effects, the VEM required more CPU time than any other flexible random effect approach considered. For the univariate random effects, numerous computational approaches have been developed to estimate univariate random effects in GLMMs (Follmann and Lambert, 1989; Lesperance and Kalbfleisch, 1992; Aitkin, 1999; Rabe-Hesketh et al., 2003), yet availability in standard software is limited. Non-parametric estimation approaches for GLMMs with higher dimensional random effects is an understudied area. The Gateaux method implemented in STATA is extendable to higher dimensions, however, as experienced in the panel survey application considered in this study, convergence to the NPML estimator may not occur. Recently two fast computational methods have been proposed to estimate bivariate random effects for GLMMs (Lesperance et al., 2014) and for LMMs (Baghfalaki and Ganjali, 2014). However neither is currently implementable in standard software. The estimation of the probability weights proposed by Baghfalaki and Ganjali (2014), using the relative frequency of the grid-points, is computationally faster than VEM. However, unlike the well-known properties of VEM (see Böhning (1999)), it is unclear whether the method of Baghfalaki and Ganjali (2014) is an NPMLE, as there is little evidence showing the mathematical or asymptotic properties of the proposed method. Therefore, even though the VEM approach may be considered computationally intensive, the simplicity and the well-known asymptotic properties enable reliable approximation to the NPMLE. If the VEM would be followed by the EM algorithm to fine tune the location of the support points, i.e. run the EM algorithm with the location and weights of the support points from the VEM solution as the initial starting values, the resulting solution would be NPMLE. However, as the

EM algorithm is computationally slow and has minimal additional improvement in regards to log-likelihood (Böhning, 1999), the NPMLE approximation derived from exclusively using the VEM algorithm is appropriate (Tsonaka et al., 2009). Development of efficient non-parametric estimation procedures for GLMMs with multiple random effects is an area gaining interest. The development and availability of computationally fast NPML estimation methods in standard software will benefit the practical implementation of non-parametric approaches, including within the sensitivity analysis framework.

The approximate standard errors for parameters unrelated to the random effects in the VEM approach were obtained from the Hessian of the log-likelihood evaluated at the estimates of $\hat{\boldsymbol{\theta}}$ and $\hat{G}$. However, as suggested by Follmann and Lambert (1989), these implicitly assume that the number of support points of G is equal to the estimated number and approximate standard errors have been shown to underestimate standard errors (Follmann and Lambert, 1989; Tsonaka et al., 2009), as the models fail to account for the uncertainty produced by estimating the random effect distribution (Butler and Louis, 1992). The standard errors need to account for the additional uncertainty and some non-parametric approaches estimate the standard errors using adjustment methods (Butler and Louis, 1992) or by utilising the computationally intensive bootstrap method (Tao et al., 1999; Baghfalaki and Ganjali, 2014). Aitkin (1999) proposed to obtain correct standard errors by calculating the absolute value of the parameter estimate divided by the square root of the deviance change on omitting variables one-by one, subsequently requiring numerous model fits. Furthermore, as standard errors for VEM have only been calculated for the model parameters unrelated to the random effects (i.e. $\boldsymbol{\beta}^1$ in Equation 7.1), estimation of standard errors for all model parameters could be calculated using the method proposed by Tsonaka et al. (2009) based on the Hessian of the log-likelihood for all model parameters. The proposed method of Tsonaka et al. (2009) produces good quality standard errors, except that it may overestimate the standard errors for parameters relating the random effects. Fast and consistent estimation of standard errors for all parameters in non-parametric maximum likelihood estimation of GLMMs is an area requiring further investigation.

Model comparison of parametric, semi-parametric and non-parametric estimation of GLMMs is not straight forward. Not only does difficulty arise when comparing the discrete solution of non-parametric approaches to methods resulting in continuous solutions (Ghidey et al., 2010), model comparison is complicated as standard asymptotic theory does not apply for non-parametric methods (Litiére et al., 2008). Within the sensitivity analysis framework considered in this study, the residual deviance has been used as an indication of model fit as previously utilised in the literature (Aitkin, 1999; McCulloch and Neuhaus, 2011a; Lesperance et al., 2014). Formal model comparison of non-parametric estimation of GLMMs based on differences in the residual deviance is lacking theoretical justification (Aitkin, 1999). Furthermore, as all residual deviances are approximations, caution should be applied when comparing the model fit based on deviance. However, given the focus has been on investigating the sensitivity of conclusions with respect to the distributional assumptions for the random effects, formal comparison of model fit is not of primary interest here.

In summary, this chapter highlights the practicality of implementing approaches to relax the parametric assumption of the random effects distribution to potentially reduce the risks associated with misspecifying distributional assumptions in logistic mixed models. By considering different approaches within a sensitivity analysis framework, potential distributional misspecification of the random effects can be identified in practice. In this application to the HILDA panel survey, the reported sensitivity of the logistic mixed model to the assumed random effects distribution highlights an example whereby the conventional model assuming normality may lead to biased estimation of model parameters. Leaving the random effects distribution completely unspecified and estimated using non-parametric maximum likelihood techniques provides an efficient approach to capture underlying heterogeneity of the random effects in a potential mover-stayer scenario. Non-parametric estimation using the VEM provides a reliable approximation to the NPMLE. The VEM performed well to capture the multimodality of the random effects, particularly for the more complex bivariate distributions, where the VEM algorithm was the only flexible approach to reach model convergence. Albeit the computational time required to estimate the VEM, the performance of the VEM to flexibly model the random effects in panel survey settings, including missing data due to attrition, is encouraging for consideration in future applications. This is particularly true for scenarios where extreme underlying heterogeneity of univariate or bivariate random effects may be suspected.

8 Discussion

8.1 Major findings

This study provides a novel insight into the inferential impact of assuming normal distributed random effects in logistic mixed models applied to panel survey settings where an underlying sub-population structure exists. In a mover-stayer scenario, the assumed normal distribution fails to adequately capture the underlying heterogeneity. Rather, the multimodality of the random intercepts is more appropriately characterised by a three component mixture of normal distributions. The three major findings of this study are highlighted and further discussed below:

1. Incorrectly assuming normally distributed random intercepts for underlying trimodal distributions can impact inference for model parameters differently, and is dependent on the type and severity of departure from normality.

In situations where there exists multimodality of the random intercept distribution, the inferential impact of incorrectly assuming normality is dependent on the type and severity of departure from the normal distribution. For minor departures from a single mode, and for departures from symmetry such as an asymmetric distribution in a mover-stayer scenario, incorrectly assuming normality had minimal impact on estimating fixed effect parameters. However, in these settings incorrectly assuming normality resulted in biased estimates of parameters related to the misspecified random intercept, the intercept constant and the random effect variance component. For large departures from normality characterised by multimodality in the form of three distinct modes, inference for fixed effects parameters of time-invariant explanatory variables and parameters related to the random effects were sensitive to distributional misspecification of the random effects. Incorrectly assuming normality in scenarios with distinct multimodality produced biased parameter estimates and poor coverage rates of confidence intervals for the intercept constant, time-invariant explanatory variables and those time-varying explanatory variables exhibiting minimal within-individual variability. Inference for the random effect variance was extremely sensitive to distributional misspecification of the random effects, resulting in biased estimates, poor coverage rates of confidence intervals and inaccurate standard errors.

2. Relaxing the parametric assumption of the random effects distribution using the non-parametric Vertex Exchange Method (VEM) is a viable approach to induce more flexibility to capture underlying heterogeneity in univariate and bivariate random effects

distributions in panel survey applications, including settings with missing data due to attrition.

Assuming more flexible distributions for the random effects within a sensitivity analysis framework can provide a practical approach to identify and potentially reduce the inferential impact of violating distributional assumptions of the random effects. Non-parametric approaches provided considerable flexibility to capture multimodal distributions. Non-parametric methods resulted in lower residual deviances when compared to existing methods, and were easily implementable in standard software. VEM is a promising method to non-parametrically estimate the random effects in logistic mixed models applied to panel survey data, including scenarios with missing data due to attrition. Albeit being computationally intensive, the VEM induced sufficient flexibility to capture the underlying heterogeneity. The VEM performed well when applied to logistic mixed models with univariate and bivariate random effects, and was the only approach to reach model convergence when compared to four existing methods (including another non-parametric method) to capture the bivariate random intercept and random slope distribution.

3. MAR attrition has minimal additional inferential impact on model parameters in the presence of distributional misspecification of the random effects.

For a similar rate of 29.5% attrition as observed in the HILDA case study, the inferential impact of incorrectly assuming normality in the presence of MAR attrition was similar to the impact for the random intercept logistic model with complete data. The minimal additional impact may be due to the potential consistency of logistic mixed models in the presence of MAR missingness. Furthermore, approaches to induce flexibility of the random effects distribution performed well when applied to the HILDA case study with complete or monotone missing data. For each approach, there were similarities between the parameter estimates for the two analysis sub-groups, giving support to the MAR assumption for the underlying missingness mechanism in the HILDA case study.

The remainder of this chapter provides more detailed discussion of the major findings of this study and potential implications to other applications in the social sciences. Firstly, Section 8.2 continues to discuss the inferential impact of misspecifying the random effects distribution in random intercept logistic models, considering the impact on each type of model parameter. Section 8.3 discusses the additional impact of missing data due to attrition, focusing on the impact of misspecifying the random effects distribution and the performance of approaches to flexibly model the random effects. Section 8.4 discusses the implications of the major findings of these studies in other applications of GLMMs to analyse longitudinal panel survey data. Section 8.5 details key computational issues that arise when implementing GLMMs assuming a normal or flexible distribution for the random effects. This is followed by a discussion of methodological issues for the statistical and simulation techniques implemented throughout the study (Section

8.6). Lastly, Section 8.7 discusses the limitations of this study and highlights areas for future work, and finally, Section 8.8 closes with some concluding remarks.

8.2 Severity of departures from normality

The impact of misspecification on maximum likelihood estimation of random intercept logistic models was dependent on the degree of departure of the true random intercept distribution from the assumed normal distribution, either in terms of departures from symmetry or departures from a single mode. Thus, the combination of the mean component distance and the component variances of the three component mixture of Gaussians leading to a skewed distribution or distinct multimodality was a contributing factor.

Depending on the severity of departure from normality, the impact of incorrectly assuming normality when the true random intercept distribution was multimodal could affect the estimation of the time-varying or time-invariant fixed effects differently. Furthermore, the degree of severity of misspecification had differing inferential impact on parameters related to the misspecified random effect, including the intercept constant and variance component of the random effects. The impact on estimating each of the model parameters is discussed further in the following sections.

8.2.1 Impact on inference about time-invariant fixed effects

The impact of misspecification on estimating the effects of time-invariant explanatory variables can depend on the type and the severity of departure of the true distribution from normality. For true random effects distributed as an asymmetric trimodal (Chapter 5) or symmetric mixture distribution with little or moderate multimodality (mean component distance less than 14, Chapter 6), estimation of time-invariant fixed effects was relatively robust to distributional misspecification of the random intercepts.

However, for true random intercepts distributions with distinct multimodality and large departures from normality (component mean distances of 14 or more, Chapter 6), incorrectly assuming normality produced biased estimates and poor coverage rates of the effects of time-invariant explanatory variables. Larger magnitudes of bias and lower coverage rates corresponded with more extreme cases of multimodality, as characterised by larger distances between the mean components and smaller component variances. This is consistent with previous literature, whereby bias and loss of efficiency have been reported when true distributions are substantially different from the assumed distribution (Agresti et al., 2004)[1] and for true random effects with large variability (Heagerty and Kurland, 2001; Litiére et al., 2008). Although the parameter estimates were impacted by misspecification of the random intercept distribution,

[1]For example, assuming a normal random intercept distribution when the true distribution is discrete with two points (Agresti et al., 2004).

the accuracy of model based standard errors was robust to misspecification as previously reported for a similar number of time-points as observed in the HILDA case study (McCulloch and Neuhaus, 2011a)[2].

Typically simulation studies investigating inferential impact on time-invariant fixed effects in the presence of misspecified random effects have considered a binary time-invariant explanatory variable (Heagerty and Kurland, 2001; Litiére et al., 2008; McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013). Often the binary variable is simulated to represent treatment effect (i.e within a clinical trial setting), and subjects are randomly allocated to one of the two treatment groups with equal probability. However, McCulloch and Neuhaus (2011a) considered a binary variable with unequal allocation of 25% and 75%, reporting similar inferential impact as for binary time-invariant explanatory variables with equal allocation. The simulation studies presented here consider a three-level categorical time-invariant explanatory variable representing baseline education level. By considering three categories with unequal allocation of 28%, 40% and 32% (Table 4.2), the results provide insight into the impact of misspecifying random effects on inference for categorical time-invariant fixed effects suggesting a similar impact as previously reported for binary time-invariant variables.

Estimation of parameters corresponding to the time-invariant explanatory variables is often considered an important inferential goal in longitudinal analysis (McCulloch and Neuhaus, 2011a), such as the effectiveness of a treatment in a clinical trial. Previously it has been conjectured that time-invariant fixed effects are more sensitive than time-varying explanatory variables to distributional misspecification of the random effects, as both time-invariant fixed effects and the random effects capture variability among individuals (Chen et al., 2002). The results from this study show that for large departures from normality in the form of distinct multimodality, misspecification can produce large biased estimates and poor coverage rates for the effects of time-invariant explanatory variables.

8.2.2 Impact on inference about time-varying fixed effects

Consistent with findings from previous theoretical and simulation studies, misspecification of the assumed random effect distribution had little impact on estimation and inference for time-varying explanatory variables. The minimal impact may be contrasted with time-invariant explanatory variables being roughly orthogonal to between-subject effects as previously postulated by Chen et al. (2002). Incorrectly assuming normality when the random effects were an asymmetric or symmetric mixture distribution generally resulted in minimal bias, and exhibited close to nominal coverage rates with accurate model based standard errors.

[2] Efficiency was reported for cluster sizes of 10. For cluster sizes of 20 or 40, McCulloch and Neuhaus (2011a) reported that incorrectly assuming normality resulted in larger standard deviation estimates of the time-invariant explanatory variable than correctly fitting the random intercept logistic model for a Tukey distribution.

The results from the simulation studies presented in Chapters 5 and 6 suggest that the degree of within-subject variability of explanatory variables may influence the magnitude of bias when estimating time-varying fixed effects in the presence of misspecification. For instance, the coefficient capturing the effect of single women was consistently underestimated and negatively biased. This particular variable exhibited substantial stability over the observational period, whereby only 5% of women transitioned when the explanatory variable was treated as a binary indicator variable. Furthermore, as only 10% of the women in the HILDA sub-sample were ever single over the observational period, the small cell size may also be a contributing factor to the stable nature of the explanatory variable. In addition to the stability, the large standard error and small magnitude of the true coefficient may have also contributed to the observed bias. The observed large standard error may also partly be explained by the relationship with dependent children. Although 26.6% (53/199) of single women had at least one dependent child at the first wave, the small cell size of single women with dependent children could lead to instabilities of the parameter estimates and subsequently, contribute to the large standard error.

The impact of misspecification on estimating the effects of categorical time-varying explanatory variables has not previously been considered. Simulation studies have predominantly focused on time-varying continuous covariates by either considering a single continuous covariate representing a linear time trend (Litiére et al., 2008; McCulloch and Neuhaus, 2011a), considering two continuous covariates representing a linear time trend and a time by group interaction (Heagerty and Kurland, 2001), or considering two orthogonal time-varying continuous covariates (Neuhaus et al., 2013). However, in comparison to linear trends of time-varying covariates, transitions of categorical variables may be complicated by the direction of possible transitions or categories with absorbing states. For instance, transitions of the categorical variable capturing marital status is restricted by certain combinations (i.e. single never married $\rightarrow$ married $\rightarrow$ divorced is suitable, whilst the transition of single never married $\rightarrow$ divorced $\rightarrow$ single never married is not permissible). By considering two time-varying categorical variables with differing within-subject variability, the simulations considered in this study provide a novel insight into the impact of misspecifying random effect distributional assumptions on both time-varying covariates and time-varying categorical variables.

Inference of time-varying fixed effects in longitudinal studies is often considered the most relevant (McCulloch and Neuhaus, 2011a), as evaluation of time-varying covariates is often a primary reason for conducting longitudinal research (McCulloch and Neuhaus, 2013). Consistent with the results of theoretical studies and simulation studies considering time-varying continuous covariates, misspecification of the random effect distribution generally had minimal impact on the estimation of coefficients for continuous and categorical time-varying explanatory variables. However, the results of this study suggest that the inferential impact of misspecification on estimation of time-varying categorical variables may depend on the degree of within-subject variability. Categorical variables exhibiting minimal within-subject variability can result in biased estimation of the corresponding regression coefficient.

8.2.3 Impact on inference about the intercept constant

The results from the simulation studies suggest that misspecification of the random intercept distribution can induce biased estimates of the intercept constant. Incorrectly assuming normality of the random intercepts when the true distribution was an asymmetric mixture (Chapter 5) or a symmetric mixture with distinct modes (mean component distance 14 or larger, Chapter 6), resulted in overestimation and substantial bias of the intercept constant. The sensitivity to distributional misspecification when estimating the intercept constant is consistent with theoretical and simulation studies that have shown that estimation of parameters directly related to the misspecified random effect may be biased when the true distribution differs substantially from the assumed (Neuhaus et al., 1992; Heagerty and Kurland, 2001; Litiére et al., 2008; McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013). Furthermore, the bias observed for the asymmetric mixture distribution (Chapter 5) is consistent with the sensitivity of the intercept constant to assuming a symmetric distribution when the true distribution is asymmetric (Neuhaus et al., 1992). As previously reported (Heagerty and Kurland, 2001; McCulloch and Neuhaus, 2011a), larger departures from normality and large true random effect variances resulted in more positively biased estimates. Furthermore, coverage of the intercept constant estimate was impacted for large departures of the symmetric mixture distribution from normality (component mean distances 14 or larger), resulting in poor coverage rates below the nominal 95%. However, as previously reported by Neuhaus et al. (1992), the accuracy of model based standard errors was robust to distributional misspecification.

Typically inference for the intercept constant is not of direct interest, however bias of the intercept constant can transfer over to estimation of the mean value of the outcome variable[3] (McCulloch and Neuhaus, 2011a). The sensitivity of the intercept constant in the presence of misspecified random effects should be taken into account when inference focuses on the mean estimation of the outcome variable or the intercept constant.

8.2.4 Impact on inference about the random effects variability

Estimation and inference for the variability of the random intercept can be severely impacted by misspecification of the random effects distribution. Departures from the assumed normal distribution, as defined by multimodality of the true random intercepts either as an asymmetric mixture (Chapter 5) or symmetric mixture with distinct modality (mean component distance 12 or larger, Chapter 6), resulted in seriously biased estimates of the variance component, with extremely poor coverage below the nominal rate and inaccurate model based standard errors. The observed sensitivity is consistent with the bias reported for scenarios where the shape of the true distribution differs from that of the assumed distribution (Neuhaus et al., 1992; Litiére

[3]As estimated at fixed values of the explanatory variables included in the model.

et al., 2008). For instance, large bias has previously been reported when incorrectly assuming normality for random intercepts distributed as an exponential, chi-squared, power-function or an asymmetric mixture of two normals (Litiére et al., 2008). Larger true random effect variances resulted in larger bias in the estimation of the variance component, and are consistent with the findings of Litiére et al. (2008). However, for the symmetric mixture distribution considered in Chapter 6, smaller component variances for the same component mean distance, corresponding to more extreme multimodality, resulted in larger bias.

Previously it has been suggested that the direction of the bias can change depending on the true distribution (Litiére et al., 2008). However unlike the underestimation previously reported for true asymmetric two component mixture of normals (Litiére et al., 2008), assuming normality for both true asymmetric and symmetric mixtures resulted in overestimation of the true random effect variability. The assumed normal distribution appears to be more sensitive to departures from symmetry, with increased skewness in the true asymmetric mixture distribution having a greater impact. This concurs with the results presented by Litiére et al. (2008). The sensitivity of an asymmetric distribution may partly be explained by the skewness of the distribution, as large bias has been reported when true skewed distributions are incorrectly assumed to be symmetric (Neuhaus et al., 1992; Litiére et al., 2008; Neuhaus et al., 2013).

Estimates of the variance component are often not of primary inferential interest (McCulloch and Neuhaus, 2011a), however it is the only measure of the true random effect variability (Litiére et al., 2008). Bias of the variance component may subsequently impact alternative summary measures of the unobserved between-subject variability, such as the intra-class correlation[4], and may complicate identifying problems in the mean structure (Litiére et al., 2008). Furthermore, biased and inaccurate estimates of the variance components can also impact the accuracy of the best predicted random effect values, particularly for situations when the true distribution has a wider range of support than assumed, and for the random effect distribution with large variances (McCulloch and Neuhaus, 2011b). The results from the simulation studies suggest that if true random intercepts are suspected to differ from the assumed normal, either as an asymmetric mixture or a symmetric mixture with distinct modes, estimation and inference for the variance component and complementary summary statistics can be impacted.

8.2.5 Misspecification of bivariate random effects distribution

This study has predominately focused on misspecification in random intercept logistic models, however as demonstrated in Chapter 7, it may be appropriate to consider more complex random effect structures. The logistic mixed models presented in Chapter 7 indicate sensitivity to the assumed random effects distribution when estimating the effects of time-invariant

[4]For random intercept logistic models, large differences may arise when the true random intercept distributions has a small true random effect variability. As the ICC for the random intercept logistic model has a fixed residual error of $\pi^2/3$, similar ICC values will arise for true random intercept distributions with large variability unless there exists substantial bias.

explanatory variables and parameters relating to the random effect. Misspecification of the bivariate random effects distribution is an understudied area, with limited literature suggesting the impact on inferential conclusions is analogous to the random intercepts scenario (McCulloch and Neuhaus, 2011a; Neuhaus et al., 2013). Thus, there is scope for future work, particularly investigating misspecification within panel survey settings (by considering attrition and categorical explanatory variables) and situations with extreme distributions (multimodality as identified in this study). Simulation studies to quantify the impact on inference and prediction will be valuable, and have been highlighted as an area for ongoing research.

8.3 Impact of attrition

Missing data resulting from non-response and attrition is a complication of longitudinal data. In practical applications of GLMMs to longitudinal data, estimation can be affected by the simultaneous impact of misspecifying the random effects distribution in the presence of missing data. This was supported by the results presented in Chapter 4, whereby 29.5% of the HILDA study population experienced attrition and assuming a three component mixture for the random intercepts provided a more adequate fit than assuming normality in a random intercept logistic model. Although maximum likelihood estimation of GLMMs can provide consistent estimators when missing data is assumed to be missing at random (MAR), valid inference requires correct specification of the mean and variance-covariance structure of the dependent variable (Hedeker and Gibbons, 2006). One of the objectives of the simulation studies was to examine the effect of misspecification of random effects distributions on parameter estimates in settings with missing data due to MAR attrition.

