# Binary Logistic Models with Partially Crossed Random Effects

by

Zizhe Zhang

**Master of Digital Experience Innovation, University of Waterloo, 2012**
**Master of Arts, Shanghai Jiao Tong University, 2009**
**Bachelor of Arts, Shanghai Ocean University, 2004**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF**

**Master of Science**

In the Graduate Academic Unit of Mathematics and Statistics

| | |
|---|---|
| Supervisor(s): | Guohua Yan, Ph.D, Statistics |
| | Renjun Ma, Ph.D, Statistics |
| Examining Board: | Vlad Tasic, Ph.D, Mathematics, Chair |
| | Jeffrey Picka, Ph.D, Statistics |
| | Murshed Chowdhury, Ph.D, Economics |

This thesis is accepted by the
Dean of Graduate Studies

**THE UNIVERSITY OF NEW BRUNSWICK**

**November, 2021**

# Abstract

Educational studies and behavioural scientists frequently encounter data with binary outcomes that have cross-classified data structures. For example, in a student admission study (success or failure), schools and areas could be treated as crossed random effects since not all students from the same school live in the same area and vice versa. It is crucial to incorporate crossed random effects into the model for data with cross-classified structures; otherwise, data analysis results might be misleading. This thesis proposes a binary logistic model with partially crossed random effects, which is further extended to a baseline-category logit model with partially crossed random effects for multinomial analysis. The random effects in our proposed models are predicted by the orthodox best linear unbiased predictor (BLUP) approach. Our models are robust because they only need to specify the first and second moments of the random effects. The simulation study shows that the estimation algorithm generally performs well. In addition, we apply these models to insurance data about motor vehicle accidents and interpret the estimates for practical references.

# Dedication

This thesis is dedicated to my family for their endless love and support.

# Acknowledgements

First, I would like to express genuine gratitude to my supervisors, Dr. Guohua Yan and Dr. Renjun Ma. They have provided precious support during my graduate studies and have guided me through my thesis writing. The regular meetings about my thesis have greatly broadened my knowledge and sharpened my research skills. Without their insightful suggestions and great assistance, this thesis would not have been finished.

Great thanks to Dr. Jeffrey Picka, who has always encouraged me and helped me think critically about statistics. I would also like to express my special thanks to Dr. Tariq Hasan, who has shown sincere and warm kindness during my graduate studies.

I would like to thank all faculty and staff in Mathematics and Statistics department, who have provided me with a friendly and supportive environment.

In the past two years, I have also got lots of help from my classmates and friends at UNB.

Finally, I would like to thank my family for their deep love and endless support.

# Table of Contents


**Abstract** ii

**Dedication** iii

**Acknowledgments** iv

**Table of Contents** v

**List of Tables** ix

**List of Figures** xi

**1 Introduction** 1

**2 Preliminaries** 6

2.1 Crossed Random Effects . . . . . . . . . . . . . . . . . . . . . . . . . . 6

2.2 Logistic Models with Crossed Random Effects . . . . . . . . . . . . . . 8

2.3 Beta-binomial Logistic Regrsssion . . . . . . . . . . . . . . . . . . . . . 12

2.4 Motor Vehicle Accident Data . . . . . . . . . . . . . . . . . . . . . . . 14

2.4.1 Data Description . . . . . . . . . . . . . . . . . . . . . . . . . 14

2.4.2 Crossed Random Effects . . . . . . . . . . . . . . . . . . . . . 15

**3 Binary Logistic Models with Partially Crossed Random Effects** 17

3.1 Model Specification . . . . . . . . . . . . . . . . . . . . . . . . . . . . 17

3.1.1 Assumption 1 . . . . . . . . . . . . . . . . . . . . . . . . . . . 18

3.1.2 Assumption 2 . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
3.1.3 Assumption 3 . . . . . . . . . . . . . . . . . . . . . . . . . . . 19
3.2 Covariance Structure . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.2.1 Derivation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
3.2.2 Matrix Expression . . . . . . . . . . . . . . . . . . . . . . . . 24
3.2.2.1 Var($\mathbf{W}$) in matrix form . . . . . . . . . . . . . . . . 24
3.2.2.2 Var($\mathbf{Y}$) in matrix form . . . . . . . . . . . . . . . . 25
3.3 The Best Linear Unbiased Predictor of Random Effects . . . . . . . . 27
3.3.1 The BLUP of Random Effect $U_i$ . . . . . . . . . . . . . . . . 27
3.3.1.1 Derivation . . . . . . . . . . . . . . . . . . . . . . . . 27
3.3.1.2 Matrix expression . . . . . . . . . . . . . . . . . . . . . 29
3.3.2 The BLUP of Random Effect $V_j$ . . . . . . . . . . . . . . . . 29
3.3.2.1 Derivation . . . . . . . . . . . . . . . . . . . . . . . . 29
3.3.2.2 Matrix expression . . . . . . . . . . . . . . . . . . . . . 31
3.4 Parameter Estimation . . . . . . . . . . . . . . . . . . . . . . . . . 32
3.4.1 Estimation of Regression Parameters . . . . . . . . . . . . . . 32
3.4.2 Estimation of Dispersion Parameters . . . . . . . . . . . . . . 33
3.5 Computational Procedure . . . . . . . . . . . . . . . . . . . . . . . 34
3.5.1 Initialization . . . . . . . . . . . . . . . . . . . . . . . . . . 34
3.5.1.1 Initialize random effects . . . . . . . . . . . . . . . . . 36
3.5.1.2 Initialize the regression parameter estimates $\widehat{\boldsymbol{\beta}}_{(0)}$ and the probabilty estimates $\widehat{\boldsymbol{\pi}}_0$ . . . . . . . . . . . . . . . 36
3.5.1.3 Initialize the dispersion parameter estimates . . . . . 37
3.5.2 Iteration . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
3.5.3 Output . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 37
3.6 Simulation Study . . . . . . . . . . . . . . . . . . . . . . . . . . . 38
3.6.1 Simulation Procedure . . . . . . . . . . . . . . . . . . . . . . 38

3.6.2 Simulation Performance . . . . . . . . . . . . . . . . . . . . . . 39

**4 Analysis of Injury Occurrence Data 41**

4.1 Exploratory Data Analysis . . . . . . . . . . . . . . . . . . . . . . 41

4.2 Model Setting . . . . . . . . . . . . . . . . . . . . . . . . . . . . 43

4.3 Data Analysis . . . . . . . . . . . . . . . . . . . . . . . . . . . . 45

4.3.1 Experiment Setting . . . . . . . . . . . . . . . . . . . . . . 45

4.3.2 Estimation Results . . . . . . . . . . . . . . . . . . . . . . 45

4.3.2.1 Full model . . . . . . . . . . . . . . . . . . . . . . 45

4.3.2.2 Final model . . . . . . . . . . . . . . . . . . . . . 47

4.3.2.3 Interpretation of parameter estimates . . . . . . . . . 49

4.3.2.4 Interpretation of random effects . . . . . . . . . . . . 51

**5 Analysis of Injury Severity Data 55**

5.1 Baseline-Category Logit Models with Partially Crossed Random Effects 55

5.1.1 Review of Baseline-Category Logit Models . . . . . . . . . . . 55

5.1.2 Baseline-Category Logit Models with Partially Crossed Random Effects . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

5.2 Data Description . . . . . . . . . . . . . . . . . . . . . . . . . . . 57

5.3 Multinomial Analysis of Injury Severity Data . . . . . . . . . . . . . 58

5.4 Analysis Results . . . . . . . . . . . . . . . . . . . . . . . . . . . 58

5.4.1 Minor Injury Subset . . . . . . . . . . . . . . . . . . . . . . 59

5.4.1.1 Interpretation of parameter estimates . . . . . . . . . 59

5.4.1.2 Interpretation of random effects . . . . . . . . . . . . 61

5.4.2 Serious Injury Subset . . . . . . . . . . . . . . . . . . . . . 64

5.4.2.1 Interpretation of parameter estimates . . . . . . . . . 64

5.4.2.2 Interpretation of random effects . . . . . . . . . . . . 66

5.4.3 Death Case Subset . . . . . . . . . . . . . . . . . . . . . . 70

5.4.4 Comparison analysis . . . . . . . . . . . . . . . . . . . . . . . . 71

**6 Discussion** **73**

6.1 Conclusion . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 73

6.2 Further Study . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74

**Bibliography** **80**

**Vita**

# List of Tables

2.1 The Penicillin dataset-fully crossed random effects . . . . . . . . . . . 7

2.2 Variable description of motor vehicle accident data . . . . . . . . . . 15

3.1 Simulation results for the proposed model parameter estimates . . . 39

4.1 Sample data of injury occurrence data . . . . . . . . . . . . . . . . . 42

4.2 Descriptive statistics for numerical variables . . . . . . . . . . . . . . 42

4.3 Descriptive statistics for categorical variables . . . . . . . . . . . . . . 43

4.4 Parameter estimation results based on GLM . . . . . . . . . . . . . . 46

4.5 Parameter estimation results based on the proposed binary model-full model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46

4.6 Parameter estimation results-final model . . . . . . . . . . . . . . . . 48

5.1 Sample data of injury severity data . . . . . . . . . . . . . . . . . . . 58

5.2 Estimation results based on the proposed baseline-category model (minor injuries)-full model . . . . . . . . . . . . . . . . . . . . . . . . . 59

5.3 Estimation results based on the proposed baseline-category model (minor injuries)-final model . . . . . . . . . . . . . . . . . . . . . . . . 60

5.4 Estimation results based on the proposed baseline-category model (serious injuries)-full model . . . . . . . . . . . . . . . . . . . . . . . . 65

5.5 Estimation results based on the proposed baseline-category model (serious injuries)-final model . . . . . . . . . . . . . . . . . . . . . . . 65

5.6 Estimation results based on the proposed baseline-category model (death cases) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 70
5.7 Comparison of the estimation results based on the proposed baseline-category models . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 71