The results presented in Chapters 5 and 6 suggest that, in the presence of MAR attrition, the impact of misspecifying the random intercept distribution was similar as that observed in the complete data scenario. In particular, distributional misspecification in the presence of MAR attrition generally resulted in similar trends and magnitudes of bias, coverage rates and accuracy of model based standard errors. The minimal additional effect of attrition was similar regardless of whether the true random intercepts were an asymmetric or symmetric mixture distribution. These results are consistent with the findings of Wang (2010b), whereby the interaction between misspecifying the bivariate random effects distribution and MCAR or MAR attrition had minimal impact on the estimated parameters and standard errors. Similarly, minimal additional impact of missingness was suggested by Neuhaus et al. (2013), when investigating misspecification of the bivariate random effects distribution for different cluster sizes. By considering cluster sizes uniformly distributed between 4 and 10, the scenario presented in Neuhaus et al. (2013) is consistent with a MCAR missingness mechanism.

The simulation studies have generated attrition assuming an underlying MAR missingness mechanism. However, as the effect of missing data on consistent estimation of GLMMs is dependent on the underlying reason for missingness, it would be valuable to explore the impact of other mechanisms. For instance, as the MNAR missingness leads to inconsistent estimation of GLMMs, the joint misspecification of the assumed missingness mechanism and random effects distribution is an area of future research. Furthermore, the simulation studies generated missing data with similar rates of wave-to-wave attrition as the 29.5% observed in the HILDA case study. However attrition rates can differ between longitudinal studies and populations under study. In the similar context of shared parameter models, the effect of misspecifying random effects distributions can become more pronounced for some parameters as the number of repeated longitudinal measurements per individual decreases (Rizopoulos et al., 2008). This is in contrast to results presented by Wang (2010b) reporting negligible differences in the impact of misspecified random effect distributions in logistic mixed models when considering three attrition rates of 10, 20 and 30%. Perhaps the interaction of higher attrition rates and substantial discrepancies of the true random effects distribution from normality are contributing factors that impact inference for model parameters in GLMMs. Preliminary results from additional simulation studies assessing the impact of 10, 30 and 50% attrition rates in the clinical trial setting presented in Chapter 6, support this idea. These results are only preliminary, and further research is required by considering different attrition rates and alternative missingness mechanisms, such as MNAR attrition and simultaneous intermittent missingness and attrition. Furthermore the simulation studies have considered a large number of 1000 individuals. Perhaps the simultaneous impact of missingness and misspecifying the assumed random effects distribution may be more pronounced in settings with fewer individuals, an area requiring further research.

To assess the feasibility of using more flexible modelling approaches to relax the parametric assumptions of the random effects distribution in practical applications, the sensitivity analysis in Chapter 7 considered applications of the logistic mixed model in settings with missing data due to attrition. Estimation of the flexible approaches applied to the logistic mixed models with univariate and bivariate random effects was not affected by the potential complexity associated with missing data due to attrition. This may partly be explained by the large number of individuals and the relatively large number of time-points in the HILDA case study (N=1927 and $n_i \leq 11$). Therefore, difficulties in regards to model convergence may occur in applications where limited data are available due to perhaps, fewer individuals and/or fewer time-points. Furthermore, loss of information due to intermittent missingness may also impact implementation and use of flexible random effects approaches. In addition to the performance of the flexible approaches within the panel survey setting, implementation of diagnostic tests in Chapter 4, highlights the practicality of informal and formal diagnostic tests to detect misspecification of the assumed random intercept distribution in the presence of attrition. Further investigation of the feasibility of flexible modelling approaches and diagnostic testing in practical applications should consider other types, mechanisms and rates of missingness.

In summary, the presence of MAR attrition in estimation of random intercept logistic models will have minimal additional impact when the random effects distribution is misspecified, provided the overall sample size is large. In particular, for similar rates of MAR attrition as in the HILDA case study, incorrectly assuming normality for true multimodal random intercept distributions can produce biased estimators of the fixed effects and parameters relating to the random intercept. By considering a range of departures from the assumed normal distribution, from minor to major departures in terms of symmetry and multimodal distributions, these results contribute to the limited literature assessing the effects of misspecification in the presence of attrition. In particular they provide an insight into the interaction of attrition and extreme random effect distributions due to potential mover-stayer scenarios in an application to panel survey data. Furthermore, implementation and estimation of approaches to induce more flexibility and detect misspecification of univariate and bivariate random effect distributions performed well in the presence of attrition in panel survey applications.

8.4 Implications to analysis of longitudinal household surveys in the social sciences

As the use of random effects in modelling increases (Agresti et al., 2004), understanding the implications of violating model based assumptions becomes crucial for accurate inference and interpretation. In practical applications of GLMMs, the conventional wisdom is that the choice of the random effects distribution is not critical (Agresti et al., 2004), and that the normality assumption is generally robust to misspecification (McCulloch and Neuhaus, 2011a). It has been shown that misspecifying the assumed random effect distribution has minimal impact on estimating the fixed effect parameters, typically the parameters of interest (McCulloch and Neuhaus, 2011a). This study shows that estimates of the fixed effects parameters from random intercept logistic models that incorrectly assume normally distributed random intercepts are generally robust to this misspecification, particularly in settings where the true underlying random effects reflect a mover-stayer scenario and in settings with missing data following the missing at random mechanism. For large departures from the assumed normal distribution characterised by distinct multimodality in a symmetric three component mixture of normal distributions, this study provides an example whereby misspecifying the random effects distribution in a two-level random intercept logistic model can produce biased estimates of parameters associated with the random effects in conjunction with time-varying and time-invariant fixed effect parameters. The impact of misspecification on estimating the time-varying parameters were restricted to categorical variables exhibiting minimal within-individual variability.

The total variance of the simulated random intercept distributions presented here may be considered more extreme than observed in practical applications of logistic mixed models[5]. Nonetheless, the illustrative example considered here was highly heterogeneous, resulting in over 50% of working aged women exhibiting constant response profiles over all eleven waves.

[5]Previously McCulloch and Neuhaus (2011a) considered simulated random intercepts with total random effect variance exceeding 4 as large (i.e. $\sigma_b^2 \geq 4$).

Taking account of the stability in the response profiles can be important, particularly for analysis of the labour market (Hyslop, 1999). The observed constant response profiles, particularly in the analysis of recurrent events, could be due to state dependence, heterogeneity in the population, or possibly a mixture of both (Davies, 1993; Hyslop, 1999). Therefore, estimation of GLMMs in applications with constant response profiles may be susceptible to misspecification of the random effects distribution. Initial exploration for constant response profiles prior to estimating the GLMM may help identify a potential non-normal random effects distribution. This highlights scenarios whereby the assumed normal distribution may not adequately capture the underlying heterogeneity and subsequently impact inferential conclusions.

Non-parametric estimation of the random effects provides a flexible approach to model the random effects, particularly in scenarios with substantial underlying heterogeneity due to a potential mover-stayer scenario. The similarity between the mover-stayer models and non-parametric characterisation of population heterogeneity has previously been noted by Davies and Crouchley (1986). In the context of a random intercept logistic model with no explanatory variables, the perceived goodness-of-fit success often observed for the three-spike Bernoulli mover-stayer model was argued to be due to it sufficiently approximating the non-parametric characterisation of the underlying heterogeneity (Davies, 1993). The results presented here suggest that the use of a non-parametric approach may accommodate an underlying mover-stayer scenario in logistic mixed models with explanatory variables. Although non-parametric estimation of the random effects distribution may be poor, it does not necessarily translate to poor estimation of the marginal distribution (Heckman and Singer, 1984; Agresti et al., 2004). Therefore, it is an appropriate method when the random effects distribution is not of primary interest (Heckman and Singer, 1984; Litiére et al., 2008). Albeit the increased computational burden and potential loss of efficiency for situations whereby the assumed normal distribution would not be badly violated (Agresti et al., 2004), non-parametric methods provide consistent estimation in data-rich settings (Chapter 12.4 in McCulloch et al., 2008). Hence, implementation of non-parametric approaches in longitudinal panel survey data may be a practical solution to relax the parametric assumptions and potential issues related to misspecifying the random effects distribution. Previously Agresti et al. (2004) and Muthén and Asparouhov (2008) among others, recommended more frequent use of non-parametric estimation of the random effects in practice. However, utilisation within panel survey settings remains limited. Perhaps the development and availability of computationally fast non-parametric maximum likelihood estimation methods in standard software, particularly for GLMMs with multiple random effects, will aid the implementation of non-parametric approaches for practical users of panel data.

As demonstrated in Chapter 7, implementation of non-parametric approaches to model the random effects within a sensitivity analysis framework can provide a practical way for users to gain confidence in the interpretation of their results. Specifically, comparing the model fit of a GLMM assuming normal random effects with the model fit of random effects fitted non-parametrically can efficiently validate the robustness of assuming normality (Agresti et al., 2004). However, model comparison based on the estimated fixed effect parameters of the parametric and non-parametric model can be misleading (Muthén and Asparouhov, 2008). Although the parameter estimates of the two approaches may be similar, higher order moments of data will be appropriately captured by non-parametric approaches (Muthén and Asparouhov, 2008). Furthermore, computational challenges associated with non-parametric approaches do not guarantee that the algorithm has converged to the maximum likelihood, and thus non-parametric approaches may lead to biased parameter estimates. Additionally, the performance of non-parametric methods is often dependent on the number of mass points, with a larger number of mass points generally resulting in improved model fit at the cost of increased computational power and issues of over-parametrisation[6]. However, the choice of the optimal number of mass points is not straight-forward. Methods such as the VEM and the Gateaux-derivative method can provide, respectively, either an approximation or an estimate of the non-parametric maximum likelihood estimator. The VEM approach had no convergence or estimation issues when implemented to estimate the more complex bivariate random effects distribution (Chapter 7), suggestive that VEM is a viable non-parametric approach in longitudinal panel applications. Simulation studies will be beneficial to further investigate the performance of VEM in applications of binary longitudinal data.

Often the greatest perceived limitation of the non-parametric approach is the discrete nature of the random effects distribution. However, if inferential interest is in density estimation of the random effects or making individual-specific predictions, alternative methodologies to flexibly model the random effects distribution are available(i.e. SNP-NLMM method by Vock et al., 2014). Furthermore, inclusion of these methods into the sensitivity framework can provide more evidence about the robustness of the assumed distribution (Agresti et al., 2004; Litiére et al., 2008).

However, flexible modelling of the random effects can be challenging in practice. Not only is the accessibility dependent on the availability of software packages, but it can also be restricted in regards to the complexity of the model (i.e. single or multiple random effects). Although the accessibility within standard statistical packages is improving, not all methods will induce sufficient flexibility to capture the extreme distribution in situations with constant response profiles. For instance, the default boundary of the assumed random effect distribution for the SNP-NLMM method (Vock et al., 2014) had limited support to capture the underlying heterogeneity of the random effects distribution. Similarly, when explanatory variables are included in the analysis, a random intercept logistic model with endpoints (Berridge and Crouchley, 2011a) may be too restrictive to capture the variability at the extremities of the distribution. In addi-

[6] For example, the number of mass points M can only be as large as the number of individuals N in the study. Or in the case of binary response data, M can only be as large as the number of distinct response and explanatory variable profiles. Therefore, in the case of a random intercept logistic model with no explanatory variables, M is restricted to be less than the number of observed distinct response profiles.

tion to the limitations of some methods to induce sufficient flexibility, flexible approaches may be susceptible to boundary solutions. For approaches where random effect distributions put appreciable probability on large magnitude values, such as the logistic model with endpoints or non-parametric approaches, boundary solutions can lead to instabilities of model parameters related to the random effects. Sensitivity of these approaches to the initial choices regarding the distributional range of the random effects should be assessed, and caution is required when interpreting parameters related to the random effects. Furthermore, implementation of the more flexible approaches may not be straightforward. For instance, using the likelihood reformulation method of Liu and Yu (2008) to model non-normal distributed random effects requires the user to reformulate the log-likelihood. Thus, correct specification of the reformulation model can be challenging in practice and requires users to understand statistical programming in SAS and theoretical aspects of the model.

Although development of diagnostic testing for distributional misspecification of random effects has recently received attention in the literature (i.e Verbeke and Molenberghs, 2013; Drikvandi et al., 2016), implementation in practice is limited. This may be partly explained by the accessibility of diagnostic tests in standard statistical software, with some methods requiring users to request syntax from the authors. As demonstrated in Chapter 4, the graphical exploratory tool of Verbeke and Molenberghs (2013) provides an easily implementable diagnostic tool to assess the distributional misspecification, however it is restricted by only using information from individuals with non-constant response profiles. The formal diagnostic test of Drikvandi et al. (2016) utilises information from all individuals, providing an easily implementable diagnostic test to identify distributional misspecification. However, both diagnostic tools were applied to the random intercept logistic model, and implementation to the bivariate random intercept and random slope scenario is not as straight forward.

The diagnostic tests considered here have focused on misspecification of the random effects distribution, nonetheless, a suite of alternative diagnostic tests is available that assess overall fit (i.e. Pan and Lin, 2005; Alonso et al., 2010b) and other aspects of the model specification (i.e. Pan and Lin (2005) describe a test for the functional form of the explanatory variables and also a test for the adequacy of the link function). Increasing the implementation of diagnostic tools in practice will be beneficial to formally detect violations of model assumptions, and thus, understand the potential impact on inferential conclusions. Sensitivity analyses can provide an informal and easily implementable tool to assess robustness of statistical methodology. This is not only limited to the distributional assumptions of the random effects, but can consider sensitivity to assumptions of underlying missing data mechanisms and other GLMM assumptions.

8.5 Computational issues

Longitudinal analysis of categorical responses using GLMMs can suffer from computational difficulties. Evaluating the likelihood is complicated by the calculation of high dimensional integrals (McCulloch et al., 2008). Therefore, not only can the choice of numerical approximation techniques impact estimation, but numerical methods can be sensitive to the number of clusters and cluster size (Hosmer et al., 2013). In this study, maximum likelihood estimation has been restricted to numerical integration techniques utilising adaptive Gaussian quadrature[7]. Adaptive quadrature is a reliable method to estimate GLMMs, however it is still susceptible to numerical issues (Hosmer et al., 2013; Capanu et al., 2013) and convergence to a global maximum can be difficult to obtain (Lesaffre and Spiessens, 2001). Although adaptive quadrature is not as sensitive to the number of quadrature points as non-adaptive quadrature (Lesaffre and Spiessens, 2001), the sensitivity to the choice of quadrature points should routinely be examined (Lesaffre and Spiessens, 2001; Hosmer et al., 2013). More quadrature points improves approximation of the log-likelihood (Capanu et al., 2013), however increasing the number of quadrature points can subsequently lead to numerical convergence issues (Hosmer et al., 2013). For the logistic mixed models assuming normal distributed random effects considered in this study, estimation was based on 20 adaptive quadrature points. Sensitivity analyses suggest stability of the parameter estimates and that 20 quadrature points provided a good approximation of the log-likelihood (results not shown).

Implementation of GLMMs in alternative software packages often differ in default settings, including the number of quadrature points and numerical method used for estimation. For instance, eight quadrature points is the default for the GLLAMM procedure in STATA, whilst the NLMIXED procedure in SAS adaptively selects the number of points. Not only does the default number of quadrature points differ, the default initial starting values and optimization procedure can vary between software procedures. The different choice of these aspects can impact inferential interpretation (as demonstrated by Lesaffre and Spiessens (2001) and Chapter 17 of Molenberghs and Verbeke (2005)), however for relatively large datasets the differences are generally minimal (Li et al., 2011). To ensure reliability of GLMM estimates, users should consider the stability of parameter estimates and model convergence to the choice of user-specified options and default settings of software packages. Furthermore, to enable reproducibility of results, users should include details of the software package and user-specified options.

As for GLMMs assuming Gaussian distributed random effects, approaches to flexibly model the random effects distribution may be sensitive to the choice of input parameters and starting values, resulting in unstable model parameters and failed model convergence. For instance, the likelihood reformulation method to model the random effects as a finite mixture of normal distributions was extremely sensitive to the initial starting values and to the number of quadrature

[7] Capanu et al. (2013) suggest that adaptive Gaussian quadrature is a good choice for settings with non-complex random effects structure and a moderate to large number of observations per random effect.

points used in the estimation (see Appendix B for further details). Model convergence issues can be more susceptible in more complex scenarios, such as modelling multiple random effects. For instance, in the bivariate random effects scenario considered in Chapter 7, both the NPMLE as estimated using the Gateaux derivative in GLLAMM and the SNP-NLMM estimated in SAS failed to converge or obtain standard errors for all model parameters (see Section 7.4.2 for details). Convergence issues are often noted in the output (Hosmer et al., 2013), however it is up to the user to assess the model output and be wary of model convergence when interpreting estimated coefficients. The sensitivity of flexible approaches to user-specified options should be assessed to ensure stability and numerical convergence.

Increases in computational power and memory allow estimation of GLMMs with normal and non-normal random effects to be implemented in practice. However, as highlighted in Chapter 7, the computational burden required to flexibly model the random effects distribution can vary substantially depending on the methodology and the number of random effects included in the model. In these scenarios, the use of high performance computers and parallel computing may alleviate some of the computational burden. The use of high performance computers was particularly beneficial when implementing the VEM to estimate the logistic mixed models with univariate and bivariate random effects. Furthermore, high performance computing and parallel processing were beneficial to reduce the computational burden associated with performing simulation studies in panel survey settings (Chapters 5 and 6).

8.6 Statistical and simulation methodology and appropriateness of measures to assess adequacy of model fit

8.6.1 Statistical methodology

This study has implicitly assumed that the observed constant response profiles in the HILDA case study are due to heterogeneity in the population. Furthermore, this study has assumed that the heterogeneity can be adequately captured by the random effects distribution in logistic mixed models. However, as constant response profiles may be due to a combination of state dependence and population heterogeneity (Davies, 1993; Hyslop, 1999), models that can additionally account for state dependence, such as transition models, may more appropriately disentangle the two effects. Therefore, it would be intriguing to investigate non-normality, and potentially multimodality, of the random effects distribution in transition models applied to mover-stayer scenarios. Furthermore, if interest was in quantifying the substructure of the population, implementation of latent mixture models may be beneficial.

Estimation of GLMMs can be sensitive to the coding of explanatory variables in the mean structure of the model. For instance, the constant intercept coefficient is interpreted as the expected value of the outcome when all explanatory variables have the value zero (Hox, 2010).

If the value zero is not a viable value for an explanatory variable, then subsequently, the value of the intercept is meaningless. Often the intercept is not of primary interest, though centering or transforming explanatory variables to make a zero value legitimate can aid model interpretation. For GLMMs that do not have random coefficient terms, i.e. random intercept models, the model is invariant to linear transformations (Hox, 2010). Thus, in the case study applying a random intercept logistic model to women aged 30 to 44 (at the first wave), changing the age term to be the difference of age and 30 years, will enable the value zero to be meaningful. However, with the exception of the constant coefficient, transforming the age term to start at zero will result in no differences of the coefficient estimates or the estimated random intercept variance (results not shown). Furthermore, by including age as a single term in the random intercept logistic models (Equation 4.1), the term captures two processes: the time-invariant effect of age at the first wave, and the time-varying effect over time as captured by the waves of the survey. Thus, the age term included in the random intercept logistic models may be confounded. Reparameterisation of the age term into the two terms (time-invariant initial age at the first wave, and a time-varying wave term) marginally alters the resulting coefficient estimates, and suggests no significant differences in initial age and a significant linear trend for each wave (results not shown).

The explanatory variables considered in the case study capture a selection of variables that are commonly utilised in modelling labour force participation (i.e. Jenkins, 2006; Parr, 2012; Tannous and Smith, 2013). Even when a more complex model, containing more explanatory variables, is considered, there still exists multimodality in the random effects (results not shown). Thus, albeit the smaller number of explanatory variables included in the motivating case study, the multimodality of the random intercepts, potentially due to the latent mover stayer scenario, is expected to be observed in practice.

8.6.2 Simulation methodology

Simulation studies aim to generate datasets with similar properties and resemblance as the original data (Burton et al., 2006). However, simulating longitudinal data to effectively preserve the correlation and temporal changes of the explanatory variables can be challenging. Methods have been developed to simulate fixed effects explanatory variables from multivariate distributions (Wicklin, 2013), however, generating correlated categorical explanatory variables can be complicated by considering both the within- and between-variable correlations. As the simulated longitudinal data in this study generates two time-varying categorical variables and one time-invariant categorical variable, resampling individuals from the HILDA data set and using their explanatory variables provides an adequate method to capture the within- and between-covariate variability. By using the HILDA data to generate the simulated data, caution is required in generalising the results to other longitudinal panel surveys and in other applications of the random intercept logistic model. However, the results for misspecification within a clinical trial setting (Appendix E) suggest similar results as presented for the longitudinal panel survey setting (Chapter 6). Furthermore, the simulation study within the clinical trial setting generated explanatory variables using multivariate normal distributions, and as such, gave support that the resampling methods utilised to simulate longitudinal panel survey data were appropriate. Additionally, all simulation studies had high convergence rates, providing increased confidence in the simulation results.

The performance measures used to assess the impact of misspecifying the assumed random intercept distribution in the simulation study are similar to the measures utilised in Neuhaus et al. (2013). The relative bias has been calculated using the average parameter estimate over the total number of simulations. However, as the sampling distribution of parameter estimates may be skewed, the median as utilised by Neuhaus et al. (2013) may provide a more accurate summary measure. Utilising the median as the summary measure for the parameter estimates will result in minimal differences in regards to substantial bias (1% and 2.3% of simulated scenarios in Chapter 5 and 6, respectively). Similarly, using the median of the standard error estimates to calculate the standard error ratio would result in minimal differences in regards to the accuracy of model based standard errors (1% and 0.2% of simulated scenarios in Chapters 5 and 6, respectively). As skewness of the sampling distributions may be related to a smaller number of simulated datasets (Chapter 15 of Green 2012), the minimal differences in overall conclusions between the median- and mean-based summary estimates were to be expected. This provides support that the large number of Monte Carlo simulations considered in the simulation studies is sufficient.

Previously it has been argued that misspecification of the random effects distribution requires simulating a single true random effect distribution and varying the assumed distribution (Neuhaus et al., 2011). By simulating a variety of true underlying random intercept distributions and only considering the normality assumption, it may be argued that the simulation studies presented here merely assess the robustness of the normality assumption (Neuhaus et al., 2011). Albeit the underlying differences, Litiére et al. (2011) considers the two approaches to be complementary. Furthermore, the choice of the approach may be dependent on the complexity of the model, and practical issues such as computational burden. For instance, to investigate distributional misspecification of bivariate random effects, Neuhaus et al. (2013) considered the impact of incorrectly assuming a bivariate normal distribution and simulated a range of true bivariate random effect distributions. Thus, to supplement the simulation results presented here, further work could consider fixing the true multimodal random effect and consider alternative assumed distributions, such as finite mixture distributions or non-parametric estimation techniques. Furthermore, to assess the consistency of the results presented here, alternative magnitudes of coefficient values for the explanatory variables could be considered (i.e. as considered by Neuhaus et al., 2013).