# List of Figures


4.1 Average risk of injury vs Seats . . . . . 49
4.2 Average risk of injury vs Insured Value . . . . . 50
4.3 Predicted random effects of agent . . . . . 52
4.4 Predicted random effects of location . . . . . 53
4.5 Predicted random effects of location with spatial coordinates . . . . 54

5.1 Average risk of minor injuries vs Insured Value . . . . . 61
5.2 Predicted random effects of agent (minor injuries) . . . . . 62
5.3 Predicted random effects of location (minor injuries) . . . . . 62
5.4 Predicted random effects of location with spatial coordinates (minor injuries) . . . . . 63
5.5 Average risk of serious injuries vs Seats . . . . . 66
5.6 Average risk of serious injuries vs Insured Value . . . . . 67
5.7 Predicted random effects of agent (serious injuries) . . . . . 67
5.8 Predicted random effects of location (serious injuries) . . . . . 68
5.9 Predicted random effects of location with spatial coordinates (serious injuries) . . . . . 69

# Chapter 1

# Introduction

Data in educational and social sciences could have cross-classified structures, where lower-level units belong to combinations of two or more higher-level units, which are formed by crossing these higher-level units with one another (Leckie, 2019). For example, one widely used cross-classified dataset is on the examination scores of 16-year olds in Fife in Scotland (Browne et al., 2001). The data are cross-classified because not all children who attended the same primary school subsequently entered the same secondary school; similarly, not all children from the same secondary school graduated from the same primary school. Therefore, the secondary school is cross-classified by the primary school. Another typical example is in healthcare, where doctors and nurses could be cross-classified in the studies assessing the effectiveness of medical treatments.

Kim et al. (2021) discussed that many problems could arise due to the misspecification of crossed factors in the cross-classified data structures (Meyers and Beretvas, 2006; Luo and Kwok, 2009; Ye and Daniel, 2017). In the comparative study of hierarchical linear model (HLM) and cross-classified random effects model, Meyers and Beretvas (2006) showed that the standard error in HLM is biased when the parameters are generated under the cross-classified random effects model. Also, Ye and

Daniel (2017) suggested that ignoring one of the crossed factors could yield underestimated standard errors in level one regression coefficients. Therefore, it is essential to incorporate the crossed random effects into the model when datasets have cross-classified data structures; otherwise, the estimation results would be misleading. It is worth noting that cross-classified data structures are different from another type of typical data structures – nested or hierarchical data structures – in which one lower level is nested within the next higher level, which could be further nested within another higher level.

Generally, the cross-classified structure could be fully or partially crossed. When a dataset has at least one combination of each level of one crossed factor and each level of the other crossed factor, we say that the dataset has a fully or completely cross-classified data structure (Bates, 2010). In contrast, if not all these theoretical combinations occur in the dataset, these two crossed factors are partially crossed instead. The difference between fully and partially crossed random effects is further briefly illustrated in the next chapter. Both fully and partially cross-classified datasets are quite common in social science fields, the research of which, thus, is of great interest. The models with partially crossed random effects proposed by this thesis can work on both fully and partially cross-classified data structures.

The research of this thesis is mainly for binary data with cross-classified structures. Binary models in the literature considering crossed factors are mainly related to generalized linear mixed models (GLMMs). A GLMM is an extension of the generalized linear model (GLM), in which the linear predictors include the random effects in addition to the usual fixed effects. The random effects in GLMMs are assumed to be normally distributed. Although the normal distribution is convenient, other distributions might be preferred in some instances. Regarding datasets with binary outcomes, a GLMM, with random intercept and using the logit link, is the mixed-effects logistic regression model (Bakbergenuly and Kulinskaya, 2018). As for the crossed random-effects model with a binary response (refer to Eq. (2.4) in Chapter 2), the high dimensional integral in the marginal likelihood that incorporates crossed random effects is daunting and considered as a major difficulty in finding distribution parameters (Ghosh et al., 2021). Jeon et al. (2017) also mentioned that estimation is computationally highly intensive because of the intractable and high-dimensional integrals involved in the likelihood functions; therefore, in many cases, the integral can only be approximated numerically (Bologa et al., 2021). As a result, some likelihood-based approximation methods are used to deal with the computational challenges, such as Laplace approximation that approximates the integral based on second-order Taylor expansion (Lindstrom and Bates, 1988; Tierney and Kadane, 1986; Wolfinger, 1993), the penalized quasi-likelihood approach (PQL)(Breslow and Clayton, 1993), and Adaptive Gaussian-Hermite quadrature (Tuerlinckx et al., 2006; Rich, 2018). For example, the PQL approach was applied to one cross-classified dataset presented by McCullagh and Nelder (1989); it is on salamander mating from two populations called whiteside (W) and rough butt (R) with a binary response (whether the salamanders mate or not). The crossed factor female and male type includes four levels: R/R, R/W, W/R, and W/W, where W/R, for example, represents that a whiteside female is crossed with a rough butt male. The salamander mating binary dataset shows that the PQL has severe biases (Lee et al., 2017).

A Bayesian approach is used as an alternative to likelihood-based approximations in which Markov Chain Monte Carlo methods (MCMC) are adopted to make inferences on the posterior distribution of the parameters (Capanu et al., 2013). Zeger and Karim (1991) introduced the use of Gibbs sampling to fit GLMMs, which can also be applied to cross-classified datasets.