8.6.3 Model fit and comparison

Comparing the adequacy of the model fit for models with alternative random effect distributions is not straight-forward, as there is generally no unrestricted model that can be used for comparison (Muthén and Asparouhov, 2008). As utilised by McCulloch and Neuhaus (2011a) and Neuhaus et al. (2013), the residual deviance, calculated as the negative of twice the log-likelihood, has been used as an indicative measure of the model fit. However, comparison tools based on the estimated log-likelihood have limitations. The reported log-likelihood value corresponds to the maximum of the approximation to the log-likelihood (Molenberghs and Verbeke, 2005), and as such, is dependent on the number of quadrature points and estimation techniques (Lesaffre and Spiessens, 2001; Molenberghs and Verbeke, 2005). This implies that the maximised log-likelihood values from different models are not necessarily comparable (Molenberghs and Verbeke, 2005), and that the differences may reflect the quality of the technique to obtain a close approximation to the model likelihood. Further, as the maximised likelihood refers to the marginal distribution, different random effects distributions can generate similar marginal distributions (Agresti et al., 2004). Thus, assessing the fit of models by comparing log-likelihood values (and subsequently, residual deviance) may not identify a better model fit unless the marginal distributions differ substantially (Agresti et al., 2004).

Alternatively, information criteria could be used to compare non-nested models. Information criteria compare models based on their maximised log-likelihood value, but penalise for the complexity of the model (i.e. number of parameters and number of individuals). To account for the differing number of parameters used to estimate the models with different assumed random effect distributions, criteria such as the AIC or BIC could be utilised (i.e. as implemented by Litiére et al., 2008). However as information criteria were originally developed for standard linear models, additional challenges arise when applying the criteria to multilevel models (Steele, 2013). For instance, specification of the number of model parameters is not straight forward (Steele, 2013), specifically for models with flexible random effects. Furthermore, as discussed for the residual deviance, information criteria may not detect differences between alternative assumed random effects distributions unless the marginal distributions are quite different (Agresti et al., 2004). To overcome this issue, performance criteria can be calculated using the conditional distribution by conditioning on the random effects distribution (Agresti et al., 2004). The conditional AIC has been developed for linear mixed models (Vaida and Blanchard, 2005) and development of conditional AIC for GLMMs is an area of ongoing research. However, model selection based on information criteria is not straight forward, as alternative information criteria may lead to different conclusions (Chapter 15 of Molenberghs and Verbeke, 2005). Furthermore, there is not one criterion that can be considered the best (Steele, 2013), as the choice is often dependent on the objective of the model comparison (Müller et al., 2013).

8.7 Limitations and scope for further research

This study has investigated the impact of misspecifying and inducing more flexible distributions for the random effects distribution in logistic mixed models when applied to panel survey data. However, there are avenues and considerations that require further research.

Firstly, the logistic mixed models considered in this study are too simple to realistically address questions about employment participation of Australian working aged women. More appropriate analyses would consider more than two employment states by distinguishing between part-time and full-time employment, and also distinguishing between unemployment and not in the labour force. By considering the four employment states, multinomial or ordinal logistic mixed models could still be susceptible to subjects exhibiting constant response profiles, and hence, an extreme random effects distribution. Distributional misspecification of the random effects in multinomial or ordinal logistic mixed models have received little attention in the literature. A limited simulation study presented by Hartzel et al. (2001) suggests minimal impact on fixed effect coefficients in random intercept ordinal logistic models. More research into inferential impact on misspecification in multinomial or ordinal logistic mixed models is required, particularly by considering similar scenarios as identified in the HILDA case study, such as attrition and potential multimodality of the random effects distribution. Furthermore, it would be of interest to assess the performance of approaches to induce more flexible random effects distributions in GLMMs for categorical response variables with more than two categories. It is of particular interest to investigate whether the VEM can sufficiently capture the potentially complex underlying random effects distribution when applied to multinomial or ordinal mixed models.

Secondly, the predominant focus of this study has been to investigate the impact of misspecifying the random effects distribution in logistic mixed effects models. In doing so, it has been implicitly assumed that other aspects of the model have been correctly specified. That is, it has been assumed that the mean structure of the model has been correctly specified, and that, with the exception of the distributional assumption, the random effects structure has also been correctly specified. However in practice, all model assumptions are violated, at least to a minor degree (McCulloch and Neuhaus, 2011a). Therefore, simultaneous misspecification of the random effect distribution and other model assumptions can occur in practice. Research investigating inferential impact of simultaneous misspecification is limited, and as discussed in more detail below, is an important area for future research.

The explanatory variables considered in the case study were a selection of variables commonly considered when modelling labour force participation. The restricted selection of explanatory variables does not extensively capture the phenomenon under study, and therefore, an important explanatory variable (or variables) could potentially be omitted from the mean structure. Furthermore, inability to capture information about certain relevant factors within the HILDA panel survey may also contribute to an important variable to be omitted from the mean structure of the model. The effects of omitting key explanatory variables in GLMMs has not been extensively investigated in the literature (McCulloch et al., 2008). Biased estimates of the coefficients in the mean structure can occur when the omitted explanatory variable is correlated with other explanatory variables in the mean structure (McCulloch et al., 2008). Furthermore, biased estimates of the included explanatory variables can occur when the omitted variable is independent of the other explanatory variables (Neuhaus and Jewell, 1993; McCulloch et al., 2008). As the random effects capture the heterogeneity of unobserved time-invariant variables, omitted variables may subsequently impact the random effect distribution. It has been suggested that if the distribution of the omitted time-invariant covariate[8] is different from the distribution of the random effects, the convolution of the distributions will not equal the random effects distribution, nor be in the same family of distributions (McCulloch et al., 2008). Thus, omitting a time-invariant covariate could be considered as misspecifying the random effects distribution (McCulloch et al., 2008). However, as omitted time-invariant categorical variables could result in polarisation (and hence, multimodality) of the random effects distribution (Agresti et al., 2004), further research should explore if the results presented here generalise to omitted time-invariant categorical variables.

Incorrect specification of the mean structure may also occur when a covariate is incorrectly assumed to have a simple linear relationship when a non-linear or more complex relationship exists. In the models considered in this study, the age term has been included as a linear term. However, it may be more appropriate to additionally include a quadratic or higher-order term to capture a non-linear changes over time. Furthermore, misspecification of the mean structure can also occur when interaction terms are excluded. Perhaps incorrect specification of these aspects of the mean structure will subsequently be captured by the random effects. Therefore, it would be beneficial to assess distributional aspects of the random effects in scenarios when the mean structure is incorrectly specified due to omission of non-linear time trends and interaction terms. Furthermore, by leaving the random effects distribution unspecified, non-parametric estimation of the random effects distributions may capture model misspecification in these scenarios. This is an area requiring further work and investigation.

Aspects of the random effect structure are also susceptible to misspecification. Incorrect specification can occur when a fixed effect for an explanatory variable should have a corresponding random effect, thus incorrectly omitting a random effect. As identified in Chapter 7, the substantial improvement in the residual deviance from considering univariate random effects to bivariate random effects indicates that the random effects structure may be misspecified when only random intercepts are included in the model. Previously, Heagerty and Kurland (2001) reported that incorrectly using a fixed coefficient for a time-varying continuous explanatory variable in a logistic mixed effects model, could result in bias of up to 30 to 50%

[8]It has been presumed that the term covariate used by (McCulloch et al., 2008) in this context refers to a time-invariant continuous explanatory variable.

in estimated regression coefficients related to the explanatory variable (including interaction terms). Additionally, the random effects structure can be misspecified by incorrectly assuming the unobserved heterogeneity is time-invariant. The assumption that the latent time-invariant variable is constant over time can be restrictive in practical applications. For instance, in the case of employment participation, the unobserved propensity to take on employment may vary over time due to changes in perceptions and attitudes to employment. Heagerty and Kurland (2001) reported biased estimates of the fixed effect coefficients and the random intercept variance when autocorrelated random intercepts were incorrectly assumed to be time-invariant. These aspects of misspecification require further work, particularly as these may additionally contribute to distributional misspecification, especially in applications of logistic mixed models with constant response profiles.

Furthermore, it has been assumed that the random effects are uncorrelated with the explanatory variables. However, in practice this assumption is often violated (Neuhaus and McCulloch, 2014). Non-zero correlations can occur as the random effects may include omitted covariates that are associated with both the response and the explanatory variables. Ignoring correlations between random effects and explanatory variables in a random intercept logistic model can produce biased estimates of the parameter coefficients and other model parameters (Neuhaus and McCulloch, 2006). Conditional likelihood (fixed effects models, Section 2.3.2.1) and hybrid approaches (decomposition methods, Section 2.3.2.2) can provide consistent estimates of the within-individual effects in settings where the random effects are correlated with time-varying explanatory variables (Neuhaus and McCulloch, 2006). However, the hybrid model produces biased estimates for the other model parameters, including both the intercept constant and random effect variance estimate (Neuhaus and McCulloch, 2006; McCulloch et al., 2008). Additionally, Neuhaus and McCulloch (2014) recently reported inconsistent estimation of conditional likelihood and hybrid models in situations where the random effects were correlated with the explanatory variables in the presence of MAR attrition. However, by decomposing the time-varying explanatory variable into the baseline value and the change over time from baseline, consistent and unbiased estimation of the within-individual effects can be produced in the presence of MAR attrition (Neuhaus and McCulloch, 2014). As the correlation between the random effects and explanatory variables can also be viewed as misspecifying the distribution of the random effects (Neuhaus and McCulloch, 2006), further work could compare the performance of decomposition methods[9] and approaches to flexibly model the random effects distribution when applied to panel survey data in the presence of MAR attrition. In particular this further work could consider non-parametric estimation of the random effects distribution, as leaving the distributional assumptions unspecified may provide an avenue to capture this type of model misspecification.

[9]Decomposing the time-varying explanatory variables using the observed baseline value is suggested by Neuhaus and McCulloch (2014). However, the work presented by Neuhaus and McCulloch (2014) only considers continuous time-varying variables. Further work is required in regards to categorical time-varying explanatory variables to determine whether it is feasible to treat them as continuous variables.

Thirdly, the focus of this study has been on maximum likelihood techniques for estimation. However, the hierarchical model formulation of the GLMM makes estimation within a Bayesian framework very appealing (Molenberghs and Verbeke, 2005; Lynch, 2007). Bayesian approaches treat all unknown parameters as random, assuming they are distributed according to a prior distribution. This provides a framework to induce flexibility to model the random effects. For instance, Bayesian Markov Chain Monte Carlo (MCMC) techniques have been developed to estimate the heterogeneity model (Ho and Hu, 2008). Furthermore, extensions of the heterogeneity model based on penalised Gaussian mixtures (Komarek and Lesaffre, 2008) and Dirichlet processes (Kleinman and Ibrahim, 1998; Jara et al., 2007) have also been developed. With the increasing accessibility of Bayesian methods and the increasing computational power to analyse longitudinal panel data, it would be valuable to explore the practicality and performance of Bayesian approaches to account for multimodal distributions within potential mover-stayer scenarios.

Finally, the heterogeneity of the underlying random intercept distribution in the HILDA case study has been postulated to be captured and represented by a three component mixture of normal distributions. However, this may be too simplistic. In a mover-stayer scenario, if the observed stayers consist of latent stayers and latent movers that have yet to transition, perhaps a five component mixture would more appropriately capture the heterogeneity. Thus, the heterogeneity model utilised here is limited by assuming the number of components is known apriori. Different models with varying number of mixture components could be considered, with the optimal number selected using goodness-of-fit tests conditional on maximum likelihood estimates (Verbeke and Lesaffre, 1996) or information criteria (Proust and Jacqmin-Gadda, 2005). However, not only do these approaches ignore the uncertainty in estimating the optimal number of components, the computational burden to estimate more than three components in the panel survey application is expected to be intensive. Approaches have been developed which implement Bayesian MCMC schemes capable of comparing models with a different number of components, such as reversible jump MCMC (RJMCMC) methodology (Richardson and Green, 1997), allowing the number of components and model parameters to be simultaneously estimated (Watier et al., 1999; Ho and Hu, 2008). However, these RJMCMC methods can be cumbersome as they often converge slowly (Carlin and Louis, 2000) and have currently only been developed for linear mixed models. Variational Bayes methods could provide an alternative to RJMCMC methods to jointly estimate the model parameters and estimate the optimal number of mixture components (McGrory and Titterington, 2007). This is an avenue that has been highlighted for future research.

8.8 Concluding remarks

As the use of longitudinal panel data increases in the health and social sciences, there is a growing need for the appropriate use and understanding of underlying assumptions of statistical models. The accuracy of model based inference is crucial for researchers and policy makers utilising results to formulate and evaluate policy initiatives. This study provides a novel insight into the impact of assuming the random effects follow a normal distribution in logistic mixed models applied to panel survey data where an underlying sub-population structure, such as a mover-stayer scenario, exists. For departures from the normal distribution characterised by multimodality with three distinct modes and skewness, incorrectly assuming normality in random intercept logistic models produced biased estimates, poor coverage rates of the confidence intervals and inaccurate model based standard errors for the intercept constant and the random intercept variance component. Estimation of the fixed effects parameters, typically the parameters of interest, is generally robust to misspecification. However for large departures from normality characterised by multimodality with three distinct modes, incorrectly assuming normality for the random effects in a random intercept logistic model can result in biased estimation of the coefficients capturing the effects of time-invariant categorical explanatory variables and time-varying categorical explanatory variables exhibiting minimal within-subject variability.

Misspecification in the presence of MAR attrition had negligible additional inferential impact. Using more flexible distributions for the random effects can provide a practical solution to reduce the impact of violating distributional assumptions in logistic mixed models. Utilisation of these approaches within a sensitivity analysis framework, can provide an easily implementable solution to identify potential misspecification of univariate or bivariate random effects in practice. In applications to panel survey data, including in the presence of attrition, the VEM algorithm of Böhning (1985) induced increased flexibility to capture multimodality of random intercepts. Furthermore, it was the only approach in comparison to existing approaches to converge and capture the complexity of the bivariate random effects. The performance of the VEM to flexibly model the random effects in logistic mixed models reported here should encourage its implementation in different applications to health and social sciences.

References

ABS (2001). Labour Statistics: Concepts, Sources and Methods, Cat 6102.0. Canberra: Australian Bureau of Statistics.

ABS (2007). Labour Force Historical Timeseries, Australia, 1966 to 1984, Cat 6204.0.55.001. Canberra: Australian Bureau of Statistics.

ABS (2014). Labour Force, Australia, Detailed, Electronic Delivery, Cat 6291.0.55.001. Canberra: Australian Bureau of Statistics.

Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons, New Jersey, 3 edition.

Agresti, A., Caffo, B., and Ohman-Strickland, P. (2004). Examples in which misspecification of a random effects distribution reduces efficiency, and possible remedies. *Computational Statistics & Data Analysis*, 47(3):639–653.

Aitkin, M. (1999). A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128.

Alderman, H., Behrman, J. R., Kohler, H.-P., Maluccio, J. A., and Watkins, S. C. (2001). Attrition in longitudinal household survey data. *Demographic Research*, 5(4):79–124.

Allison, P. D. (2009). *Fixed Effects Regression Models*. SAGE Publications, Inc., Thousand Oaks.

Alonso, A., Geys, H., Molenberghs, G., Kenward, M. G., and Vangeneugden, T. (2004). Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics*, 60(4):845–853.

Alonso, A., Litiere, S., and Laenen, A. (2010a). A note on the indeterminacy of the random-effects distribution in hierarchical models. *The American Statistician*, 64(4):318–324.

Alonso, A., Litiere, S., and Molenberghs, G. (2008). A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models. *Computational Statistics & Data Analysis*, 52(9):4474–4486.

Alonso, A., Litiere, S., and Molenberghs, G. (2010b). Testing for misspecification in generalized linear mixed models. *Biostatistics*, 11(4):771–786.

Andress, H.-J., Schmidt, A. W., and Golsch, K. (2013). *Applied Panel Data Analysis for Economic and Social Surveys*. Springer, Berlin.

Austen, S. and Ong, R. (2010). The employment transitions of mid-life women: health and care effects. *Ageing & Society*, 30(2):207–227.

Baghfalaki, T. and Ganjali, M. (2014). A fast algorithm for using semi-parametric random effects model for analyzing longitudinal data. *Statistics, Optimization & Information Computing*, 2(4):339–351.

Barry, J., Francis, B. J., and Davies, R. B. (1989). SABRE: Software for the analysis of binary recurrent events. In Decarli, A., Francis, B. J., Gilchrist, R., and Seeber, G., editors, *Statistical Modelling: Proceedings of GLIM 89 and the 4th International Workshop on Statistical Modelling*, pages 56–63, Trento, Italy. Springer Science & Business Media.

Baxter, J. and Renda, J. (2011). Lone and couple mothers in the Australian labour market: differences in employment transitions. *Australian Journal of Labour Economics*, 14(2):103–122.

Bell, A. and Jones, K. (2015). Explaining fixed effects: random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods*, 3(1):133–153.

Berridge, D. M. and Crouchley, R. (2011a). Chapter 13: Stayers, non-susceptibles and end-points. In *Multivariate Generalized Linear Mixed Models using R*, pages 157–168. CRC Press, Boca Raton.

Berridge, D. M. and Crouchley, R. (2011b). *Multivariate Generalized Linear Mixed Models using R*. CRC Press, Boca Raton.

Blumen, I., Kogan, M., and Holland, P. J. (1955). *The Industrial Mobility of Labor as a Probability Process*. Cornell University Press, New York.

Böhning, D. (1985). Numerical estimation of a probability measure. *Journal of Statistical Planning and Inference*, 11(1):57–69.

Böhning, D. (1999). *Computer-assisted Analysis of Mixtures and Applications: Meta-analysis, Disease Mapping and Others*. Chapman and Hall, London.

Bonate, P. L. (2011). *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer Science and Business Media, New York.

Burniaux, J., Duval, R., and Jaumotte, F. (2003). Coping with ageing: a dynamic approach to quantify the impact of alternative policy options on future labour supply in OECD countries. Technical Report 371, OECD Economics Department Working Papers, Paris.

Burton, A., Altman, D. G., Royston, P., and Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25(24):4279–4292.

Butler, S. M. and Louis, T. A. (1992). Random effects models with non-parametric priors. *Statistics in Medicine*, 11(14-15):1981–2000.

Caffo, B., An, M.-W., and Rohde, C. (2007). Flexible random intercept models for binary outcomes using mixtures of normals. *Computational Statistics & Data Analysis*, 51(11):5220–5235.

Cai, L. and Kalb, G. (2006). Health status and labour force participation: evidence from Australia. *Health Economics*, 15(3):241–261.

Capanu, M., Gnen, M., and Begg, C. B. (2013). An assessment of estimation methods for generalized linear mixed models with binary outcomes. *Statistics in Medicine*, 32(26):4550–4566.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis, Second Edition*. Chapman & Hall/CRC, New York, 2 edition.

Carlin, J. B., Wolfe, R., Brown, C. H., and Gelman, A. (2001). A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes. *Biostatistics*, 2(4):397–416.

Chen, J., Zhang, D., and Davidian, M. (2002). A monte carlo EM algorithm for generalized linear mixed models with flexible random effects distribution. *Biostatistics*, 3(3):347–360.

Clark, T. S. and Linzer, D. A. (2015). Should I use fixed or random effects? *Political Science Research and Methods*, 3(2):399–408.

Cook, R. J. and Lawless, J. F. (2014). Statistical issues in modeling chronic disease in cohort studies. *Statistics in Biosciences*, 6(1):127–161.

Crouchley, R. (2007). *sabreR: Provides SABRE functionality from within R*. R package version 1.0.

Crouchley, R. and Davies, R. B. (2001). A comparison of GEE and random effects models for distinguishing heterogeneity, nonstationarity and state dependence in a collection of short binary event series. *Statistical Modelling*, 1(4):271–285.

Davies, R., Elias, P., and Penn, R. (1992). The relationship between a husband's unemployment and his wife's participation in the labour force. *Oxford Bulletin of Economics & Statistics*, 54(2):145–171.

Davies, R. B. (1993). Nonparametric control for residual heterogeneity in modelling recurrent behaviour. *Computational Statistics & Data Analysis*, 16(2):143–160.

Davies, R. B. and Crouchley, R. (1986). The mover-stayer model requiescat in pace. *Sociological Methods & Research*, 14(4):356–380.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (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. and Kenward, M. G. (1994). Informative drop-out in longitudinal data analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 43(1):49–93.

Drikvandi, R., Verbeke, G., and Molenberghs, G. (2016). Diagnosing misspecification of the random-effects distribution in mixed models. *Biometrics*, Accepted.

Efendi, A., Drikvandi, R., Verbeke, G., and Molenberghs, G. (2014). A goodness-of-fit test for the random-effects distribution in mixed models. *Statistical Methods in Medical Research*, page Advance online publication. doi:10.1177/0962280214564721.

Farewell, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics*, 38(4):1041–1046.

Fedorov, V. V. (1972). *Theory of optimal experiments*. Academic, New York.

Feeny, S., Ong, R., Spong, H., and Wood, G. (2012). The impact of housing assistance on the employment outcomes of labour market programme participants in Australia. *Urban Studies*, 49(4):821–844.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (2009). *Longitudinal Data Analysis*. CRC Press, Boca Raton.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*. John Wiley & Sons, Hoboken, 2 edition.

Follmann, D. and Wu, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, 51(1):151–168.

Follmann, D. A. and Lambert, D. (1989). Generalizing logistic regression by nonparametric mixing. *Journal of the American Statistical Association*, 84(405):295–300.

Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge University Press, New York.

Frijters, P., Johnston, D. W., and Shields, M. A. (2014). The effect of mental health on employment: Evidence from Australian panel data. *Health Economics*, 23(9):1058–1071.

Gallant, A. R. and Nychka, D. W. (1987). Semi-nonparametric maximum likelihood estimation. *Econometrica*, 55(2):363–390.

Ghidey, W., Lesaffre, E., and Verbeke, G. (2010). A comparison of methods for estimating the random effects distribution of a linear mixed model. *Statistical Methods in Medical Research*, 19(6):575–600.

Gibbons, R. D., Hedeker, D., and DuToit, S. (2010). Advances in analysis of longitudinal data. *Annual Review of Clinical Psychology*, 6:79–107.

Goldstein, H. (1979). *The Design and Analysis of Longitudinal Studies: Their Role in the Measurement of Change*. Academic Press, London.

Goldstein, H. (2011). *Multilevel Statistical Models*. John Wiley & Sons, Chichester, 4 edition.

Goldstein, H. and Healy, M. J. R. (1995). The graphical presentation of a collection of means. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 158(1):175–177.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association*, 56(296):841–868.

Graham, J. W. (2012). *Missing Data: Analysis and Design*. Springer, New York.

Green, W. (2012). *Econometric Analysis*. Prentice Hall, Boston, 7 edition.

Hartford, A. and Davidian, M. (2000). Consequences of misspecifying assumptions in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, 34(2):139–164.

Hartzel, J., Agresti, A., and Caffo, B. (2001). Multinomial logit random effects models. *Statistical Modelling*, 1(2):81–102.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46(6):1251–1271.

Heagerty, P. J. and Kurland, B. F. (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika*, 88(4):973–985.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, 52(2):271–320.

Hedeker, D. and Gibbons, R. D. (2006). *Longitudinal Data Analysis*. Wiley, Hoboken.

Ho, H. J. and Lin, T.-I. (2010). Robust linear mixed models using the skew t distribution with application to schizophrenia data. *Biometrical Journal*, 52(4):449–469.

Ho, R. K. W. and Hu, I. (2008). Flexible modelling of random effects in linear mixed models-a Bayesian approach. *Computational Statistics & Data Analysis*, 52(3):1347–1361.

Honey, A., Kariuki, M., Emerson, E., and Llewellyn, G. (2014). Employment status transitions among young adults, with and without disability. *Australian Journal of Social Issues*, 49(2):151–170.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley, Hoboken, 3 edition.

House of Representatives Standing Committee on Employment and Workplace Relations (2009). Making it fair: Pay equity and associated issues related to increasing female participation in the workforce. Technical report, Australian Parliament, Canberra. House of Representatives Committees.

Hox, J. (2010). *Multilevel Analysis : Techniques and Applications*. Taylor and Francis, Hoboken, 2 edition.

Hsiao, C. (2007). Panel data analysis - advantages and challenges. *TEST*, 16(1):1–22.

Huang, X. (2009). Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response. *Biometrics*, 65(2):361–368.

Huang, X. (2011). Detecting random-effects model misspecification via coarsened data. *Computational Statistics & Data Analysis*, 55(1):703–714.

Hyslop, D. R. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica*, 67(6):1255–1294.

Imhof, J. P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3-4):419–426.

Jara, A., Garcia-Zattera, M. J., and Lesaffre, E. (2007). A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics & Data Analysis*, 51(11):5402–5415.

Jaumotte, F. (2003). Female labour force particpation: past trends and main determinants in OECD countries. Technical Report 376, OECD Economics Department Working Papers, Paris.

Jenkins, A. (2006). Women, lifelong learning and transitions into employment. *Work, Employment & Society*, 20(2):309–328.

Kleinman, K. P. and Ibrahim, J. G. (1998). A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine*, 17(22):2579–2596.

Knott, M. and Tzamourani, P. (2007). Bootstrapping the estimated latent distribution of the two-parameter latent trait model. *British Journal of Mathematical and Statistical Psychology*, 60(1):175–191.

Komarek, A. and Lesaffre, E. (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Computational Statistics & Data Analysis*, 52(7):3441–3458.

Kullback, S. (1959). *Information Theory and Statistics*. John Wiley & Sons, New York.

Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811.

Lange, N. and Ryan, L. (1989). Assessing normality in random effects models. *The Annals of Statistics*, 17(2):624–642.

Lee, K. J. and Thompson, S. G. (2008). Flexible parametric models for random-effects distributions. *Statistics in Medicine*, 27(3):418–434.

Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random effects model: an example. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(3):325–335.

Lesperance, M. L. and Kalbfleisch, J. D. (1992). An algorithm for computing the nonparametric MLE of a mixing distribution. *Journal of the American Statistical Association*, 87(417):120–126.

Lesperance, M. L., Saab, R., and Neuhaus, J. (2014). Nonparametric estimation of the mixing distribution in logistic regression mixed models with random intercepts and slopes. *Computational Statistics & Data Analysis*, 71:211–219.

Li, B., Lingsma, H. F., Steyerberg, E. W., and Lesaffre, E. (2011). Logistic random effects regression models: a comparison of statistical packages for binary and ordinal outcomes. *BMC Medical Research Methodology*, 11(1):77.

Lin, K.-C. and Chen, Y.-J. (2015). Detecting misspecification in the random-effects structure of cumulative logit models. *Computational Statistics & Data Analysis*, 92:126–133.

Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory. *The Annals of Statistics*, 11(1):86–94.

Lindsey, J. K. (1997). *Applying Generalized Linear Models*. Springer, New York.

Litiére, S. (2007). *The Impact of Random-effects Misspecification on Maximum Likelihood Estimators in Generalized Linear Mixed Models*. PhD thesis, Universiteit Hasselt, Belgium.

Litiére, S., Alonso, A., and Molenberghs, G. (2007). Type I and Type II error under random-effects misspecification in generalized linear mixed models. *Biometrics*, 63(4):1038–1044.

Litiére, S., Alonso, A., and Molenberghs, G. (2008). The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 27(16):3125–3144.

Litiére, S., Alonso, A., and Molenberghs, G. (2011). Rejoinder to a note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics*, 67(2):656–660.

Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, 88(421):125–134.

Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, 90(431):1112–1121.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical analysis with missing data*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Liu, L. and Yu, Z. (2008). A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine*, 27(16):3105–3124.

Lynch, S. M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer, New York.

Maas, C. J. and Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, 46(3):427–440.

Magder, L. S. and Zeger, S. L. (1996). A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association*, 91(435):1141–1151.

Marshall, A., Altman, D. G., and Holder, R. L. (2010). Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. *BMC Medical Research Methodology*, 10(1):112.

McCulloch, C. E. and Neuhaus, J. M. (2011a). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, 26(3):388–402.

McCulloch, C. E. and Neuhaus, J. M. (2011b). Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics*, 67(1):270–279.

McCulloch, C. E. and Neuhaus, J. M. (2013). Generalized Linear Mixed Models: Estimation and Inference. In Scott, M. A., Simonoff, J. S., and Marx, B. D., editors, *The SAGE Handbook of Multilevel Modeling*, pages 271–286. SAGE Publications Ltd, London.

McCulloch, C. E., Searle, S. R., and Neuhaus, J. M. (2008). *Generalized, Linear, and Mixed Models*. John Wiley & Sons, Hoboken, 2 edition.

McGrory, C. A. and Titterington, D. M. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis*, 51(11):5352–5367.

Molenberghs, G., Fitzmaurice, G. M., Kenward, M. G., Tsiatis, A. A., and Verbeke, G. (2015). *Handbook of Missing Data Methodology*. CRC Press, Boca Raton.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer, New York.

Müller, S., Scealy, J. L., and Welsh, A. H. (2013). Model selection in linear mixed models. *Statistical Science*, 28(2):135–167.

Muthén, B. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In Kaplan, D., editor, *The SAGE Handbook of Quantitative Methodology for the Social Sciences*, pages 346–369. SAGE Publications, Inc., Thousand Oaks.

Muthén, B. and Asparouhov, T. (2008). Growth mixture modeling: Analysis with non-Gaussian random effects. In Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G., editors, *Longitudinal Data Analysis*. CRC Press, Boca Raton.

Nelder, J. A. and Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society. Series A (General)*, 135(3):370–384.

Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M., and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational & Graphical Statistics*, 15(1):39–57.

Neuhaus, J., Kalbfleisch, J., and Hauck, W. (1994). Conditions for consistent estimation in mixed-effects models for binary matched-pairs data. *Canadian Journal of Statistics*, 22(1):139–148.

Neuhaus, J. M., Hauck, W. W., and Kalbfleisch, J. D. (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika*, 79(4):755–762.

Neuhaus, J. M. and Jewell, N. P. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, 80(4):807–815.

Neuhaus, J. M. and Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, 54(2):638–645.

Neuhaus, J. M. and McCulloch, C. E. (2006). Separating between- and within-cluster covariate effects by using conditional and partitioning methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(5):859–872.

Neuhaus, J. M. and McCulloch, C. E. (2011). Estimation of covariate effects in generalized linear mixed models with informative cluster sizes. *Biometrika*, 98(1):147–162.

Neuhaus, J. M. and McCulloch, C. E. (2014). Covariate decomposition methods for longitudinal missing-at-random data and predictors associated with subject-specific effects. *Australian & New Zealand Journal of Statistics*, 56(4):331–345.

Neuhaus, J. M., McCulloch, C. E., and Boylan, R. (2011). A note on type II error under random effects misspecification in generalized linear mixed models. *Biometrics*, 67(2):654–660.

Neuhaus, J. M., McCulloch, C. E., and Boylan, R. (2013). Estimation of covariate effects in generalized linear mixed models with a misspecified distribution of random intercepts and slopes. *Statistics in Medicine*, 32(14):2419–2429.

Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York, 2 edition.

O'Keeffe, A. G., Tom, B. D., and Farewell, V. T. (2013). Mixture distributions in multi-state modelling: Some considerations in a study of psoriatic arthritis. *Statistics in Medicine*, 32(4):600–619.

Pan, Z. and Lin, D. Y. (2005). Goodness-of-fit methods for generalized linear mixed models. *Biometrics*, 61(4):1000–1009.

Parr, N. (2012). Trends in differentials in the workforce participation of mothers with young children in Australia 2002-2008. *Journal of Population Research*, 29(3):203–227.

Proust, C. and Jacqmin-Gadda, H. (2005). Estimation of linear mixed models with a mixture of distribution for the random effects. *Computer Methods and Programs in Biomedicine*, 78(2):165–173.

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

Rathouz, J., P. (2004). Fixed effects models for longitudinal binary data with drop-outs missing at random. *Statistica Sinica*, 14:969–988.

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):731–792.

Rizopoulos, D., Verbeke, G., and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika*, 95(1):63–74.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *The Annals of Statistics*, 4(6):1200–1209.

Singer, B. and Spilerman, S. (1976). Some methodological issues in the analysis of longitudinal surveys. *Annals of Economic and Social Measurement*, 5(4):447–474.

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

Skrondal, A. and Rabe-Hesketh, S. (2008). Multilevel and related models for longitudinal data. In de Leeuw, J. and Meijer, E., editors, *Handbook of Multilevel Analysis*, pages 275–299. Springer, New York.

Skrondal, A. and Rabe-Hesketh, S. (2014). Protective estimation of mixed-effects logistic regression when data are not missing at random. *Biometrika*, 101(1):175–188.

Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 171(1):5–19.

Steele, F. (2011). Multilevel discrete-time event history models with applications to the analysis of recurrent employment transitions. *Australian & New Zealand Journal of Statistics*, 53(1):1–20.

Steele, R. (2013). Model selection for multilevel models. In Scott, M. A., Simonoff, J. S., and Marx, B. D., editors, *The SAGE Handbook of Multilevel Modeling*, chapter 7, pages 109–127. SAGE Publications Ltd, London.

Tannous, K. and Smith, M. (2013). Access to full-time employment: Does gender matter? *Australian Journal of Labour Economics*, 16(2):237–257.

Tao, H., Palta, M., Yandell, B. S., and Newton, M. A. (1999). An estimation method for the semiparametric mixed effects model. *Biometrics*, 55(1):102–110.

Tchetgen, E. J. and Coull, B. A. (2006). A diagnostic test for the mixing distribution in a generalised linear mixed model. *Biometrika*, 93(4):1003–1010.

Torczon, V. (1991). On the convergence of the multidirectional search algorithm. *SIAM Journal on Optimization*, 1(1):123–145.

Tsonaka, R., Verbeke, G., and Lesaffre, E. (2009). A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, 65(1):81–87.

Vaida, F. and Blanchard, S. (2005). Conditional Akaike information for mixed-effects models. *Biometrika*, 92(2):351–370.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91(433):217–221.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, 23(4):541–556.

Verbeke, G. and Molenberghs, G. (2009). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Verbeke, G. and Molenberghs, G. (2013). The gradient function as an exploratory goodness-of-fit assessment of the random-effects distribution in mixed models. *Biostatistics*, 14(3):477–490.

Verbeke, G., Molenberghs, G., and Rizopoulos, D. (2010). Random effects models for longitudinal data. In van Montfort, K., Oud, J., and Satorra, A., editors, *Longitudinal Research with Latent Variables*, pages 37–96. Springer, Berlin.

Verbeke, G., Spiessens, B., and Lesaffre, E. (2001). Conditional linear mixed models. *The American Statistician*, 55(1):25–34.

Vieira, M. D. and Skinner, C. J. (2008). Estimating models for panel survey data under complex sampling. *Journal of Official Statistics*, 24(3):343 – 364.

Vock, D. M., Davidian, M., and Tsiatis, A. A. (2014). SNP_NLMM: A SAS macro to implement a flexible random effects density for generalized linear and nonlinear mixed models. *Journal of Statistical Software*, 56(2):1–21.

Waagepetersen, R. (2006). A simulation-based goodness-of-fit test for random effects in generalized linear mixed models. *Scandinavian Journal of Statistics*, 33(4):721–731.

Wang, Y. (2007). On fast computation of the non-parametric maximum likelihood estimate of a mixing distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):185–198.

Wang, Y. (2010a). Maximum likelihood computation for fitting semiparametric mixture models. *Statistics and Computing*, 20(1):75–86.

Wang, Y. (2010b). *Simultaneous Effects of Misspecification and Missingness on Parameter Estimation in Generalized Multilevel Models with Longitudinal Data*. PhD thesis, University of California, Los Angeles.

Watier, L., Richardson, S., and Green, P. J. (1999). Using Gaussian mixtures with unknown number of components for mixed model estimation. In Friedl, H., Berghold, A., and Kauermann, G., editors, *Proceedings of the 14th International Workshop on Statistical Modelling*, pages 394–401, Graz, Austria.

Watson, N. and Wooden, M. (2009). Identifying factors affecting longitudinal survey response. In Lynn, P., editor, *Methodology of Longitudinal Surveys*, pages 157–181. John Wiley & Sons, Ltd, Chichester.

Watson, N. and Wooden, M. P. (2012). The HILDA survey: a case study in the design and development of a successful household panel survey. *Longitudinal and Life Course Studies*, 3(3):369–381.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.

Wicklin, R. (2013). *Simulating Data with SAS*. SAS Institute, Cary.

Wooden, M. and Watson, N. (2007). The HILDA survey and its contribution to economic and social research (so far). *Economic Record*, 83(261):208–231.

Wu, C.-F. (1978). Some algorithmic aspects of the theory of optimal designs. *The Annals of Statistics*, 6(6):1286–1301.

Wynn, H. P. (1970). The sequential generation of the theory of optimal designs. *The Annals of Mathematical Statistics*, 41(5):1655–1664.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42(1):121–130.

Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3):795–802.

Appendix A: SAS syntax for likelihood reformulation method

The following SAS code uses the likelihood reformulation method (Liu and Yu, 2008) to fit a random intercept logistic model assuming the random intercepts are distributed as a three component mixture of normal distributions with unequal component variances to the case study in Chapter 4. The corresponding model estimates for the 1359 women with complete case data and the 1927 women with monotone missing data are presented in Table 4.3.

A.1 Complete case data

```
proc nlmixed data=hildadata2LR qpoints=54 cov;
 parms  beta1=0.09 beta2=-0.3 beta3=-0.12 beta4=-1.53 beta5=-2.8 beta6=-2.33 beta7=-0.4
          mu1=-4.37 mu2=0.34 mu3=4.40 p1=0.1 p2=0.42 sd1=1.1 sd2=1.49 sd3=0.79;
 bounds sd1 sd2 sd3 p1 p2>=0;
 bounds p1 p2<1;
 where Completers=1;
 /*conditional likelihood for observed data given random effects;*/
 eta = a + beta1*hgage +  beta2*maritalStat3cat1 + beta3*maritalStat3cat2 +
         beta4*BaselineEduc12 + beta5*BaselineEduc3 + beta6*DependChild1 +
         beta7*DependChild2;
expeta = exp(eta);
p=expeta/(1+expeta);

 loglik=employment_2cat*log(p) + (1-employment_2cat)*log(1-p);

 /*log finite mixture density*/
  /* To satisfy the restriction: p1+p2+p3=1*/
  p3=1-p1-p2;
logmixden=log((p1/sd1)*exp(-0.5*((a-mu1)/sd1)**2) + (p2/sd2)*exp(-0.5*((a-mu2)/sd2)**2) +
          (p3/sd3)*exp(-0.5*((a-mu3)/sd3)**2));
 /*log standard normal density*/
          lognormalden=-(a**2)/2;
 /*lastid=1 for the last observation for the same id, otherwise lastid=0;*/
 if lastid=1 then loglik=loglik+logmixden-lognormalden;
 model employment_2cat~general(loglik);
 random a~normal(0,1) subject=xwaveid;
 estimate 'prob3' 1-p1-p2;
 estimate 'beta0' p1*mu1 + p2*mu2 + (1-p1-p2)*mu3;
 estimate 'RE_Var' p1*((mu1 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd1**2)
        +  p2*((mu2 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd2**2)
         +  (1-p1-p2)*((mu3 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd3**2);
 estimate 'mean1' mu1 -  (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 estimate 'mean2' mu2 -  (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 estimate 'mean3' mu3 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 run;
```

A.2 Monotone missing data

```
proc nlmixed data=hildadata2LR qpoints=61 cov;
 parms beta1=0.09 beta2=-0.25 beta3=-0.2 beta4=-1.6 beta5=-2.9 beta6=-2.3 beta7=-0.4
                 mu1=-4.1 mu2=0.3 mu3=3.8 p1=0.14 p2=0.37 sd1=1.2 sd2=1.4 sd3=0.76;
 bounds sd1 sd2 sd3 p1 p2>0;
 bounds p1 p2<1;

 /*conditional likelihood for observed data given random effects;*/
 eta =  a + beta1*hgage +  beta2*maritalStat3cat1 + beta3*maritalStat3cat2 + beta4*BaselineEduc12
       + beta5*BaselineEduc3 + beta6*DependChild1 + beta7*DependChild2    ;
 expeta = exp(eta);
 p=expeta/(1+expeta);

 loglik=employment_2cat*log(p) + (1-employment_2cat)*log(1-p);

 /*log finite mixture density*/
  /* To satisfy the restriction: p1+p2+p3=1*/
  p3=1-p1-p2;
 logmixden=log((p1/sd1)*exp(-0.5*((a-mu1)/sd1)**2) + (p2/sd2)*exp(-0.5*((a-mu2)/sd2)**2)
       + (p3/sd3)*exp(-0.5*((a-mu3)/sd3)**2));

 /*log standard normal density*/
  lognormalden=-(a**2)/2;
 /*lastid=1 for the last observation for the same id, otherwise lastid=0;*/
 if lastid=1 then loglik=loglik+logmixden-lognormalden;
 model employment_2cat~general(loglik);
 random a~normal(0,1) subject=xwaveid;
 estimate 'prob3' 1-p1-p2;
 estimate 'beta0' p1*mu1 + p2*mu2 + (1-p1-p2)*mu3;
 estimate 'RE_Var' p1*((mu1 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd1**2)
       +  p2*((mu2 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd2**2)
       +  (1-p1-p2)*((mu3 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3))**2 + sd3**2);
 estimate 'mean1' mu1 -  (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 estimate 'mean2' mu2 -  (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 estimate 'mean3' mu3 - (p1*mu1 + p2*mu2 + (1-p1-p2)*mu3);
 run;
```

Appendix B: Sensitivity analyses of likelihood reformulation method applied to the HILDA case study in Chapter 4

The likelihood reformulation method (Liu and Yu, 2008) applied to estimate logistic models assuming mixture distributed random intercepts in the HILDA case study (Chapter 4) appeared to be sensitive to the number of adaptive quadrature points. To explore the impact of the choice of the adaptive quadrature points used in the estimation, a sensitivity analysis was performed by re-fitting the likelihood reformulation method with adaptive quadrature points ranging from 10 to 80. The sensitivity analyses were performed for both data scenarios: the women with complete case data, and the women with monotone missing data.

B.1 Complete case data

For the random intercept logistic models that converged, the parameter coefficients and standard errors for the women with complete case data estimated using the likelihood reformulation method for quadrature points ranging from 24 to 80 are presented in Table B.1. The residual deviance ($-2ll$) was similar for all converged models ranging from 9685.7 to 9690.6. The deviance increased slightly as the number of adaptive quadrature points increased. The model parameters, standard errors and the deviance appeared to stabilise after 54 quadrature points, with the deviance for the models with 54 to 80 quadrature points ranging between 9690.4 to 9690.6. Therefore, the results in Table 4.3 for the Complete Data scenario are based on 54 adaptive quadrature points.

Table B.1: Fitting the three component mixture of normal distributions as the random intercept for the 1359 women with complete case data. The same starting values for all models were used, however the number of quadrature points (Points) varied and ranged from 10 to 80. The following results are the estimates (Est) and corresponding standard errors (SE) of converged models for the fixed effect coefficients ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$) and the estimates of the random intercept three component mixture distribution (mixing proportions: π_1, π_2, π_3, mean components: μ_1, μ_2, μ_3), standard deviations: $\sigma_1, \sigma_2, \sigma_3$, as well as the estimated intercept ($\hat{\beta}_0 = \hat{\pi}_1\hat{\mu}_1 + \hat{\pi}_2\hat{\mu}_2 + \hat{\pi}_3\hat{\mu}_3$) and the total random effect variance (σ_b^2)