Beta-binomial models are a natural choice for binary data with a clustered structure, and they have certain advantages. In beta-binomial regression models, the probability parmeter is modelled with a beta distribution and can thus handle the extra-binomial variation. In addition, since beta distribution is conjugate for the binomial distribution, the marginal distribution could be expressed in explicit form. Therefore, the likelihood function for beta-binomial is in a relatively simple formula (Dohoo et al., 2003), and it is simpler to compute than GLMMs, which need to approximate the high-dimensional integral. On the other hand, Najera-Zuloaga et al. (2018) mentioned that beta-binomial model could be considered as a special case of hierarchical generalized linear models (HGLMs)(Lee and Nelder, 1996), in which the distribution of random effects is not necessarily normal. Based on our best knowledge, beta-binomial models cannot deal with crossed random effects with both crossed random effects assumed to come from beta distributions. Our proposed models could be considered as a generalization of beta-binomial models because our models could deal with crossed random effects where the conditional response follows a Bernoulli distribution and do not need to specify the distribution of random effects.

Specifically, the thesis proposes a binary logistic model with partially crossed random effects, which is further extended to a baseline-category logit model with partially crossed random effects for multinomial analysis. The thesis is built on work done by Ma (1999), and Ma and Jørgensen (2007). The random effects in our proposed models are predicted by the orthodox best linear unbiased predictor (BLUP) approach, which is the best and unbiased because it minimizes the mean square distance between the random effects and their predictor (Ma, 1999). Concretely, there are several advantages of our proposed models. First, as in Ma and Jørgensen (2007), our models are robust against the misspecification of random effects distributions (Ha and Lee, 2005) because our models only specify the first two moments of random effects. Second, in general, our models could be computationally efficient since they do not need to integrate over the random effects such as in GLMMs. Last, the marginal mean could be easily expressed since the crossed random effects are in the multiplicative form.

The remainder of the thesis is organized as follows. Chapter 2 first briefly reviews the basic concepts of random effects and the difference between partially and fully crossed random effects. It then introduces logistic models with crossed random effects and beta-binomial logistic regression, followed by description of motor vehicle accident data.

Chapter 3 includes model assumptions and the derivation of the moment structure of the response and random effects. In addition, it discusses the orthodox best linear unbiased predictor (BLUP) of random effects and estimation methods of regression and dispersion parameters. Finally, it presents details of the estimation algorithm.

In Chapter 4, we first apply our proposed binary logistic model to injury occurrence data with binary outcomes. Then, we interpret parameter estimates and predicted random effects. Finally, we perform a simulation study accessing the performance of the estimation algorithm.

Chapter 5 proposes a baseline-category logit model with partially crossed random effects for multinomial analysis. We apply this model to injury severity data, the results of which are interpreted.

Lastly, simple conclusions and brief further discussions are made in Chapter 6.

# Chapter 2

# Preliminaries

This chapter reviews some basic concepts of crossed random effects, logistic models with crossed random effects as well as beta-binomial logistic regression, followed by the brief description of the dataset used to study our proposed models – insurance data about motor vehicle accidents.

## 2.1 Crossed Random Effects

The thesis considers partially crossed factors in a dataset with binary outcomes, which are also called partially crossed random effects. Before we talk about crossed random effects, we first briefly review random effects. As in McCulloch and Searle (2001), the effects of a factor are either fixed or random in terms of parameters. When parameters are considered fixed constants, they are called fixed effects. In contrast, when the parameters are considered random, they are random effects, where only a random sample occurs in the data. The typical example used to illustrate the difference is the baking bread example: four loaves of bread are taken from six batches of bread baked at three different temperatures. Here the temperature effects are considered fixed, whereas the batch effects are random for the following reasons. First, the statistical interest or even the experiment design is to estimate those temperature effects, and it is also not sensible to assume that temperate effects are random since the temperature is defined on a continuum, e.g., the effect of a 450° temperature is almost always likely to be similar to the effect of a 449° temperature. In contrast, batches cannot be defined on a continuum and depend on different circumstances, just as clinics or agents, thus being random.

Generally, random effects could be fully or partially crossed. Here we illustrate their difference using the Penicillin data (Davies and Goldsmith, 1972). Using the B.subtilis method, six penicillin samples were tested on each of 24 Petri dishes of approximately 90 mm diameter, known as plates. The diameter (mm) of the zone of inhibition of the organisms' growth is considered as the response in this experiment.

Table 2.1: The Penicillin dataset-fully crossed random effects

| sample \ plate | a | b | c | d | e | f | g | h | i | j | k | l | m | n | o | p | q | r | s | t | u | v | w | x |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| B | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| C | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| D | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| E | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| F | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

There are three variables in the Penicillin dataset: one is numerical response variable diameter, and the other two are categorical variables, sample and plate. The random effects of sample and plate factors are considered here. Table 2.1 shows that sample has six different levels, A, B, ..., F, and plate has 24 levels, a, b, ..., x. We could describe that these two crossed factors are fully crossed, which means that we have at least one combination for each level of the plate and each level of the sample (Bates, 2010). Here there is one observation for each combination of sample and plate.