Points		Fixed Effects								Variance Component										$-2ll$
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	π_1	π_2	π_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	σ_b^2	
24	Est	0.632	0.093	-0.337	0.049	-1.439	-2.567	-2.268	-0.369	0.143	0.593	0.265	-5.159	-0.223	3.287	2.309	1.749	0.649	9.377	9680.5
	SE	0.483	0.009	0.130	0.275	0.290	0.299	0.149	0.121	0.082	0.168	0.104	1.618	0.316	0.552	1.037	0.400	0.225	0.924	
28	Est	0.441	0.097	-0.359	0.061	-1.391	-2.554	-2.262	-0.375	0.156	0.542	0.302	-4.963	-0.336	3.160	2.338	1.592	0.662	9.277	9680.2
	SE	0.457	0.009	0.138	0.274	0.270	0.291	0.150	0.122	0.074	0.132	0.079	1.378	0.333	0.351	0.695	0.316	0.646	0.953	
32	Est	0.454	0.096	-0.339	0.065	-1.388	-2.500	-2.269	-0.376	0.151	0.561	0.288	-5.000	-0.277	3.170	2.452	1.675	0.629	9.310	9680.3
	SE	0.458	0.009	0.156	0.276	0.252	0.269	0.149	0.121	0.059	0.134	0.086	1.105	0.316	0.353	0.820	0.332	0.319	0.919	
34	Est	0.373	0.098	-0.363	0.059	-1.374	-2.549	-2.248	-0.368	0.172	0.516	0.312	-4.674	-0.325	3.110	2.400	1.526	0.660	9.161	9680.6
	SE	0.456	0.009	0.139	0.274	0.268	0.294	0.149	0.121	0.150	0.228	0.100	2.560	0.372	0.354	1.298	0.516	0.747	1.004	
42	Est	0.374	0.098	-0.362	0.056	-1.372	-2.547	-2.249	-0.369	0.171	0.519	0.310	-4.683	-0.317	3.118	2.411	1.539	0.657	9.176	9680.6
	SE	0.456	0.009	0.139	0.274	0.268	0.293	0.149	0.122	0.147	0.219	0.096	2.544	0.376	0.352	1.305	0.485	0.743	1.004	
48	Est	1.017	0.091	-0.480	-0.065	-1.637	-2.761	-2.365	-0.401	0.126	0.567	0.307	-5.465	-0.540	3.238	1.450	1.589	0.773	9.023	9685.8
	SE	0.459	0.009	0.141	0.268	0.263	0.291	0.151	0.121	0.057	0.168	0.121	0.913	0.378	0.544	0.518	0.603	0.672	1.123	
50	Est	0.992	0.092	-0.494	-0.062	-1.630	-2.771	-2.349	-0.410	0.128	0.570	0.303	-5.446	-0.506	3.256	1.466	1.594	0.771	9.050	9685.7
	SE	0.458	0.009	0.141	0.268	0.263	0.295	0.151	0.121	0.066	0.193	0.135	1.049	0.391	0.587	0.566	0.689	0.585	1.093	
52	Est	0.990	0.092	-0.497	-0.062	-1.623	-2.778	-2.337	-0.417	0.128	0.570	0.302	-5.445	-0.502	3.258	1.467	1.593	0.770	9.041	9685.7
	SE	0.459	0.009	0.141	0.268	0.263	0.297	0.150	0.121	0.067	0.196	0.137	1.071	0.399	0.596	0.564	0.700	0.680	1.148	
54	Est	1.068	0.090	-0.310	-0.116	-1.521	-2.824	-2.330	-0.396	0.124	0.553	0.323	-5.437	-0.700	3.291	1.121	1.502	0.839	9.074	9690.6
	SE	0.464	0.009	0.140	0.261	0.247	0.267	0.150	0.119	0.044	0.145	0.109	0.662	0.342	0.639	0.417	0.515	0.664	1.287	
56	Est	1.073	0.090	-0.310	-0.116	-1.524	-2.821	-2.331	-0.395	0.124	0.552	0.324	-5.442	-0.705	3.286	1.122	1.502	0.839	9.071	9690.6
	SE	0.464	0.009	0.140	0.262	0.247	0.267	0.150	0.119	0.044	0.144	0.109	0.657	0.343	0.637	0.414	0.511	0.668	1.289	
58	Est	1.066	0.090	-0.310	-0.116	-1.527	-2.819	-2.333	-0.396	0.124	0.553	0.323	-5.435	-0.697	3.292	1.122	1.502	0.841	9.072	9690.6
	SE	0.464	0.009	0.140	0.261	0.247	0.267	0.150	0.119	0.045	0.146	0.110	0.668	0.343	0.645	0.421	0.520	0.661	1.291	
60	Est	1.066	0.090	-0.310	-0.116	-1.528	-2.817	-2.334	-0.396	0.125	0.553	0.323	-5.435	-0.697	3.292	1.122	1.502	0.841	9.074	9690.6
	SE	0.465	0.009	0.140	0.261	0.247	0.267	0.150	0.119	0.045	0.147	0.111	0.673	0.344	0.649	0.424	0.523	0.662	1.296	
62	Est	1.070	0.090	-0.310	-0.116	-1.530	-2.815	-2.336	-0.396	0.124	0.553	0.323	-5.439	-0.700	3.288	1.121	1.502	0.840	9.073	9690.6
	SE	0.465	0.009	0.140	0.262	0.247	0.267	0.150	0.120	0.045	0.146	0.110	0.669	0.344	0.644	0.422	0.520	0.664	1.293	
64	Est	1.068	0.090	-0.310	-0.116	-1.532	-2.813	-2.337	-0.396	0.124	0.553	0.323	-5.437	-0.698	3.290	1.121	1.502	0.842	9.067	9690.6
	SE	0.464	0.009	0.140	0.262	0.247	0.267	0.150	0.119	0.045	0.147	0.111	0.669	0.344	0.647	0.422	0.521	0.660	1.293	
66	Est	1.069	0.090	-0.312	-0.115	-1.534	-2.813	-2.339	-0.396	0.125	0.551	0.324	-5.439	-0.696	3.285	1.125	1.506	0.837	9.097	9690.4
	SE	0.465	0.009	0.140	0.262	0.248	0.268	0.150	0.120	0.047	0.149	0.111	0.697	0.346	0.645	0.439	0.533	0.676	1.299	
68	Est	1.069	0.090	-0.310	-0.116	-1.535	-2.810	-2.339	-0.397	0.124	0.552	0.323	-5.439	-0.699	3.288	1.121	1.503	0.841	9.075	9690.5
	SE	0.465	0.009	0.140	0.262	0.247	0.267	0.150	0.120	0.045	0.147	0.111	0.676	0.345	0.649	0.426	0.524	0.664	1.298	
70	Est	1.073	0.090	-0.310	-0.116	-1.535	-2.810	-2.340	-0.397	0.124	0.552	0.324	-5.442	-0.702	3.285	1.121	1.503	0.841	9.071	9690.5
	SE	0.465	0.009	0.140	0.262	0.248	0.267	0.150	0.120	0.045	0.146	0.110	0.670	0.345	0.645	0.423	0.520	0.666	1.296	
72	Est	1.070	0.090	-0.310	-0.116	-1.536	-2.809	-2.340	-0.397	0.124	0.553	0.323	-5.439	-0.699	3.287	1.121	1.503	0.842	9.073	9690.5

Continued on next page

Points		Fixed Effects								Variance Component										$-2ll$
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	π_1	π_2	π_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	σ_b^2	
	SE	0.465	0.009	0.140	0.262	0.248	0.267	0.150	0.120	0.045	0.147	0.111	0.676	0.345	0.650	0.426	0.524	0.663	1.299	
74	Est	1.067	0.090	-0.311	-0.116	-1.538	-2.808	-2.339	-0.391	0.124	0.554	0.322	-5.437	-0.698	3.290	1.122	1.503	0.843	9.058	9690.5
	SE	0.464	0.009	0.140	0.262	0.247	0.267	0.150	0.119	0.045	0.147	0.111	0.668	0.345	0.654	0.422	0.522	0.657	1.299	
76	Est	1.070	0.090	-0.310	-0.115	-1.537	-2.808	-2.341	-0.397	0.124	0.552	0.323	-5.439	-0.700	3.287	1.121	1.503	0.842	9.074	9690.5
	SE	0.465	0.009	0.140	0.262	0.248	0.267	0.150	0.120	0.045	0.148	0.111	0.678	0.345	0.651	0.427	0.525	0.664	1.301	
78	Est	1.071	0.090	-0.310	-0.116	-1.537	-2.808	-2.341	-0.397	0.124	0.552	0.323	-5.440	-0.701	3.286	1.121	1.503	0.842	9.072	9690.5
	SE	0.465	0.009	0.140	0.262	0.248	0.267	0.150	0.120	0.045	0.147	0.111	0.674	0.345	0.649	0.425	0.523	0.664	1.300	
80	Est	1.071	0.090	-0.310	-0.116	-1.537	-2.808	-2.341	-0.398	0.124	0.552	0.323	-5.441	-0.701	3.286	1.121	1.503	0.842	9.072	9690.5
	SE	0.465	0.009	0.140	0.262	0.248	0.267	0.150	0.120	0.045	0.147	0.111	0.675	0.345	0.649	0.425	0.523	0.664	1.299	

B.2 Monotone missing data

For the random intercept logistic models that converged, the parameter coefficients and standard errors for the women with monotone missing data estimated using the likelihood reformulation for quadrature points ranging from 15 to 80 are presented in Table B.2. The residual deviance ($-2ll$) for the converged models were similar, ranging between 11526 to 11534. The parameter estimates for the fixed effects and variance components were relatively stable for all quadrature points. However, there appeared to be variability in the standard errors for parameters in the variance component. For instance, the standard error for estimate for the mixing proportion of the first component (π_1) ranged from 0.020 (21 points) to 1.789 (79 points), and the standard error for the mean of the first component (μ_1) ranged from 0.235 (21 points) to 24.899 (79 points). Furthermore, some models had consistently large standard errors for all variance component parameters, such as, models with 75 and 79 quadrature points. The observed variability in the results may indicate that the model estimates are local maxima of the likelihood. This may be an attribute of the starting values selected for the likelihood reformulation method, as the estimation of the variance components appeared sensitive to the choice of starting values for the variance component. The observed variability could perhaps also be a consequence of the missing data. The parameter estimates and standard errors for the models estimated with 57 and 61 quadrature points were similar. The results in Table 4.3 for the monotone missing data are based on 61 quadrature points, as it produced consistently small standard errors.

Table B.2: Fitting the three component mixture of normal distributions as the random intercept for the 1927 women with monotone missing data. The same starting values for all models were used, however the number of quadrature points (Points) varied and ranged from 10 to 80. The following results are the estimates (Est) and corresponding standard errors (SE) of converged models for the fixed effect coefficients ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7$) and the estimates of the random intercept three component mixture distribution (mixing proportions: π_1, π_2, π_3, mean components: μ_1, μ_2, μ_3, standard deviations: $\sigma_1, \sigma_2, \sigma_3$) as well as the estimated intercept ($\hat{\beta}_0 = \hat{\pi}_1\hat{\mu}_1 + \hat{\pi}_2\hat{\mu}_2 + \hat{\pi}_3\hat{\mu}_3$) and the total random effect variance (σ_b^2)

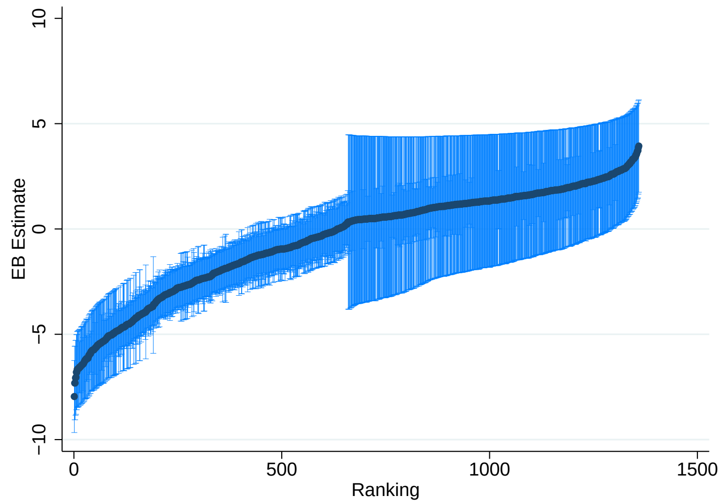
Points		**Fixed Effects**								**Variance Component**										$-2ll$
		β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	π_1	π_2	π_3	μ_1	μ_2	μ_3	σ_1	σ_2	σ_3	σ_b^2	
15	Est	0.876	0.091	-0.253	-0.200	-1.610	-2.895	-2.299	-0.400	0.149	0.499	0.353	-4.974	-0.589	2.933	1.208	1.369	0.735	8.227	11534
	SE	0.461	0.009	0.151	0.234	0.222	0.238	0.150	0.137	0.040	0.210	0.174	0.333	0.681	0.471	0.392	0.877	0.508	0.620	
21	Est	0.952	0.089	-0.255	-0.199	-1.605	-2.901	-2.299	-0.408	0.142	0.492	0.366	-5.051	-0.643	2.833	1.214	1.392	0.814	8.175	11534
	SE	0.389	0.008	0.133	0.242	0.223	0.256	0.130	0.104	0.020	0.061	0.049	0.235	0.225	0.175	0.352	0.239	0.212	0.604	
22	Est	0.896	0.090	-0.253	-0.200	-1.604	-2.899	-2.301	-0.398	0.148	0.495	0.357	-4.994	-0.610	2.915	1.208	1.369	0.732	8.236	11534
	SE	0.422	0.009	0.134	0.244	0.220	0.237	0.140	0.113	0.075	0.129	0.062	0.937	0.210	0.212	0.788	0.528	0.201	0.629	
30	Est	0.909	0.090	-0.254	-0.200	-1.605	-2.898	-2.301	-0.399	0.147	0.494	0.359	-5.007	-0.621	2.899	1.208	1.371	0.738	8.223	11534
	SE	0.422	0.009	0.136	0.236	0.220	0.232	0.142	0.113	0.052	0.196	0.160	0.598	0.634	0.474	0.559	0.554	1.526	1.305	
33	Est	0.682	0.095	-0.317	-0.085	-1.560	-2.863	-2.310	-0.416	0.168	0.517	0.315	-4.932	-0.375	3.238	1.935	1.448	0.667	9.305	11527
	SE	0.425	0.009	0.137	0.242	0.241	0.267	0.141	0.116	0.124	0.268	0.151	1.758	0.353	0.525	0.618	0.935	0.568	0.916	
36	Est	0.908	0.090	-0.253	-0.200	-1.605	-2.898	-2.301	-0.399	0.147	0.494	0.359	-5.005	-0.621	2.904	1.208	1.368	0.731	8.228	11534
	SE	0.426	0.009	0.140	0.247	0.225	0.231	0.145	0.113	0.159	0.498	0.344	1.758	0.969	0.921	1.240	1.676	2.404	1.825	
39	Est	0.858	0.092	-0.390	-0.160	-1.655	-2.799	-2.324	-0.383	0.154	0.497	0.349	-5.086	-0.556	3.039	1.551	1.388	0.923	8.981	11529
	SE	0.430	0.009	0.132	0.241	0.230	0.245	0.141	0.114	0.078	0.225	0.161	1.039	0.545	0.539	0.554	0.734	1.082	1.206	
46	Est	0.885	0.090	-0.254	-0.199	-1.619	-2.886	-2.300	-0.410	0.147	0.501	0.352	-4.984	-0.569	2.892	1.212	1.409	0.834	8.219	11534
	SE	0.211	0.001	0.131	0.234	0.219	0.222	0.135	0.112	0.067	0.112	0.050	0.806	0.198	0.105	0.614	0.471	0.229	0.619	
51	Est	0.616	0.095	-0.311	-0.068	-1.564	-2.810	-2.277	-0.396	0.171	0.530	0.299	-4.805	-0.288	3.257	2.088	1.530	0.623	9.265	11526
	SE	0.427	0.009	0.167	0.252	0.287	0.384	0.140	0.117	0.388	0.791	0.407	5.435	0.615	1.141	1.618	2.613	0.559	1.140	
55	Est	0.632	0.096	-0.341	-0.066	-1.569	-2.834	-2.298	-0.409	0.173	0.518	0.310	-4.824	-0.345	3.267	2.047	1.466	0.630	9.348	11526
	SE	0.424	0.009	0.139	0.243	0.243	0.271	0.140	0.115	0.159	0.320	0.168	2.248	0.337	0.610	0.827	1.075	0.372	0.927	
57	Est	0.617	0.095	-0.312	-0.065	-1.563	-2.809	-2.277	-0.396	0.171	0.530	0.299	-4.804	-0.291	3.256	2.109	1.533	0.618	9.276	11526
	SE	0.425	0.009	0.138	0.245	0.253	0.267	0.140	0.116	0.108	0.261	0.160	1.501	0.428	0.529	0.698	0.894	0.477	0.891	
59	Est	0.616	0.095	-0.310	-0.068	-1.563	-2.810	-2.277	-0.397	0.171	0.530	0.299	-4.801	-0.289	3.256	2.093	1.529	0.615	9.259	11526
	SE	0.425	0.009	0.147	0.245	0.256	0.305	0.140	0.116	0.239	0.482	0.249	3.370	0.433	0.735	1.106	1.586	0.448	0.957	
61	Est	0.611	0.095	-0.311	-0.067	-1.563	-2.809	-2.275	-0.395	0.172	0.530	0.299	-4.792	-0.284	3.258	2.100	1.531	0.618	9.267	11526
	SE	0.425	0.009	0.137	0.245	0.254	0.264	0.141	0.117	0.096	0.251	0.161	1.314	0.456	0.536	0.673	0.867	0.535	0.900	
64	Est	0.641	0.095	-0.312	-0.076	-1.544	-2.846	-2.289	-0.405	0.172	0.520	0.308	-4.840	-0.318	3.239	2.002	1.467	0.650	9.258	11526
	SE	0.427	0.009	0.146	0.245	0.258	0.313	0.140	0.116	0.221	0.474	0.259	3.038	0.503	0.769	0.878	1.623	0.646	0.964	
69	Est	0.604	0.094	-0.317	-0.156	-1.466	-2.704	-2.303	-0.419	0.186	0.509	0.304	-4.559	-0.255	3.220	2.219	1.478	0.656	9.224	11526
	SE	0.429	0.009	0.138	0.244	0.260	0.276	0.141	0.115	0.121	0.298	0.184	1.587	0.529	0.582	0.715	0.984	0.615	0.931	
70	Est	0.894	0.091	-0.399	-0.177	-1.662	-2.808	-2.277	-0.420	0.152	0.493	0.355	-5.099	-0.579	2.982	1.456	1.352	0.885	8.766	11530
	SE	0.426	0.009	0.131	0.238	0.226	0.237	0.139	0.113	0.036	0.133	0.112	0.492	0.465	0.401	0.384	0.383	0.943	1.085	
75	Est	0.588	0.096	-0.320	-0.066	-1.562	-2.814	-2.277	-0.398	0.175	0.527	0.299	-4.743	-0.280	3.261	2.125	1.526	0.599	9.265	11526
	SE	0.430	0.010	0.207	0.272	0.328	0.496	0.141	0.120	0.691	1.362	0.673	9.595	0.878	1.822	3.061	4.361	1.054	1.232	
79	Est	0.590	0.096	-0.318	-0.068	-1.558	-2.803	-2.277	-0.397	0.174	0.529	0.297	-4.747	-0.268	3.260	2.139	1.538	0.619	9.275	11526
	SE	0.479	0.016	0.458	0.403	0.733	1.285	0.143	0.148	1.789	3.585	1.796	24.899	2.387	4.723	7.255	11.542	1.792	2.935	
80	Est	0.678	0.095	-0.323	-0.087	-1.558	-2.864	-2.314	-0.418	0.169	0.515	0.316	-4.922	-0.381	3.247	1.919	1.437	0.677	9.326	11527
	SE	0.426	0.009	0.136	0.242	0.238	0.260	0.141	0.115	0.109	0.235	0.134	1.529	0.343	0.497	0.590	0.815	0.585	0.926	

Appendix C: Caterpillar plots of Empirical Bayes estimates for the HILDA case study considered in Chapter 4

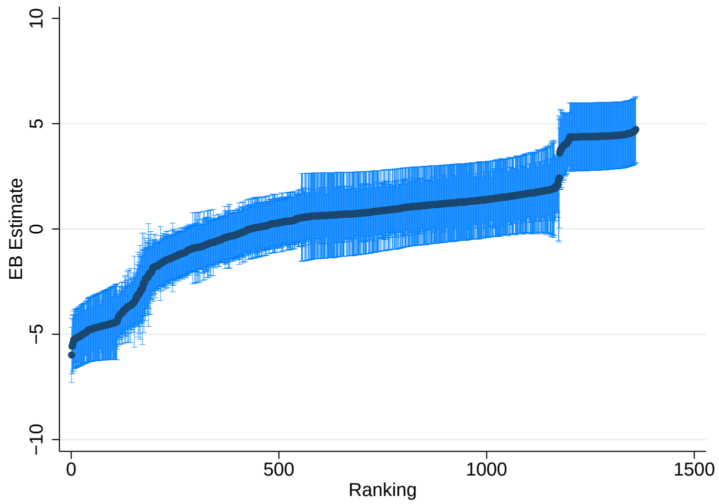
The predicted random intercept (Empirical Bayes estimates) estimated from the logistic models fit in the case study (Chapter 4) are presented graphically in Figure C.1. This plot, sometimes called a 'caterpillar plot', displays the Empirical Bayes estimate of the random intercept in rank order along with an error bar. The error bar represents the 95% confidence interval around each of the Empirical Bayes estimates, and is calculated by multiplying the standard error by a factor of 1.39 instead of the conventional 1.96 multiplication factor (Hox, 2010). The value 1.39 ($1.96/\sqrt{2}$) results in confidence intervals that can be used for pairwise comparisons. For example, if the error bars of two individuals do no overlap, the two individuals are interpreted as having significantly different random intercepts at the 5% significance level (Goldstein and Healy, 1995). Furthermore, caterpillar plots can be used to identify extreme residuals of the intercept by assessing for significant deviations from the average value of zero.

Figure C.1 (a) and (b) presents the caterpillar plots for the predicted random intercepts for the complete case data scenario when assuming a normal and a three component mixture of normals, respectively. Figure C.1 (c) and (d) presents the corresponding caterpillar plots for the monotone missing data scenario. The caterpillar plots in Figure C.1 show that the standard errors of the predicted random intercepts were generally larger when assuming a normal random intercept (Figure C.1(a) and (c)) than assuming a three component mixture of normal distributions (Figure C.1(b) and (d)).

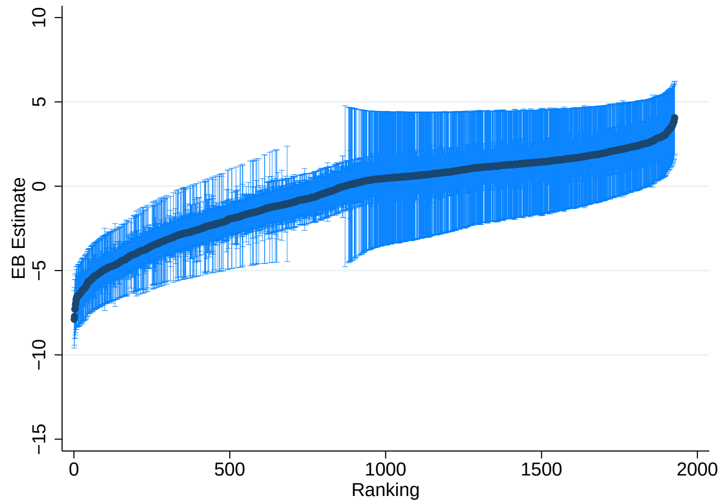
For the 1359 women with complete case data, the estimated standard errors of the predicted random intercepts when assuming normal random intercepts were always larger than the assumed mixture distribution for the observed stayers, i.e. the 103 women never employed over the 11 waves and the 631 women observed to always employed over the 11 waves. For the 1926 women with monotone missing data, the standard error of the predicted random intercepts when assuming normal random intercepts were always larger than the assumed mixture distribution for the 934 women observed to always be employed, and sometimes larger (34.6%) for the 243 women who were observed to never be employed. Some predicted random intercepts had large standard errors, particularly for assuming a mixture distribution for women with monotone missing data (Figure C.1 (d)), which may partly be explained by the instability of the likelihood reformulation method used to estimate the model with mixture distribution for the random intercepts (see Table B.2 in Appendix B for further details).



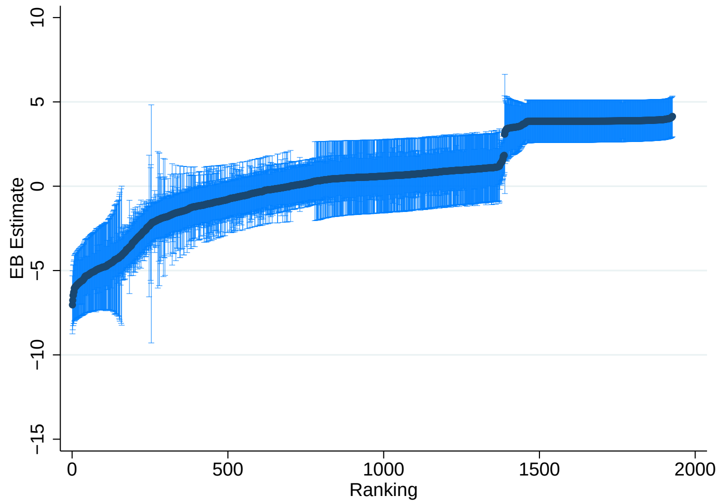
(a) Complete Cases - Assume Normal



(b) Complete Cases - Assume Mixture



(c) Monotone Missing - Assume Normal



(d) Monotone Missing - Assume Mixture

Figure C.1: Caterpillar plots of the Empirical Bayes (EB) estimates (dark blue dots) and corresponding and 95% confidence interval bands (blue error bars) for the fitted logistic model in Chapter 4 to (a) Complete cases assuming a normal distribution for the random intercept, (b) Complete cases assuming three component mixture of normal distribution for the random intercept, (c) Monotone missing data assuming a normal distribution for the random intercept, and (d) Monotone missing assuming a three component mixture of normal distribution for the random intercept.