Suppose we replace several 1s with 0s in Table 2.1, indicating that certain combinations of sample and plate do not exist in the dataset, then we could describe the two factors plate and sample as being partially crossed. This could be immediately relevant for education research, where teachers and students have random effects associated with them. For example, some students have been attending a math tutoring class for several years. We want to study if this tutoring could improve their math performance. If we consider 100 students and 50 teachers as the crossed random effects, in that case, they cannot be fully crossed because it is unusual for each student to be observed with each teacher. In most cases, as the one indicated here, students and teachers are partially crossed random effects.

## 2.2 Logistic Models with Crossed Random Effects

The generalized linear mixed model (GLMM) is an extension to the generalized linear model (GLM) (e.g., logistic regression) in which linear predictors are allowed to include both the usual fixed effects and the desired random effects. Conditioned on random effects $\boldsymbol{u}$, the expectation of response $\mathbf{Y}$ is related to the linear predictor via a link function $g$. The linear predictor specifying the modeling for fixed and random effects is expressed as,

$$g(E(\mathbf{Y}|\boldsymbol{u}))= \boldsymbol{X\beta} + \boldsymbol{Zu}$$

where $\boldsymbol{u}$ is the set of random effects, $\boldsymbol{Z}$ is the design matrix for the random part of the model, and the random effects are normally distributed.

For the binary data where the response $Y_k$ can only take on values 0 and 1, the common link function is the logit link function,

$$g(p_k) = \log \frac{p_k}{1 - p_k}$$

where $p_k$ represents the probability of $Y_k$ taking on the value 1. A GLMM, with random intercept and using a logit link function, is a mixed effects logistic regression model (Bakbergenuly and Kulinskaya, 2018). A crossed random effects model with the binary response $Y_{ij} \in \{0,1\}$ has

$$\log\left(\frac{\Pr(Y_{ij}=1|a_i,b_j)}{\Pr(Y_{ij}=0|a_i,b_j)}\right) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j \quad 1 \le i \le R, 1 \le j \le C \tag{2.1}$$
$$a_i \sim N(0,\sigma_a^2) \quad b_j \sim N(0,\sigma_b^2)$$

for unobserved random effects $a_i$ and $b_j$, which are assumed to be independent of each other. $\boldsymbol{\beta}$ is a vector of fixed regression parameters, and $\boldsymbol{x}_{ij}$ is a vector of covariates. By algebraic transformation, we get

$$\Pr(Y_{ij}=1|\mathbf{a},\mathbf{b}) = \Pr(Y_{ij}=1|a_i,b_j) = \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)}. \tag{2.2}$$

As in Eq. (2.1), $\boldsymbol{a}$ and $\boldsymbol{b}$ are incorporated in the model as random intercepts, and the probability incorporating these random effects is expressed as in Eq. (2.2); therefore, the marginal mean expressed as follows cannot be expressed in a simple form.

$$\begin{aligned}
\mathrm{E}(Y_{ij}) &= \mathrm{E}(\mathrm{E}(Y_{ij}|\mathbf{a},\mathbf{b})) \\
&= \mathrm{E}\left(\frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)}\right) \\
&= \iint_{\mathbb{R}^2} \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+a_i+b_j)} \frac{1}{\sigma_a}\varphi\left(\frac{a_i}{\sigma_a}\right)\frac{1}{\sigma_b}\varphi\left(\frac{b_j}{\sigma_b}\right) \mathrm{d}a_i\ \mathrm{d}b_j,
\end{aligned}$$

where $\varphi(\cdot)$ is the $N(0,1)$ probability density function, and $\mathbf{a} \in \mathbb{R}^R$ and $\mathbf{b} \in \mathbb{R}^C$ are vectors with components $a_i$ and $b_j$. Unlike the expression here, our proposed models can express the marginal mean relatively simply (refer to Eq. (3.5)).

Based on Eq. (2.2), we obtain the conditional likelihood of $\boldsymbol{\beta}$ given $\mathbf{a}$ and $\mathbf{b}$,

$$
\begin{aligned}
L(\boldsymbol{\beta}|\mathbf{a},\mathbf{b}) &= \prod_{(i,j)} \left( \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)} \right)^{Y_{ij}} \left( 1 - \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)} \right)^{1-Y_{ij}} \\
&= \prod_{(i,j)} \frac{\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)^{Y_{ij}}}{1+\exp(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + a_i + b_j)}, \qquad (2.3)
\end{aligned}
$$

where the product is obtained by taking over pairs $(i, j)$ for which $(\boldsymbol{x}_{ij}, Y_{ij})$ is observed, and the marginal likelihood considering random effects is (Ghosh et al., 2021)

$$
L(\boldsymbol{\beta}, \sigma_a^2, \sigma_b^2) = \int_{\mathbb{R}^{R+C}} L(\boldsymbol{\beta}|\mathbf{a},\mathbf{b}) \prod_{i=1}^{R} \frac{1}{\sigma_a}\varphi\left(\frac{a_i}{\sigma_a}\right) \prod_{j=1}^{C} \frac{1}{\sigma_b}\varphi\left(\frac{b_j}{\sigma_b}\right) \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b}. \qquad (2.4)
$$

Ghosh et al. (2021) pointed it out that the high dimensional integral in Eq. (2.4) is daunting and regarded as the primary difficulty in finding estimates for $\boldsymbol{\beta}$, and Pawitan (2001) also mentioned that integrating over the random effects could be highly computationally intensive. Solutions for marginal likelihoods are typically available for a small class of distribution, in particular when the integrated-out parameter is the conjugate prior of the distribution of the data. Bologa et al. (2021) argued that the integral as in Eq. (2.4) can only be approximated numerically. Several likelihood-based approximation methods have been developed to handle the computational difficulty related to the integral, and one popular way is Laplace approximation which approximates an integral by expanding the logarithm of the integrand in Taylor series (Jeon et al., 2017). Another common approach is penalized quasi-likelihood approach (PQL), which uses a Laplace approximation to the integrated mixed model likelihood (Dean et al., 2004), giving an approximated likelihood function with normal distribution (Breslow and Clayton, 1993). However, PQL has been shown to give biased estimates especially when the distribution of the response variable is far from normal (Agresti, 2003; Rodríguez and Goldman, 1995). In most statistical computing environments, integration of GLMMs is approximated based on Adaptive Gauss-Hermite quadrature (Liu and Pierce, 1994; Pinheiro and Bates, 1995, 2000; Pinheiro and Chao, 2006). Bologa et al. (2021) claimed Adaptive Gauss-Hermite quadrature centers the location of the nodes by finding the random effects values that maximize the log-likelihood function and scales them based on the curvature of that function. Jeon et al. (2017) stated that despite the ongoing efforts, there are some limitations for these methods, and they cannot be widely applicable for various situations.

On the other hand, predictions of random effects are also important in the models. Based on Ma and Jørgensen (2007) and Bologa et al. (2021), we introduce the joint log-likelihood function for crossed random effects. Let $\mathbf{Y}$ be the response, and $\boldsymbol{a}$ and $\boldsymbol{b}$ be the unobserved crossed random effects. The joint log-likelihood could be written as

$$l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}; \mathbf{Y}, \mathbf{a}, \mathbf{b}) = l(\boldsymbol{\beta}, \boldsymbol{\alpha}; \mathbf{Y}|\mathbf{a}, \mathbf{b}) + l(\boldsymbol{\gamma}; \mathbf{a}) + l(\boldsymbol{\delta}; \mathbf{b}), \tag{2.5}$$

where $l(\boldsymbol{\beta}, \boldsymbol{\alpha}; \mathbf{Y}|\mathbf{a}, \mathbf{b})$ is the conditional log-density of $\mathbf{Y}$ given $\boldsymbol{a}$ and $\boldsymbol{b}$ with regression parameter $\boldsymbol{\beta}$ and dispersion parameter $\boldsymbol{\alpha}$, and $l(\boldsymbol{\gamma}; \mathbf{a})$ and $l(\boldsymbol{\delta}; \mathbf{b})$ are the log-densities for $\boldsymbol{a}$ and $\boldsymbol{b}$ with parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$, respectively. Here the crossed random effects are not necessarily normally distributed. For a normal theory linear mixed model with fixed $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, predictions of random effects for crossed factor $\boldsymbol{a}$, for example, can be obtained by solving the equation, $\partial l(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}; \mathbf{Y}, \mathbf{a}, \mathbf{b})/\partial \mathbf{a} = 0$, which indicates that the distribution of the random effects must be known beforehand. However, by adopting the orthodox best linear unbiased predictor (BLUP) approach, our models do not need to specify the distribution because our assumptions on random effects are based on the first two moments. In addition, our models can predict random effects explicitly.

## 2.3 Beta-binomial Logistic Regrsssion

Now we briefly review the basic concepts of beta-binomial logistic regression and beta-binomial HGLM. Assume that we have $Y_k$, a set of variables, $k = 1, \ldots, N$, conditioned on the probability parameter $p$, are independent and follow a Bernoulli distribution with parameter $p$. In addition, a logical choice for the distribution of the random variable $p$ is a beta distribution with parameters $\alpha_1$ and $\alpha_2$ since it is flexible on (0,1) and has mathematically tractable results (McCulloch and Searle, 2001). Summing up all the variables, we have $Y = \sum_{k=1}^{N} Y_k$ and have the beta-binomial model:

$$\mathbf{Y}|\boldsymbol{p} \sim \text{Bin}(n, \boldsymbol{p}) \quad \boldsymbol{p} \sim \text{Beta}(\alpha_1, \alpha_2), \tag{2.6}$$

where $n$ denotes the number of trials, and $\mathbf{Y}$ follows a beta-binomial distribution with

$$\mu = \frac{n\alpha_1}{\alpha_1 + \alpha_2}, \quad \sigma^2 = \frac{n\alpha_1\alpha_2(\alpha_1 + \alpha_2 + n)}{(\alpha_1 + \alpha_2)^2(\alpha_1 + \alpha_2 + 1)}.$$