Appendix D: Derivation of the drop-out model used to generate missing at random attrition in simulation studies presented in Chapters 5 and 6

An ordinary logistic regression was used to investigate the associations of observed variables with the conditional probability of drop-out ($p_{ij}(\alpha)$) in the HILDA case-study. For the 1927 women with monotone missingness, the data was set up such that an indicator variable, d_{ij}, represents the first time subject i drops out of the study. Therefore, $d_{ij} = 1$ when subject i drops out of the study between time $j-1$ and j (for $j = 2, ..., 11$). As the focus is on attrition, all previous time-points prior to drop-out will be observed, such that $d_{i1} = ... = d_{i,j-1} = 0$. It is assumed that all subjects are observed at the first time-point, i.e. $d_{i1} = 0$ for all $i = 1, ..., N$. If the subject was observed for all time-points, then $d_{ij} = 0$ for all $j = 1, ..., 11$. An ordinary logistic regression was fit to the dataset to model $\Pr(d_{ij} = 1)$. The variables included in the model were based on the same variables considered in the case study. The final drop-out model was determined by using forward and backward selection. The levels of categorical variables included in the drop-out model were selected based on the cell frequencies and chi-squared test comparing the model likelihood ratios. Four categories for the highest level of education were used as there were no differences in the model fit for having two, three or four categories. The final drop-out model captured wave-specific effects by including linear and quadratic wave terms (w_{1j} and w_{1j}^2), and included the following explanatory variables: employment status at the previous wave $y_{i,j-1}$, age at first wave w_{2i}, highest level of education attained at first wave (w_{3i}, w_{4i}, w_{5i}), and dependent children at the previous wave ($w_{6i,j-1}$ and $w_{7i,j-1}$). The final drop-out model was given by the following logistic model:

$$\text{logit}(\Pr(d_{ij} = 1 | d_{i,j-1} = 0)) = \alpha_0 + \alpha_1 w_{1j} + \alpha_2 w_{1j}^2 + \alpha_3 w_{2i} + \alpha_4 w_{3i} + \alpha_5 w_{4i} + \alpha_6 w_{5i} + \alpha_7 w_{6i,j-1} + \alpha_8 w_{7i,j-1} + \alpha_9 y_{i,j-1}.$$

The estimated coefficients (Estimate) and corresponding standard errors (SE) of fitting the ordinary logistic regression to the 1927 women with monotone missing data in HILDA case study is shown in Table D.1.

These coefficient estimates (Table D.1) motivated the choice of the coefficients used to simulate the MAR attrition in the simulation studies, by generating the following conditional probabilities:

$$p_{ij} = \frac{1}{1 + \exp(-\psi_{ij})}$$

where

$$\psi_{ij} = 0.28 - 0.69 w_{1j} + 0.04 w_{1j}^2 - 0.035 w_{2i} + 0.33 w_{3i} + 0.45 w_{4i} + 0.55 w_{5i}$$
$$-0.59 w_{6i,j-1} - 0.15 w_{7i,j-1} - 0.33 y_{i,j-1}$$

Data set-up and analysis of drop-out was performed in STATA Version 13 .

Table D.1: Coefficient estimates and corresponding standard errors (SE) of the ordinary logistic regression to model drop-out in the HILDA case study.

Parameter	**Coefficient**	**Estimate**	**SE**
Intercept	α_0	0.282	0.472
Wave	α_1	-0.687	0.075
Wave^2	α_2	0.040	0.006
Age at wave 1	α_3	-0.035	0.011
Highest Education at wave 1			
Bachelor or higher		0	
Diploma/Certificate	α_4	0.329	0.135
Year 12	α_5	0.451	0.145
Year 11 or less	α_6	0.551	0.125
Dependent Children at previous wave $(j-1)$			
None		0	
Youngest aged < 5	α_7	-0.585	0.137
Youngest aged 5-24	α_8	-0.154	0.100
Previous Employment Status (Y_{j-1})	α_9	-0.332	0.096

Appendix E: Simulation study assessing impact of misspecifying the assumed random intercept distribution within a clinical trial setting

The aim of this secondary simulation study is to assess the impact of misspecified multimodal random intercept distributions with and without attrition in a clinical trial setting. Considering the same multimodal distributions as in Chapter 6, the random intercepts in this secondary simulation were generated as symmetric three component mixture distributions. Using the same simulation set-up detailed in Litiére et al. (2008), data were generated to represent a binary response and a treatment effect in a randomised clinical trial.

The simulation considered by Litiére et al. (2008) is based on a case study comprising of patient data from a randomised clinical trial, comparing two treatments for chronic schizophrenia (Alonso et al., 2004). The randomised clinical trial compared the effect of risperidone with conventional antipsychotic agents. The primary aim of the statistical analysis was to assess the evolution of improvement in a patient's global mental condition over the course of treatment. Clinical Global Impression (CGI) is generally used to measure a subject's mental condition, and is based on a seven-grade scale. The primary outcome variable as considered by Litiére et al. (2008) is a dichotomous version of CGI, classifying patient i at time j as normal to mildly ill ($y_{ij} = 1$ if CGI $\in [1, 3]$) or moderately to severely ill ($y_{ij} = 0$ if CGI $\in [4, 7]$). The clinical trial comprised of 128 patients, whereby 64 were randomly assigned to receive risperidone treatment ($z_i = 1$) and the remaining 64 to an active control ($z_i = 0$), for a total of 8 weeks. The outcome was assessed at six fixed time-points, at 0, 1, 2, 4, 6 and 8 weeks. In the clinical trial, by the end of the treatment, 31 (48.4%) of the subjects in the control group and 37 (57.8%) of the subjects in the risperidone group had dropped out of the study. As in Litiére et al. (2008), the missing data generating mechanism is assumed to be missing at random (MAR).

E.1 Simulation study design

Binary response data representing the severity of schizophrenia were generated from the following model as considered in Litiére et al. (2008):

$$\text{logit}(\Pr(y_{ij} = 1|b_i)) = \beta_0 + \beta_1 z_i + \beta_2 t_j + b_i. \tag{E.1}$$

This model includes an intercept, a time-invariant binary covariate z_i representing treatment for subject i (randomly assigned with equal probability), and a continuous time covariate t_j with values 0, 1, 2, 4, 6 and 8. The coefficients were the same as in the simulation study considered by Litiére et al. (2008), and were based on the results of fitting the same random intercept logistic model to the real data (Table I in Litiére et al. 2008): $\beta_0^0 = -8$, $\beta_1^0 = 2$ and $\beta_2^0 = 1$.

The same random intercept distributions considered in Chapter 6 were included in the simulation study. Therefore, the random intercept (b_i) in Equation E.1 was generated from a symmetric three component mixture of normals. Twenty-one different random effect distributions of increasing component mean distances, each with component variances $\sigma^2 = 1, 2$ and 4, were considered. The different mean combinations for μ_1, μ_2 and μ_3 have a symmetric distribution with mean zero. The specific case was considered where $\mu_1 = -\mu_3$ and $\mu_2 = 0$, with μ_3 ranging from 0 to 10, increasing in increments of 0.5.

Simulations were performed for each of the 63 combinations of component mean distance and component variance, to generate complete data for 1000 subjects. To investigate the additional impact of missingness, two missingness scenarios were considered: complete data and incomplete data due to attrition. Attrition was assumed to be generated by the missing at random (MAR) mechanism, with the overall attrition of 30% by the end of the observation period. The MAR missingness was simulated using similar methodology as described in Section 3.4.2. The probability of drop-out for subject i at time j was generated using the following logistic model,

$$\text{logit}(\Pr(d_{ij} = 1|d_{i(j-1)} = 0)) = 3.5 + 0.95 \times y_{i,j-1} - 0.6 \times v_j \tag{E.2}$$

where $y_{i,j-1}$ is the outcome at the previous time-point, and $v_j = 1, 2, 3, 4, 5$ is the visit number, corresponding to week $t_j = 1, 2, 4, 6, 8$. The values of the coefficients were based on the observed coefficients of fitting the above model to the real data analysed by Litiére et al. (2008). Using the observed fitted coefficients as starting values, the final coefficients were obtained by adjusted each coefficient iteratively to acquire an overall attrition rate of 30%.

For each of the 21 random intercept distributions, 3 variance component settings and two missingness scenarios (126 combinations in total), 1000 datasets each containing 1000 subjects

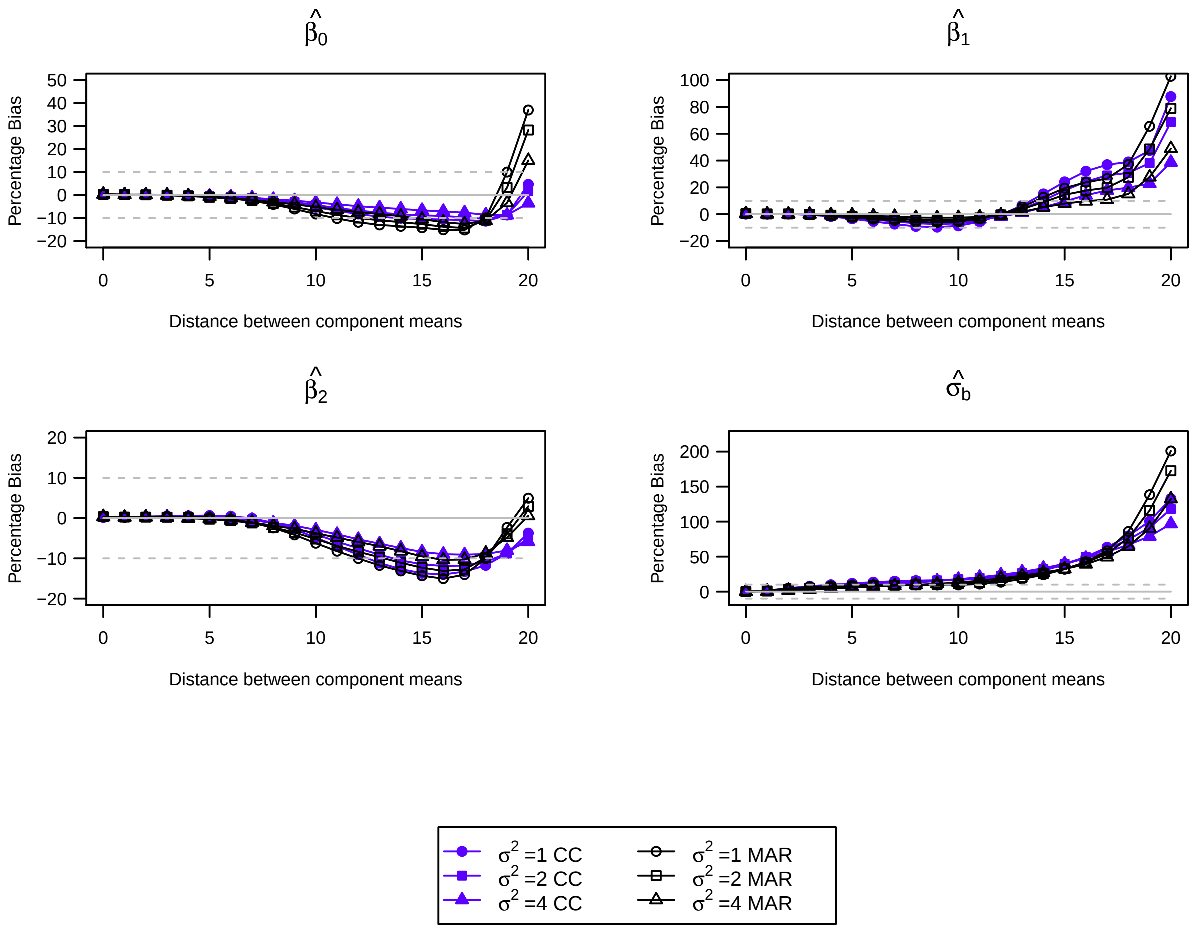


Figure E.1: Percentage Bias for parameter coefficients of random intercept logistic model for complete data (C) and MAR attrition of 30% (MAR)for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under three component variance scenarios ($\sigma^2 = 1, 2$ or 4) in the clinical trial simulation study. Grey horizontal solid line at 0 percentage bias and grey horizontal dashed lines at ± 10%.

were generated. A random intercept logistic model (Model in Equation E.1) assuming Gaussian distributed random effect was fitted to each simulated dataset. The same performance measures as detailed in Chapter 3 and used in Chapters 5 and 6 were used (i.e. percentage bias, coverage of 95% confidence intervals and standard error ratio). As bias based on the average value is sensitive to extreme values, calculation of the percentage bias is based on the median value of the 1000 simulations.

Simulations and analyses were conducted in SAS (Version 9.4, SAS Institute, Cary NC). All random intercept logistic models were fitted using the SAS procedure NLMIXED with adaptive Gaussian Quadrature using 20 quadrature points.

E.2 Results and discussion

Figure E.1 shows the percentage bias as a function of increasing component mean distances for the three variance component scenarios and both data scenarios. With increasing component mean distances, Figure E.1 shows that misspecification produced larger biased estimators for the random intercept standard deviation (σ_b). For component mean distances exceeding 10, the percentage bias was outside the acceptable limit of ±10%. Incorrectly assuming normally distributed random intercepts when the true random intercept was a three component mixture of normals generally produced unbiased estimates for fixed effect parameters (β_0, β_1 and β_2) for component mean distances less than 12. As the component mean distances exceeded 12, bias in the estimates of the slope effect β_1 increased, exceeding the acceptable threshold of 10% for distances larger than 15 for all scenarios. Non-linear trends in the percentage bias of β_0 and β_2 were observed for distances exceeding 12, coinciding with the large exponential increase in the empirical standard error (results not shown). Patterns in percentage bias were similar for the three variance components and for the two missing data scenarios.

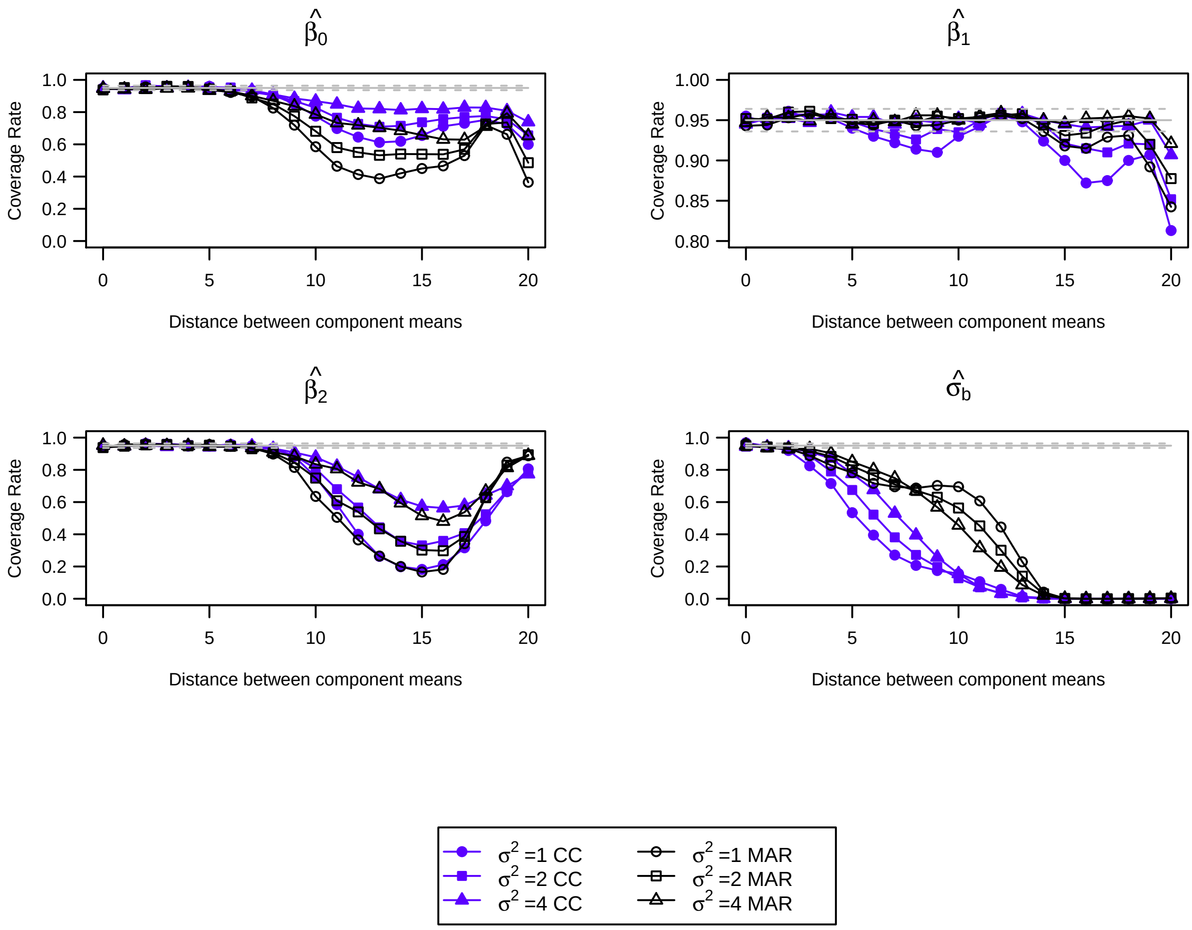


Figure E.2: Coverage of 95% confidence intervals for parameter coefficients of random intercept logistic model for complete data (C) and MAR attrition of 30% (MAR) for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under three component variance scenarios ($\sigma^2 = 1, 2$ or 4) in the clinical trial simulation study. Grey horizontal solid line at nominal coverage rate 0.95 and grey horizontal dashed lines at coverage rate 0.936 and 0.964.

Figure E.2 presents the coverage rate as a function of increasing component mean distances for the three variance component scenarios and both data scenarios. As the component mean distance increased, misspecifying the random intercept distribution produced poor coverage rates for all parameters. There were poor coverage rates for estimates of the varaince component (σ_b) for all scenarios as the component mean distance exceeded 3. Misspecification resulted in close to nominal coverage rates for β_1, with poor coverage rates observed for all scenarios for the most extreme component mean distance considered. The non-linear trend in the coverage rate for β_0 and β_2 can partially be explained by the exponential increase in the estimated standard errors for the larger component mean distances (results not shown). This suggests that, as the distance between the mean components increases, the width of the confidence interval also increases, consequently leading to apparent improvements in the coverage rates. Coverage rates had similar patterns for the three component variance scenarios. Similar coverage rates and trends as for the complete data scenario were observed for the 30% MAR attrition scenario.

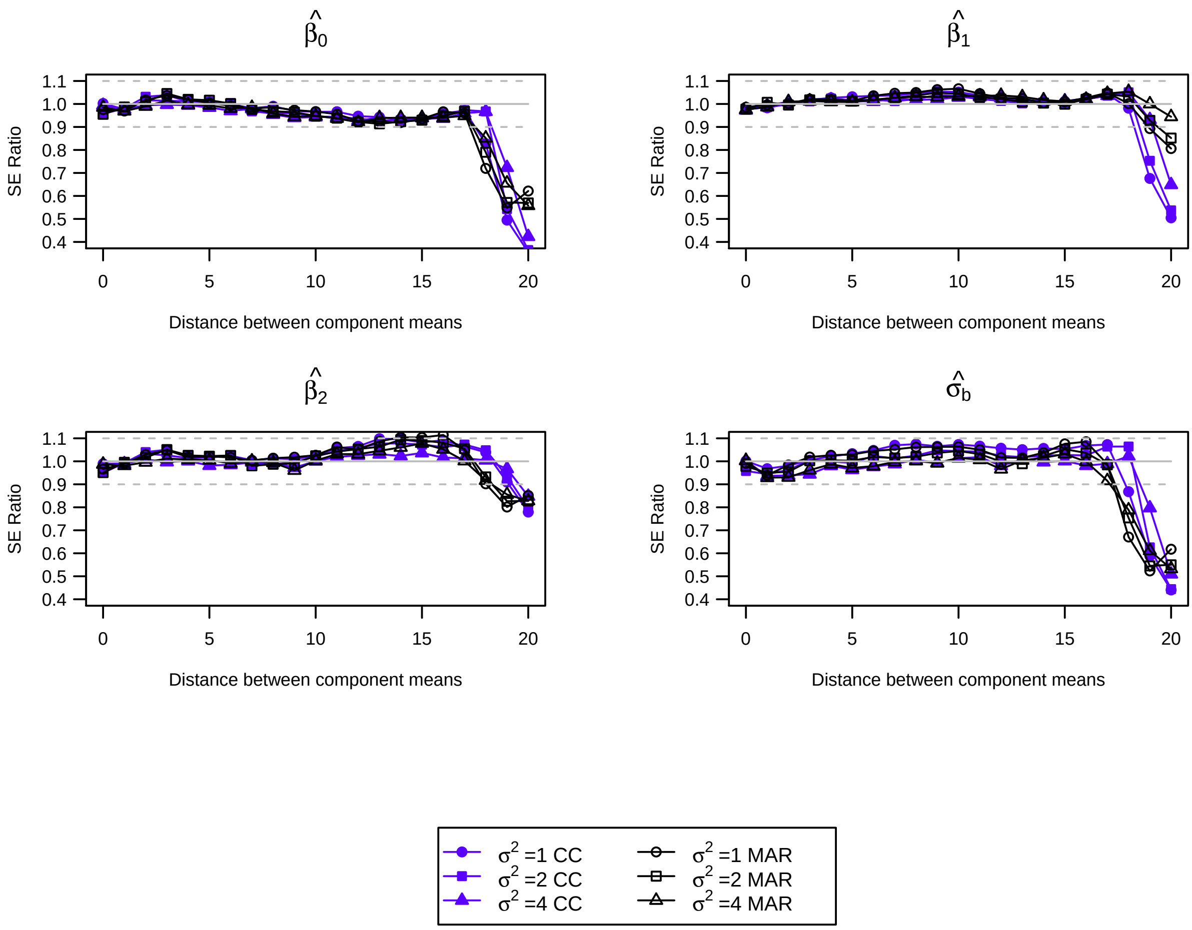


Figure E.3: Ratio of mean model-based standard error to the empirical standard error for parameter coefficients of random intercept logistic model for complete data (C) and MAR attrition of 30% (MAR) for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under three component variance scenarios ($\sigma^2 = 1, 2$ or 4) in the clinical trial simulation study. Grey horizontal solid line at ratio=1 and grey horizontal dashed lines at ratio of 0.9 and 1.1.

Figure E.3 presents the ratio of the mean standard error to the empirical standard error for the increasing distance between the component means. Assuming normal random intercepts in the presence of misspecification generally resulted in accurate model based standard errors. However, for severe departures from normality ($\mu_3 - \mu_1 \geq 18$), inaccurate model based standard errors were produced. These large decreases in accuracy correspond with the exponential increase of the empirical standard errors for all parameters as the component mean distances exceeded 18. The large variability of the parameter estimates and associated standard errors, resulted in skewed sampling distribution, and hence non-parametric summary measures of the model based standard error and empirical standard error may be better performance measures for these extreme true random intercept distributions. Again, patterns for the accuracy of the model based standard errors were similar across the three variance component scenarios and the two missing data scenarios.