Najera-Zuloaga et al. (2018) pointed it out that by applying certain reparameterization (for details refer to (Najera-Zuloaga et al., 2018)) of the beta-binomial distribution parameters, $p_k$ could be considered as the probability of success in each Bernoulli observation, allowing to link it with some given covariates by some link functions. Many functions can be used for binary data, and logit is the most common one. Forcina and Franconi (1988) assumed that the probability parameter $p_k$ is connected to regression parameters by a logit link model,

$$\text{logit}(p_k) = \log\left(\frac{p_k}{1 - p_k}\right) = \boldsymbol{x}_k^T\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a vector of fixed regression parameters, and $\boldsymbol{x}_k^T$, $k = 1, \ldots, n$, is a vector of covariates. Applying maximum likelihood approach, we can obtain the maximum likelihood estimators of the regression as well as the dispersion parameters by the iteration method.

Lee and Nelder (1996) extended GLMMs to hierarchical generalized linear models (HGLMs), where the random effects may come from a conjugate exponential-family distribution. For HGLMs, the expectation of response $\mathbf{Y}$ is related to the linear predictor via a link function $g$, and the linear predictor is written as,

$$g(E(\mathbf{Y}|\boldsymbol{u})) = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{v},$$

where $\boldsymbol{v} = \boldsymbol{v}(\boldsymbol{u})$, the scale, is a monotone function of $\boldsymbol{u}$, and $\boldsymbol{Z}$ is the design matrix for the random part of the model. We know that the beta-binomial models have conditional response following a binomial distribution as well as some given random effects following beta distributions; therefore, the beta-binomial model can be treated as a special case of the HGLMs with $Y_k|u_k \sim \text{Bin}(n, p_k)$ and $u_k \sim \text{Beta}(\alpha_1, \alpha_2)$, where $p_k$ is connected to $u_k$ by the linear predictor of the beta-binomial HGLM expressed as follows,

$$\text{logit}(p_k) = \boldsymbol{x}_k^T\boldsymbol{\beta} + v_k, \tag{2.7}$$

where $v_k = \text{logit}(u_k)$ is the random effect attributed to individual $k$, $k = 1, \ldots, n$ (Najera-Zuloaga et al., 2018).

From Eq. (2.7), we can tell that it is not easy to extend it to crossed random effects. As far as we know, beta-binomial models cannot incorporate crossed random effects with both crossed factors following beta distributions. Different from this model, our proposed model is a generalization of the beta-binomial model, in which the conditional response follows a Bernoulli distribution and beta distributed random effects are a special case of the model framework which only specifies the first two moments of random effects.

## 2.4 Motor Vehicle Accident Data

### 2.4.1 Data Description

We iillustrate our proposed models on analyzing a confidential insurance dataset about motor vehicle accidents. It consists of 80 variables, for example, driver ID, insurance institution as well as its branch, compensation number, claim number, insurance product, accident description, agent, location, number of seats in a vehicle, Insured Value, good driver or not, accident date, registration date, injured or not, injury severity, and many other after-accident variables. After removing those after-accident variables, ID, agent and location related variables, and observations with missing data, we have a dataset with 12,195 observations. Variable descriptions are summarized in Table 2.2, and more details about covariates are explained in Chapter 4. At the bottom of the table, Injury is the binary response, and Severity is the multi-category response. For model studies, when the dataset is with response Injury, it is called injury occurrence dataset, and when it is with response Severity, it is called injury severity dataset. The two datasets are used in the analyses of the binary logistic model with partially crossed random effects in Chapter 4 and baseline-category logit model with partially crossed random effects for multinomial analysis in Chapter 5, respectively.

There are 71 agents defined here who represent those insurance companies or individual persons who sell motor vehicle insurance products. In the data shown in Table 2.2 , motor vehicle accidents are associated with three types of agency: individual persons, employees from insurance companies, and insurance companies. These agency belong to three different tertiary institutions, A, B, and C. We first combined those employees from the same insurance company into one agent. Second, for both individual persons and insurance companies who are associated with less than 20 accidents (frequency less than 20), we combined them into their tertiary institutions as one agent unitfor example, Person A represents an individual person who belongs to tertiary institution A, and Company B represents one insurance company that belongs to tertiary institution B. In this way, adding up the insurance companies and individuals whose frequencies are equal to or higher than 20, we have 71 agents in total.

There are 127 locations that are unique places where motor accidents occurred, as shown in the data.

Table 2.2: Variable description of motor vehicle accident data

| **Variable** | **Description** | **Type** |
|---|---|---|
| Agent | ID number for each agent (71 in total) | Random effect $U_i$ |
| Location | ID number for each location (127 in total) | Random effect $V_j$ |
| Insurance Product* | Different insurance products | Categorical |
| Purpose | Commercial or non-commercial vehicles | Categorical |
| Good Driver | Good driver yes/no | Categorical |
| Insured Value | The value of the insurance product | Numerical |
| Seats | The number of seats in a vehicle | Numerical |
| Vehicle's Age | The age of the vehicle at the time of the accident | Numerical |
| Accident Date | The date of the accident | Datetime |
| Registration Date | Initial date of insurance registration | Datetime |
| Injury | The occurrence of injuries, yes/no | Binary response |
| Severity | Injuries with 4 severity categories: Death, Serious injuries, Minor injuries and No injuries | Multi-category Response |

* Insurance Product includes six levels: Compulsory Motorcycle Insurance, Compulsory Motor Vehicle Insurance, Commercial Motor Vehicle Insurance, Comprehensive Motor Vehicle Insurance, Special Commercial Vehicle Insurance, and Telemarketing Commercial Insurance.