The simulated attrition rate of 30% was similar to the observed attrition rate in the HILDA panel survey, and the additional impact of MAR attrition was minimal as observed in the HILDA simulations considered in Chapters 5 and 6.

There were excellent convergence rates for all but eight iterations in the simulation study. Out of the 1000 simulations for the most extreme distance between the component means (i.e. $\mu_1, \mu_2, \mu_3 = -10, 0, 10$), there were four instances each when analysing MAR data for variance components $\sigma^2 = 1$ and 2.

E.3 Summary

Within a clinical trial scenario, the impact of incorrectly assuming normally distributed random intercepts when the true distribution is a symmetrical three component mixture of normals was similar to that for the panel survey setting. Severe departures from the assumed normal distribution (component mean distances ≥ 18) resulted in large variability of parameter estimates and the corresponding standard errors for all variance and missing data scenarios. These extreme true random intercept distributions resulted in biased estimates, poor coverage rates and inaccurate model based standard errors. Similar to the simulation studies considered in Chapters 5 and 6, estimation of the random intercept variance component in the presence of misspecified random intercept distributions were biased and had extremely poor coverage rates.

Appendix F: Simulation study assessing impact of misspecified random intercept distributions for true symmetric three component mixture with fixed total random effect variances

The aim of this additional secondary simulation study is to assess whether the varying total random effect variance in the simulation study presented in Chapter 6 subsequently affects the observed results. For instance, the simulation study in Chapter 6 kept the component variances fixed at either $\sigma^2 = 1, 2$ or 4, resulting in unequal total variance (as σ_b^2 is a function of the component means and component variances, as derived by $\sigma_b^2 = \sigma^2 + \frac{2}{3}\mu_3^2$). To investigate the potential impact, the simulation study presented here considers the same panel survey setting motivated by the HILDA case study and the same random intercept distribution as considered in Chapter 6. However, by altering the component variances, this secondary simulation study fixes the total random effect variance at $\sigma_b^2 = 20, 25, 30$ and 35.

F.1 Simulation study design

Utilising the same simulation study design as described in Section 6.2, clustered binary response data, representing employment status over 11 years, were generated for 1000 women using the random intercept logistic model presented in Equation 4.1. The same design matrix and fixed effect parameter values detailed in Secion 6.2 were used to generate the response data.

In this secondary simulation study, the random intercept b_i was simulated from a symmetric three component mixture of normals (Equation 6.1). The variance components of the mixture distribution used to generate the random intercepts were selected such that the total random effect variance was fixed at $\sigma_b^2 = 20, 25, 30$ and 35. To ensure a variety of distributions that would allow the total true random effect variances to be fixed, the component means were restricted to μ_3 ranging from 0 to 5, increasing in increments of 0.5. Higher values of μ_3 would generate higher σ_b^2 values than those considered here. The four variance scenarios were selected to represent variability observed in panel survey settings, and are similar to the magnitudes considered in the simulation study of Litiére et al. (2008).

Table F.1: Component variances (σ^2) used to generate the random intercepts in the secondary simulation study for true random effect variances fixed at $\sigma_b^2 = 20, 25, 30$ and 35 for component mean distances ranging from 0 to 10. The component variances were determined by $\sigma^2 = \sigma_b^2 - \frac{2}{3}\mu_3^2$

Component Mean	True Random Variance			
Distance	$\sigma_b^2 = 20$	$\sigma_b^2 = 25$	$\sigma_b^2 = 30$	$\sigma_b^2 = 35$
$\mu_3 - \mu_1 = 0$	20.00	25.00	30.00	35.00
$\mu_3 - \mu_1 = 1$	19.83	24.83	29.83	34.83
$\mu_3 - \mu_1 = 2$	19.33	24.33	29.33	34.33
$\mu_3 - \mu_1 = 3$	18.50	23.50	28.50	33.50
$\mu_3 - \mu_1 = 4$	17.33	22.33	27.33	32.33
$\mu_3 - \mu_1 = 5$	15.83	20.83	25.83	30.83
$\mu_3 - \mu_1 = 6$	14.00	19.00	24.00	29.00
$\mu_3 - \mu_1 = 7$	11.83	16.83	21.83	26.83
$\mu_3 - \mu_1 = 8$	9.33	14.33	19.33	24.33
$\mu_3 - \mu_1 = 9$	6.50	11.50	16.50	21.50
$\mu_3 - \mu_1 = 10$	3.33	8.33	13.33	18.33

For the four true random variance scenarios considered in this simulation study, the values of the variance components ($\sigma_1^2 = \sigma_2^2 = \sigma_3^2$) for the 11 mean component scenarios are shown in Table F.1. The density distributions of the simulated random intercepts for selected mean scenarios ($\mu_1 = -\mu_3 = 0, 5$ and 10) are shown in Figure F.1. In comparison to the primary simulation study considered in Chapter 6, for similar component mean distances of $\mu_3 - \mu_1 = 0$ to 10, the total random effect variance and component variances are larger (i.e. see Table 6.1) and subsequently, the distributions are not as extreme as those with distinct modes shown in Figure 6.1.

As in the primary study, simulations were performed under two missing data scenarios: complete data and incomplete data due to attrition. Attrition was generated from a MAR mechanism by implementing the same drop-out model detailed in Sections 3.4.2 and 6.2.

Across the eleven component mean scenarios, four total variance scenarios and the two missing data scenarios (complete data and incomplete data due to attrition), 4400 datasets each containing 1000 subjects were generated. A random intercept logistic model assuming Gaussian random effects was fitted to each simulated dataset. The robustness of assuming normality in the presence of misspecified random intercepts was assessed by using the same performance measures as described in Section 6.2.

Simulations and analyses were conducted in SAS (Version 9.4, SAS Institute, Cary NC). All random intercept logistic models were fitted using the SAS procedure NLMIXED with adaptive Gaussian Quadrature using 20 quadrature points.

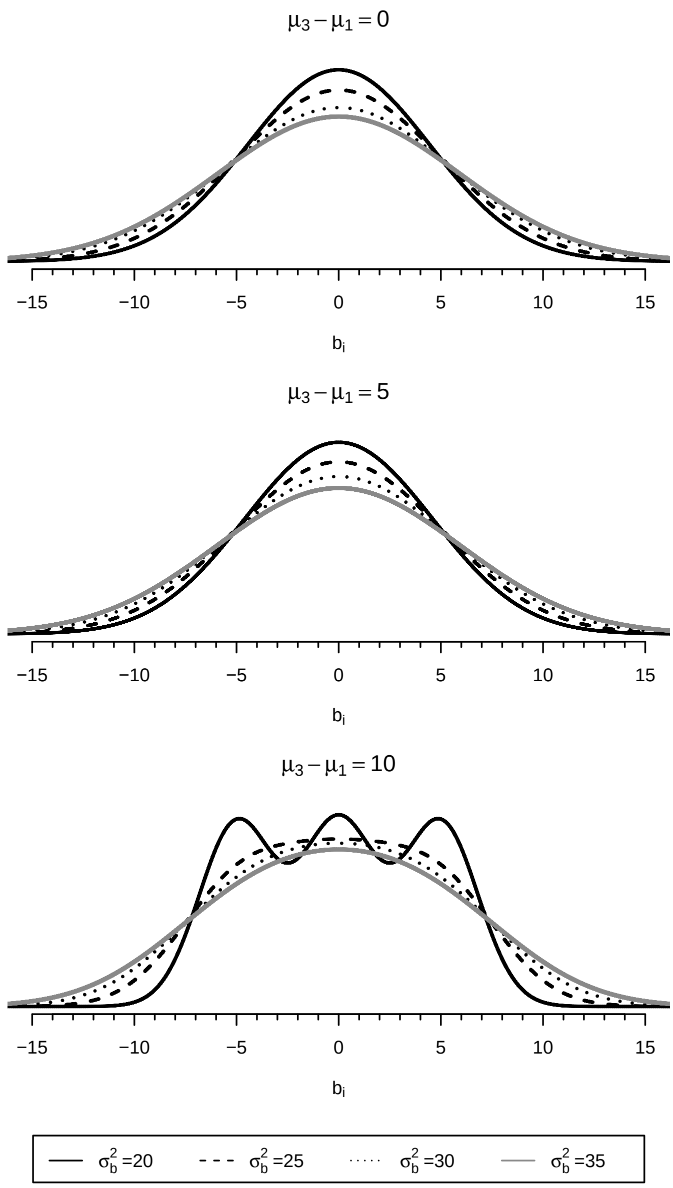


Figure F.1: Density of the true random intercept distributions for component mean distances $\mu_3 - \mu_1 = 0, 5$ and 10 for the four true total variance scenarios: $\sigma_b^2 = 20, 25, 30$ and 35.

F.2 Results and discussion

The performance measures of percentage bias, coverage rates of the 95% confidence intervals and the ratio of the mean model based standard errors to the empirical standard errors for the complete data scenario are shown in Figures F.2, F.3 and F.4 respectively. The results are similar as observed in Chapter 6 for minor departures in the true random intercept distribution from the assumed normal distribution (i.e. for comparable component mean distances of $\mu_3 - \mu_1 = 0$ to 10). The results presented here suggest misspecification of the random intercept distribution has some impact on the inference of parameters associated with the random effect.

Bias in the estimation of the intercept constant β_0 was observed for all component mean distances and variance component scenarios (Figure F.2). As observed in Chapter 6, minimal bias was observed for the other parameters with the exception of β_3. Similarly, negligible bias in estimating the random intercept standard deviation σ_b was observed for component mean distances $\mu_3 - \mu_1 < 10$.

Close to nominal coverage rates for all parameters were observed, with the exception of σ_b where the larger departures from the assumed normal distribution (i.e. $\mu_3 - \mu_1 > 8$ and $\sigma_b^2 = 20, 25$) resulted in low coverage rates (Figure F.3). The lower coverage rates for σ_b are consistent with the trend in the coverage rate observed in Chapter 6 for similar component mean distances.

Similarly, for small deviations from the shape of normality, accurate model based standard errors were produced for all parameters with the exception of σ_b (Figure F.4). As observed in Chapter 6 for similar component mean distances, there were some fluctuations of the standard error ratio for σ_b, with inaccurate model based errors produced for the most extreme true random intercept distribution considered ($\mu_3 - \mu_1 > 8$ and $\sigma_b^2 = 20$).

The magnitude and trend in the performance measures were similar for the four total random effect variances. There were some deviations in the trend for coverage and standard error ratio for smaller total random effect variances, corresponding to the largest departures from normality. For instance, some deviations in the magnitude of coverage and standard error ratio were observed for the smallest total random effect variance of $\sigma^2 = 20$ and the most extreme component mean distance of $\mu_3 - \mu_1 = 10$.

As in Chapters 5 and 6, MAR attrition had minimal additional impact. The impact of misspecifying the random intercept distribution in the presence of MAR attrition was similar to that in the complete data scenario, producing similar magnitudes and trends for percentage bias, coverage rates and accuracy of model based standard errors (results not shown).

The actual rate of attrition in the simulated datasets for all scenarios averaged 32.8% (range:

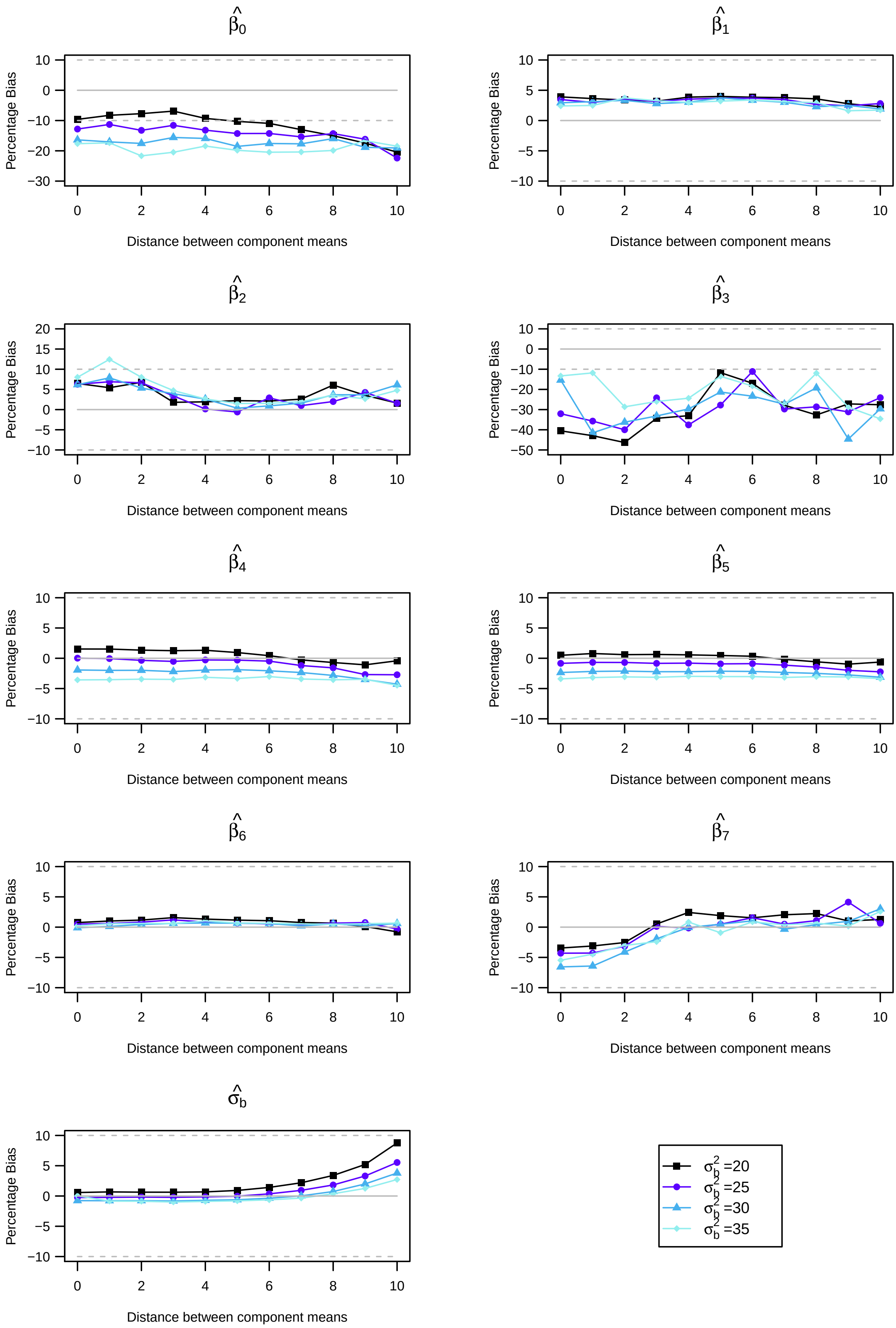


Figure F.2: Percentage bias for parameter coefficients of random intercept logistic model applied to the Complete data for increasing distances of random intercept component means $(\mu_3 - \mu_1)$ under four total random effect variance scenarios ($\sigma_b^2 = 20, 25, 30$ or 35). Grey horizontal solid line at percentage bias=0 and grey horizontal dashed lines at $\pm 10\%$.

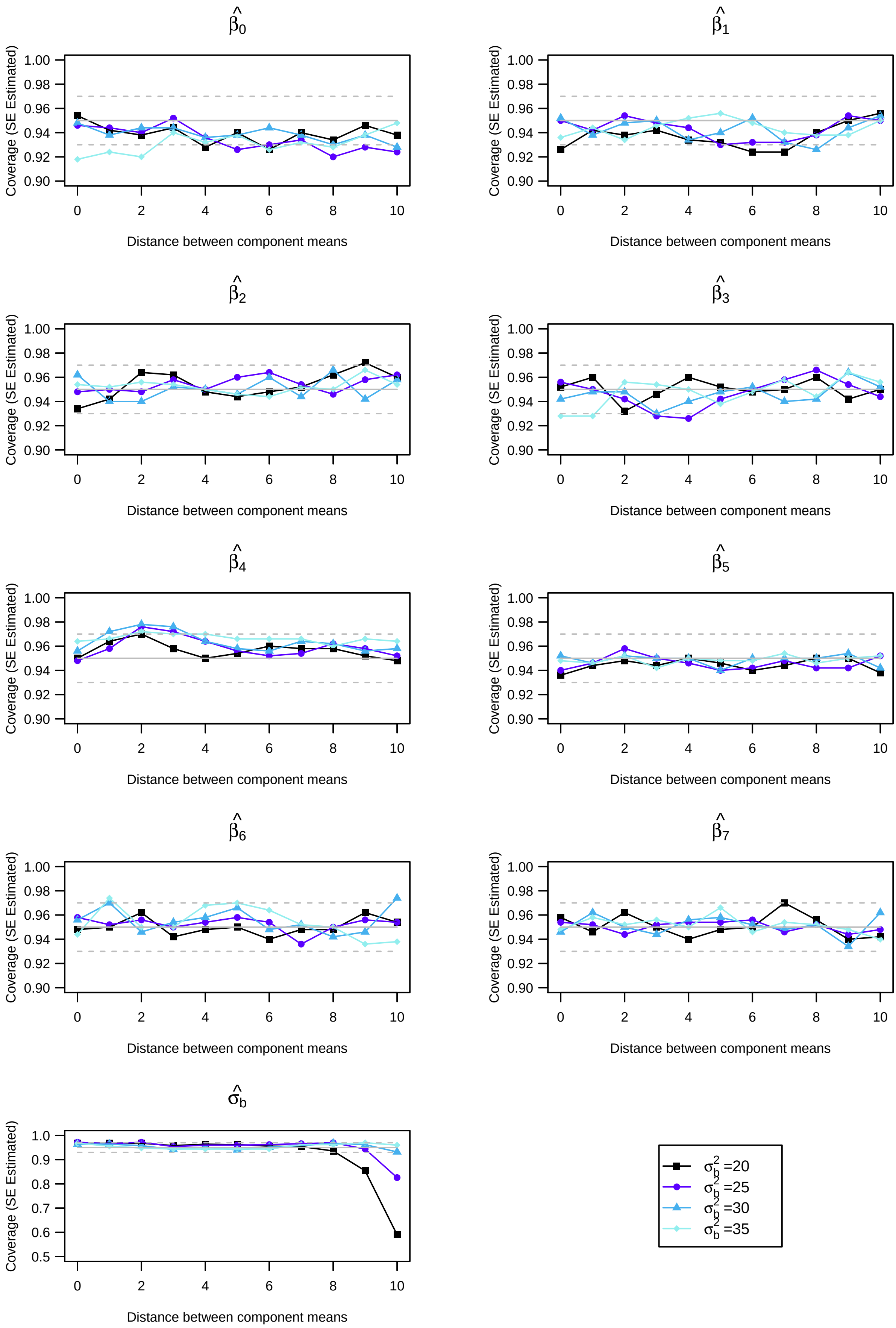


Figure F.3: Coverage rates of 95% confidence intervals for parameter coefficients of random intercept logistic model applied to the Complete data for increasing distances of random intercept component means ($\mu_3 - \mu_1$) under four total random effect variance scenarios ($\sigma_b^2 = 20, 25, 30$ or 35). Grey horizontal solid line at nominal coverage rate of 95% and grey horizontal dashed lines at coverage rates of 0.93 and 0.97.

27.9% to 37.3%) and was similar to the observed rate of 29.5% in the HILDA subgroup of working aged women (Table 4.1). The attrition rates for the eleven component means and the four total random effect variance scenarios were similar (results not shown).

The rate of subjects staying in unemployment for all observed time-points for the complete data and the MAR attrition data averaged 12.6% (range: 6.8% to 19.8%) and 17.2% (range: 10.4 to 24.3%), respectively. As in the primary simulation study, the number of observed stayers in the unemployed group was influenced by the true random effect distribution, with higher proportion of continuously unemployed for more extreme distributions, such as smaller total random effect variance and larger component mean distances (results not shown). As in the primary simulation study, the rate of subjects staying employed for all observed time-points averaged 51.1% (range: 44.8% to 57.2%) for the complete data, and 53.6% (47.0% to 60.5%) for the MAR attrition data scenario.

There were excellent convergence rates of 100% for all iterations of the simulation study.

F.3 Summary

By restricting the total random effect variance to $\sigma_b^2 = 20$ to 35, the resulting simulated random intercepts considered in this secondary simulation study were not as extreme as the multimodal distributions considered in the primary simulation. Thus, the negligible impact of incorrectly assuming normally distributed random intercepts for minor departures from normality were similar regardless of whether the total random effect variance was fixed or not. The results from this simulation study suggest that the varying total random effect variance for the simulated random effect distributions considered in Chapter 6 does not additionally influence the observed results.

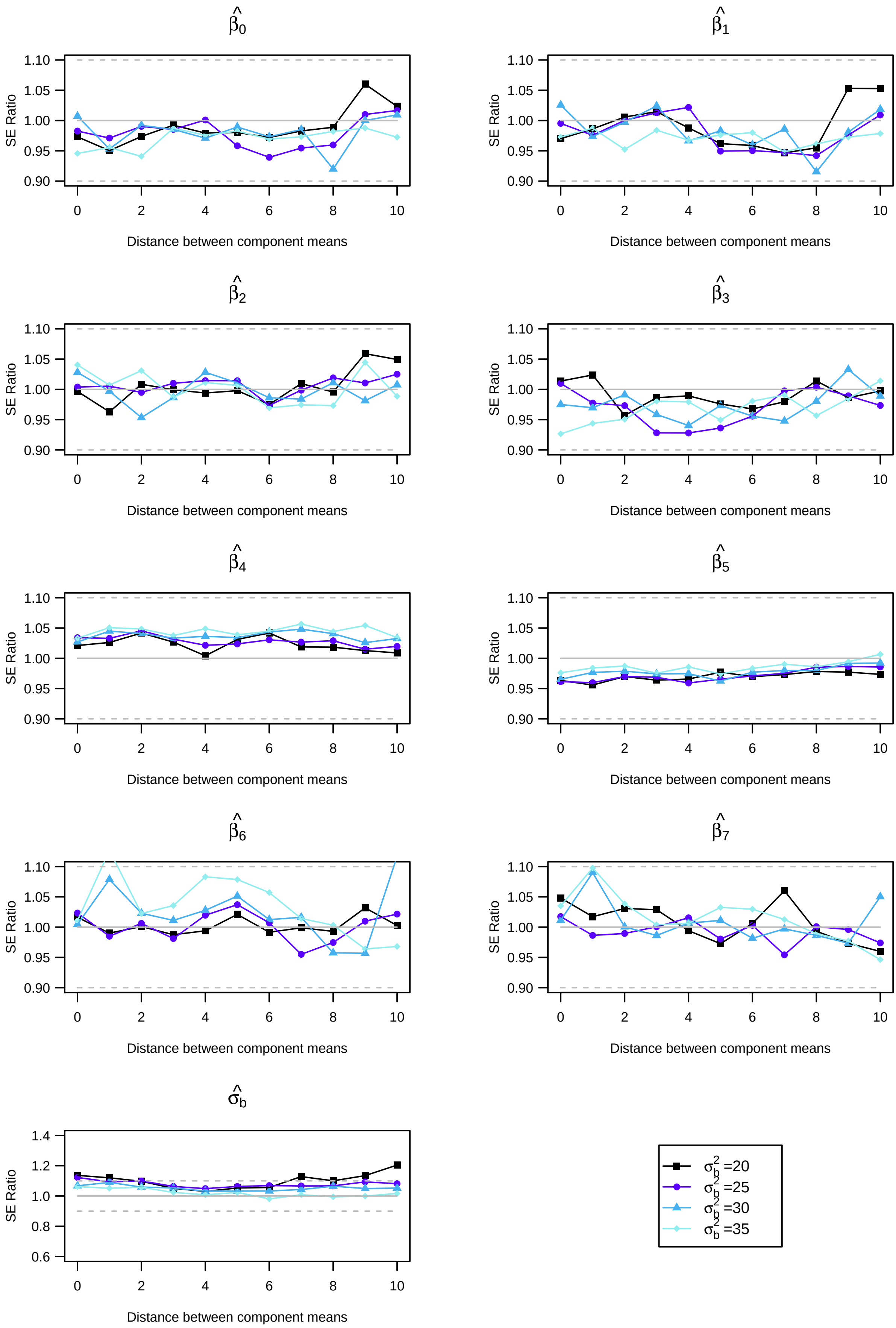


Figure F.4: Ratio of mean model-based standard error to the empirical standard error for parameter coefficients of random intercept logistic model applied to the Complete data for increasing distances of random intercept component means $(\mu_3 - \mu_1)$ under four total random effect variance scenarios ($\sigma_b^2 = 20, 25, 30$ or 35). Grey horizontal solid line at ratio=1 and grey horizontal dashed lines at ratio of 0.9 and 1.1.

Appendix G: Sensitivity of the Vertex Exchange Method in random intercept logistic models

The Vertex Exchange Method (VEM) was applied in Chapter 7 to estimate the logistic models assuming unspecified distributions of the random intercept. It was based on an initial grid consisting of 301 equally spaced support points based on the Cholesky decomposition, with the range set at ± 5 standard deviations of the assumed normal random intercept distribution. Furthermore, the starting values for the parameter coefficients were set to the estimates from the equivalent random intercept logistic model assuming normally distributed random intercepts. To assess the impact of the choice of the initial starting values and the initial grid used in the estimation, a sensitivity analysis was performed by re-fitting the VEM algorithm to the HILDA case study with alternative starting values and initial grids. Two sets of starting values were considered, parameter coefficients of the equivalent model with random intercepts assumed to be normally distributed (random intercept logistic-normal model), or no random intercepts (binary logistic model). A total of six different combinations of initial grid size and grid range were considered, consisting of three initial number grid points, K=101, 301 or 501, and the range of the grid was either 5 or 7 standard deviations of the equivalent model with assumed normal random intercepts. The sensitivity analyses were performed for both missing data scenarios: the women with complete cases and the women with monotone missing data. In total, 24 scenarios were considered.

G.1 Complete case data

The parameter estimates and standard errors of the 12 fitted VEM random intercept logistic models for women with complete cases are presented in Tables G.1 and G.2. Table G.1 contains the results for when the starting values are set to parameter estimates of the random intercept logistic-normal model, and Table G.2 contains the results for when the starting values are set to the parameter estimates of the standard binary logistic model. The VEM algorithm was robust to the choice of starting values and the initial grid choice, producing similar residual deviance (-2ll) for all models, ranging from 9663 to 9682. There was more variability in the residual deviance for models with the grid size of 101, particularly for the starting values based on the random intercept logistic-normal model, however the residual deviance stabilised for larger grid sizes of 301 and 501 ($-2ll$ ranging from 9663 to 9667). The parameter estimates and standard errors for the fixed effects not related to the random intercept were similar (when rounded to one decimal place) across the number of grid points and the grid range. There were some minor differences observed for the constant coefficient (estimates ranging between -0.401 and -0.782) and the random intercept variability (estimates ranging from 12.135 to 13.589). The computational time for calculating the model substantially increased with the number of initial grid points. Models with 101 initial grid points required less than an hour of CPU time, whilst the models with 501 initial grid points required over 13 hours.

G.2 Monotone missing data

The parameter estimates and standard errors of the 12 fitted VEM random intercept logistic models for women with monotone missing data are presented in Tables G.3 and G.4. Table G.3 contains the results for when the starting values are set to parameter estimates of the random intercept logistic-normal model, and Table G.4 contains the results for when the starting values are set to the parameter estimates of the standard binary logistic model. Similar robustness to the complete case data scenario was observed for the monotone missing scenario, with similar residual deviance for all models (ranging from 11515 to 11527). There were some minor differences observed for coefficients relating to the random intercept, the intercept coefficient ranged from -0.434 to -1.100, and the random intercept variance estimate ranged from 11.100 to 13.536. For the two different starting values, there were minimal differences for the coefficients and standard errors for the parameters unrelated to the random intercept, particularly for the larger grid points of 301 and 501. For the same choice of initial grid, differences in the residual deviance between the two starting values ranging between -2 to 4 (random intercept logistic model - ordinary logistic model). As the sample size for the monotone missing sub-group is larger (n=1927), the computational time for estimating the model took longer than for the complete cases. The models with 101 initial grid points required approximately an hour of CPU time, compared to over 16 hours required for the models with 501 initial grid points.

G.3 Summary

The VEM applied to estimate the unspecified distribution of the random intercept logistic model is robust to the choice of initial grid, including the grid size and the range of grid points, and the choice of starting values. The parameters related to the random intercept were sensitive to the number of grid points, with variability of the parameter estimates and standard errors shown for 101 grid points. Similar robustness was demonstrated for both of the missing data scenarios. Therefore the results presented in Chapter 7, based on 301 grid points with the range of grid points set at 5 standard deviations and starting values based on the random intercept logistic-normal model, are similar to those reported for another set of starting values and models based on similar, or larger, number of grid points and range of grid points.

Table G.1: Sensitivity of VEM to the choice of the initial grid, including the number of grid points (K=101, 301 or 501) and the range of the grid points (± 5 or 7 standard deviations), when applied to the 1359 women with complete case data in the HILDA case study with starting values based on the random intercept logistic model with normal distributed random intercepts. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Random intercept logistic-normal model**											
Grid Points	**101**				**301**				**501**			
Grid Range	$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-1.124		-1.645		-0.895		-0.877		-0.806		-0.809	
Age	0.127	0.003	0.139	0.003	0.120	0.003	0.121	0.003	0.118	0.003	0.120	0.003
Marital Status												
Married/Defacto												
Sep/Div/Wid	-0.351	0.140	-0.371	0.140	-0.347	0.140	-0.350	0.139	-0.346	0.139	-0.346	0.139
Single	0.229	0.238	0.295	0.255	0.201	0.232	0.200	0.234	0.188	0.231	0.193	0.232
Highest Education												
Bachelor or higher												
Year 12/Dip/Cert	-1.082	0.160	-1.054	0.166	-1.092	0.159	-1.091	0.158	-1.096	0.159	-1.094	0.158
Year 11 or less	-2.435	0.169	-2.457	0.175	-2.428	0.168	-2.439	0.167	-2.431	0.168	-2.432	0.167
Dependent Children												
None												
Youngest<5	-2.135	0.116	-2.073	0.118	-2.169	0.116	-2.172	0.116	-2.184	0.116	-2.179	0.116
Youngest 5-24	-0.282	0.103	-0.247	0.105	-0.303	0.103	-0.301	0.103	-0.311	0.103	-0.307	0.103
Random Effect												
Variance	12.415		11.949		14.197		12.674		14.409		12.711	
$-2ll$	9671		9682		9667		9667		9666		9666	
CPU Time (hh:mm)	0:44		0:51		6:28		5:51		13:46		15:29	

Table G.2: Sensitivity of VEM to the choice of the initial grid, including the number of grid points (K=101, 301 or 501) and the range of the grid points (± 5 or 7 standard deviations), when applied to the 1359 women with complete case data in the HILDA case study with starting values based on the ordinary binary logistic model (with no random effect). The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Binary logistic model**											
Grid Points	**101**				**301**				**501**			
Grid Range	$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-0.732		-0.651		-0.517		-0.672		-0.401		-0.782	
Age	0.118	0.003	0.118	0.003	0.112	0.003	0.116	0.003	0.110	0.003	0.119	0.003
Marital Status												
Married/Defacto												
Sep/Div/Wid	-0.345	0.140	-0.345	0.140	-0.344	0.139	-0.344	0.139	-0.340	0.139	-0.346	0.139
Single	0.189	0.229	0.185	0.230	0.160	0.227	0.177	0.230	0.150	0.226	0.189	0.232
Highest Education												
Bachelor or higher												
Year 12/Dip/Cert	-1.097	0.158	-1.107	0.162	-1.108	0.159	-1.101	0.159	-1.114	0.160	-1.093	0.159
Year 11 or less	-2.428	0.167	-2.454	0.171	-2.435	0.167	-2.432	0.168	-2.436	0.169	-2.429	0.168
Dependent Children												
None												
Youngest<5	-2.186	0.115	-2.199	0.117	-2.224	0.116	-2.198	0.116	-2.235	0.116	-2.180	0.116
Youngest 5-24	-0.312	0.102	-0.318	0.103	-0.332	0.103	-0.317	0.103	-0.341	0.103	-0.308	0.103
Random Effect												
$\sigma^2_{b_0}$	12.538		12.135		13.440		12.508		13.589		12.935	
-2*ll*	9665		9666		9663		9665		9663		9666	
CPU Time (hh:mm)	0:46		0:43		4:45		6:05		13:55		15:48	

Table G.3: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=101, 301 or 501) and the range of the grid points (± 5 or 7 standard deviations), when applied to the 1927 women with monotone missing data in the HILDA case study with starting values based on the random intercept logistic model with normal distributed random intercepts. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Random intercept logistic-normal model**											
Grid Points	**101**				**301**				**501**			
Grid Range	$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-0.659		-1.100		-0.686		-0.589		-0.706		-0.434	
Age	0.118	0.003	0.128	0.003	0.118	0.003	0.116	0.003	0.118	0.003	0.113	0.003
Marital Status												
Married/Defacto												
Sep/Div/Wid	-0.292	0.134	-0.297	0.136	-0.289	0.135	-0.292	0.135	-0.291	0.134	-0.288	0.135
Single	0.054	0.219	0.105	0.233	0.069	0.219	0.060	0.218	0.066	0.219	0.054	0.214
Highest Education												
Bachelor or higher												
Year 12/Dip/Cert	-1.162	0.152	-1.150	0.157	-1.166	0.152	-1.172	0.153	-1.163	0.152	-1.181	0.153
Year 11 or less	-2.551	0.158	-2.561	0.163	-2.541	0.159	-2.544	0.157	-2.542	0.158	-2.544	0.158
Dependent Children												
None												
Youngest<5	-2.199	0.111	-2.132	0.112	-2.186	0.111	-2.201	0.111	-2.187	0.111	-2.213	0.111
Youngest 5-24	-0.344	0.100	-0.312	0.101	-0.343	0.100	-0.349	0.100	-0.341	0.100	-0.358	0.100
Random Effect												
Variance	11.520		11.591		11.937		12.000		12.329		11.973	
$-2ll$	11518		11527		11518		11517		11518		11515	
CPU Time (hh:mm)	1:21		1:03		8:10		6:56		18:12		16:44	

Table G.4: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=101, 301 or 501) and the range of the grid points (± 5 or 7 standard deviations), when applied to the 1927 women with monotone missing data in the HILDA case study with starting values based on the ordinary binary logistic model (with no random effect). The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance with the CPU computational time.

Starting Values	**Binary Logistic Model**											
Grid Points	**101**				**301**				**501**			
Grid Range	$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$		$\pm 5\sigma_b$		$\pm 7\sigma_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	-0.811		-0.823		-0.604		-0.508		-0.609		-0.520	
Age	0.120	0.003	0.122	0.003	0.115	0.003	0.115	0.003	0.115	0.003	0.114	0.003
Marital Status												
Married/Defacto												
Sep/Div/Wid	-0.300	0.136	-0.294	0.136	-0.292	0.135	-0.289	0.135	-0.290	0.135	-0.288	0.134
Single	0.094	0.231	0.085	0.233	0.069	0.217	0.058	0.216	0.070	0.217	0.057	0.216
Highest Education												
Bachelor or higher												
Year 12/Dip/Cert	-1.163	0.155	-1.152	0.158	-1.179	0.153	-1.174	0.153	-1.178	0.153	-1.172	0.153
Year 11 or less	-2.549	0.160	-2.521	0.164	-2.544	0.159	-2.541	0.158	-2.543	0.159	-2.535	0.158
Dependent Children												
None												
Youngest<5	-2.183	0.110	-2.145	0.111	-2.205	0.111	-2.205	0.111	-2.202	0.111	-2.200	0.111
Youngest 5-24	-0.337	0.100	-0.324	0.101	-0.353	0.100	-0.353	0.100	-0.352	0.100	-0.350	0.100
Random Effect												
σ^2_{b0}	12.047		11.100		13.536		11.907		13.244		12.101	
-2ll	11520		11523		11516		11516		11516		11516	
CPU Time (hh:mm)	1:21		0:57		7:42		6:35		18:44		17:18	

Appendix H: Sensitivity of the Vertex Exchange Method in random intercept and random slope logistic models

The Vertex Exchange Method (VEM) applied in Chapter 7 to estimate logistic mixed models with random intercepts and random slopes was based on the initial grid consisting of 31 equally spaced support points in both dimensions. Thus, the initial grid consisted of 961 support points with the range of the grid set at ± 5 times the Cholesky decomposition of the variance-covariance matrix corresponding to the equivalent model assuming bivariate normal random effects ($\hat{\boldsymbol{S}}_{\boldsymbol{b}}$). To assess the impact of the initial grid and starting values used in VEM estimation, a sensitivity analysis was performed by re-fitting the VEM algorithm to the HILDA case study with alternative initial grids and starting values of the fixed effect parameters. The initial grids consisted of either 31 or 51 initial grid points in each dimension, such that the two dimensional grid for the two random effects were either K=31 × 31 or 51 × 51, and the range of the grid was based on the Cholesky decomposition multiplied by a factor of 5 or 7, $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ or $\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$. Furthermore, two alternative sets of starting values for the fixed effect parameters were considered, either the starting values were the estimated coefficients of the equivalent logistic mixed model assuming bivariate normal random effects, or the standard binary logistic model. Therefore a total of four different combinations of the initial grid and two initial starting values were considered. The sensitivity analyses were performed for both missing data scenarios, the women with complete cases and the women with monotone missing data. The corresponding results for the two data analysis sub-groups are presented in Section H.1 and H.2, respectively.

H.1 Complete case data

The parameter estimates and standard errors of the logistic mixed models estimated by the VEM approach for women with complete case data are presented in Table H.1. The VEM algorithm was robust to the initial grid choice, particularly to the number of grid points used. A comparison of the initial grid values within each starting value subset shows stability of the fixed effect coefficients, producing similar residual deviance for the models. For models using the logistic mixed model as starting values the residual deviance ranged from 8989 to 8995, and the residual deviance for models utilising the standard logistic model as starting values ranged from 8990 to 9007. The VEM algorithm applied to the logistic mixed model with bivariate random effects was susceptible to boundary issues. As such, larger initial grids ($\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$) resulted in larger magnitude of estimates for parameters relating to the random effects, with the exception of the constant coefficient. For instance, all estimates of the variance-covariance matrix were larger for the corresponding model based on the $\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ than on the $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ model. The choice of starting values did produce differences in the magnitude of the coefficient estimates, however at the conventional 5% significance level, the same inference would be made for all models. Furthermore, the absolute difference in the coefficient and standard error estimates for the two sets of starting values generally decreased as the initial number of grid points and range of grid points increased. The computational time for calculating the model substantially increased with the number of initial grid points, with marginal differences in CPU time observed for the two choices of starting values.

H.2 Monotone missing data

The parameter estimates and standard errors of the logistic mixed models estimated by the VEM approach for women with monotone missingness are presented in Table H.2. Similar conclusions as for the complete case data scenario were observed for the monotone missing scenario. The VEM approach was robust to the initial grid choice, though coefficient estimates differed depending on the initial starting values of the coefficients. In comparison to the complete case scenario, the absolute differences between the equivalent models with the different initial starting values were not as extreme, potentially due to the larger number of observations in the Monotone missing data scenario. Albeit the differences in the coefficients and standard error estimates, at the 5% significance level, the same inferential conclusions would be made for all VEM models. Within each initial starting value subgroup, the VEM approach was robust to the number of initial grid points resulting in similar residual deviances. The residual deviance ranged from 10736 to 17041 for models with starting values based on the logistic mixed model, and ranged from 10736 to 10754 for models with starting values based on the standard logistic model. As observed for the complete cases, the VEM approach was susceptible to boundary solutions, and therefore, the choice of the initial grid range impacted the estimation of parameters related to the random effects. In comparison to the grid range of $\pm 5\hat{\boldsymbol{S}}_{\boldsymbol{b}}$, the larger grid range of $\pm 7\hat{\boldsymbol{S}}_{\boldsymbol{b}}$ resulted in larger estimates for all components of the variance-covariance matrix. As expected, the larger sample size for the monotone missing data scenario required more CPU time that for the complete case scenario. The CPU time required for model convergence increased with the initial number of grid points and the grid range.

H.3 Summary

The VEM used to estimate the random intercept and random slope logistic model was robust to the choice of the initial grid, including the number of grid points and the range of the initial grid, and to the choice of starting values. Increasing the number of initial grid points to consist of 51 equally spaced support points resulted in a marginal improvement of the residual deviance, however the CPU time required to fit the model was approximately 4 times longer. The results presented in Chapter 7 adequately represent the performance of the VEM to estimate the bivariate random effects distribution in the HILDA case study, and the VEM is computationally practical.

Table H.1: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=31 or 51) and the range of the grid points ($\pm 5\hat{S}_b$ or $7\hat{S}_b$) in addition to the choice of starting values, either based on coefficients of a random intercept and random slope logistic model assuming bivariate normal random effects or the standard logistic model, when applied to the 1359 women with complete case data in the HILDA case study. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, along with the CPU computational time.

Starting Values	**Logistic-normal mixed model**								**Binary logistic model**							
Grid Points	**31 x 31**				**51 x 51**				**31 x 31**				**51 x 51**			
Grid Range	$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	2.964		2.747		2.977		2.549		2.645		2.236		2.671		2.547	
(Wave-1)/10	2.597		3.367		2.564		2.892		2.187		2.703		2.599		2.787	
(Age Baseline-30)/10	1.252	0.144	0.798	0.149	1.226	0.140	1.281	0.146	1.215	0.142	1.283	0.160	1.215	0.138	1.040	0.147
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.445	0.174	-0.406	0.178	-0.409	0.173	-0.440	0.177	-0.391	0.167	-0.495	0.177	-0.423	0.168	-0.415	0.175
Single	0.492	0.272	0.345	0.257	0.390	0.243	0.300	0.256	0.413	0.268	0.250	0.286	0.457	0.247	0.427	0.258
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-1.154	0.172	-1.147	0.175	-1.149	0.166	-1.178	0.170	-0.978	0.168	-1.001	0.191	-1.089	0.162	-1.115	0.171
Year 11 or less	-2.812	0.193	-2.760	0.177	-2.854	0.181	-2.840	0.187	-2.646	0.186	-2.648	0.212	-2.692	0.181	-2.715	0.186
Dependent Children																
None																
Youngest<5	-2.647	0.130	-2.628	0.130	-2.641	0.126	-2.619	0.128	-2.496	0.132	-2.505	0.149	-2.558	0.128	-2.597	0.126
Youngest 5-24	-0.710	0.128	-0.674	0.126	-0.681	0.123	-0.715	0.127	-0.620	0.127	-0.654	0.142	-0.645	0.125	-0.651	0.126
Random Effects																
$\sigma^2_{b_0}$	35.4		54.6		35.3		54.7		36.1		65.4		39.3		58.1	
$\sigma^2_{b_1}$	78.1		114.7		80.5		118.4		73.5		116.3		73.1		109.2	
σ_{b_0,b_1}	-22.9		-37.2		-23.5		-33.0		-13.7		-30.9		-22.4		-27.8	
-2ll	8995		8995		8989		8990		9001		9007		8994		8990	
CPU Time (hh:mm)	43:12		47:59		199:12		236:09		37:44		48:03		228:37		232:43	

Table H.2: Sensitivity of VEM to the choice of initial grid, including the number of grid points (K=31 or 51) and the range of the grid points ($\pm 5\hat{S}_b$ or $7\hat{S}_b$) in addition to the choice of starting values, either based on coefficients of a random intercept and random slope logistic model assuming bivariate normal random effects or the standard logistic model, when applied to the 1927 women with monotone missing data in the HILDA case study. The parameter estimates (Est) and corresponding standard errors (SE) are presented for the fixed effects and the random effect variance, along with the CPU computational time.

Starting Values	**Logistic-normal mixed model**								**Binary logistic model**							
Grid Points	**31 x 31**				**51 x 51**				**31 x 31**				**51 x 51**			
Grid Range	$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$		$\pm 5\hat{S}_b$		$\pm 7\hat{S}_b$	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
Constant	3.599		3.332		3.492		3.309		3.092		3.434		3.375		3.303	
(Wave-1)/10	2.820		2.725		2.381		2.703		2.035		2.205		2.407		2.625	
(Age Baseline-30)/10	0.705	0.146	0.742	0.149	0.915	0.139	0.923	0.144	1.236	0.140	0.930	0.167	0.800	0.143	0.860	0.151
Marital Status																
Married/Defacto																
Sep/Div/Wid	-0.466	0.162	-0.468	0.175	-0.422	0.167	-0.435	0.167	-0.420	0.165	-0.369	0.166	-0.427	0.168	-0.410	0.171
Single	0.260	0.228	0.270	0.253	0.286	0.230	0.206	0.229	0.254	0.239	0.235	0.258	0.297	0.226	0.184	0.242
Highest Education																
Bachelor or higher																
Year 12/Dip/Cert	-1.316	0.162	-1.273	0.167	-1.265	0.157	-1.333	0.164	-1.204	0.160	-1.440	0.183	-1.306	0.158	-1.350	0.167
Year 11 or less	-2.750	0.171	-2.861	0.170	-2.865	0.165	-2.915	0.168	-2.772	0.175	-2.970	0.222	-2.849	0.165	-2.907	0.171
Dependent Children																
None																
Youngest<5	-2.746	0.123	-2.771	0.124	-2.777	0.118	-2.799	0.122	-2.612	0.122	-2.721	0.140	-2.778	0.118	-2.776	0.123
Youngest 5-24	-0.812	0.123	-0.847	0.124	-0.818	0.118	-0.834	0.120	-0.749	0.120	-0.804	0.138	-0.813	0.119	-0.834	0.121
Random Effects																
$\sigma^2_{b_0}$	32.4		44.0		31.6		43.3		30.5		39.0		34.3		46.8	
$\sigma^2_{b_1}$	98.4		123.9		88.0		118.9		86.9		114.4		85.3		120.6	
σ_{b_0,b_1}	-25.5		-39.0		-27.1		-38.6		-20.9		-31.9		-24.3		-41.5	
$-2ll$	10740		10741		10736		10736		10750		10754		10738		10736	
CPU Time (hh:mm)	50:52		58:26		194:36		251:33		57:19		53:16		243:55		251:59	