### 2.4.2 Crossed Random Effects

The data is cross-classified by agent and location because not all motor vehicles with the insurance bought from the same agent are involved in vehicle accidents in the same location; on the other hand, vehicle accidents in some locations are associated with more than one agents.

As for agents, it seems that they are only professionals who sell insurance products to their clients and are not related to the injury occurrence/severity of vehicle accidents. However, many professional agents have their own standards for selecting clients. For example, some professional agents might do strict quality reviews before selling insurance products while others do not. In this sense, the random effects of agents cannot be ignored, and those hidden factors could unknowingly influence the results of vehicle accidents. As for locations, it is known that there are certain places where car crashes occur on a more regular basis; in addition, some car crashes are only minor, while some are with bad injuries involved. Therefore, the random effects of locations should be considered too. As a result, Agent and Location are selected as the crossed random effects for our models.

# Chapter 3

# Binary Logistic Models with Partially Crossed Random Effects

In this chapter, we present our binary logistic model with partially crossed random effects. Specifically, we first propose three model assumptions and then investigate the moment structure of the model. Next, we introduce the orthodox best linear unbiased predictor (BLUP) of random effects and discuss the estimation for regression and dispersion parameters. Finally, we present the details of computational procedure and simulation study.

## 3.1 Model Specification

In this section, we consider the binary logistic model with partially crossed random effects, where each response is associated with the partially crossed random effects. Let $\mathbf{Y} = (Y_1, ..., Y_k, ..., Y_N)^T$ be a response vector, with $Y_k$ denoting the $k$th response and $N$ denoting the total number of observations in a dataset. In addition, let $\mathbf{U} = (U_1, ..., U_i, ..., U_I)^T$ represent the vector of random effects for one crossed factor, with $I$ components in total, and $\mathbf{V} = (V_1, ..., V_j, ..., V_J)^T$ represent the vector of random effects for the other crossed factor, with $J$ components in total. Here $i$ and

$j$ are determined for each $k$; in other words, given the $k$th observation, it must have the property with specific combination of $U_i$ and $V_j$. To avoid potential confusion in this thesis, letters $i$ and $I$ are only used for $\mathbf{U}$ related concepts, while $j$ and $J$ are only used for $\mathbf{V}$ related concepts. We characterize our model by the following three assumptions.

### 3.1.1 Assumption 1

The random effects $U_1, \cdots, U_i, \cdots, U_I$ are positive, independently and identically distributed on $(0, 1)$ with

$$\mathrm{E}(U_i) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(U_i) = \sigma^2.$$

### 3.1.2 Assumption 2

The random effects $V_1, ..., V_j, ..., V_J$ are positive, independently and identically distributed on $(0, 1)$ with

$$\mathrm{E}(V_j) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(V_j) = \tau^2.$$

In addition, let $\mathbf{W}$ represent the multiplicative combination of random effects $\mathbf{U}$ and $\mathbf{V}$, which is a vector with $I \times J$ components

$$\mathbf{W} = (U_1V_1, ..., U_1V_j, ..., U_1V_J, ..., U_iV_j, ..., U_IV_1, ..., U_IV_j, ..., U_IV_J)^T, \tag{3.1}$$

based on which we create an indication matrix $\mathbf{B}$ of size $N \times (I \times J)$ as

$$\mathbf{B} = (\mathbf{B}_1, ..., \mathbf{B}_k, ..., \mathbf{B}_N)^T,$$

where row $\mathbf{B}_k$ indicates the multiplicative combination of $U_iV_j$ related to the $k$th observation (Cui, 2019). $\mathbf{B}_k$ only has a value 1 at the position corresponding to $U_iV_j$, and 0 otherwise, illustrated as follows in the matrix,

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \end{pmatrix}_{N\times(I\times J)}.$$

**Example.** Let $I = 2$, $J = 3$, and $N = 4$. We have

$$\mathbf{W} = (U_1V_1, U_1V_2, U_1V_3, U_2V_1, U_2V_2, U_2V_3)^T, \tag{3.2}$$

which has 6 ($2 \times 3$) components. If the multiplicative combinations of the random effects for the 4 observations are $U_1V_3, U_2V_3, U_1V_2$, and $U_1V_3$ respectively, then we have a $4 \times (2 \times 3)$ matrix $\mathbf{B}$ as

$$\mathbf{B} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}_{4\times(2\times3)},$$

according to $\mathbf{W}$ in Eq. (3.2).

### 3.1.3 Assumption 3

$\mathbf{U} = (U_1, ..., U_i, ..., U_I)$ and $\mathbf{V} = (V_1, ..., V_j, ..., V_J)$ are independent with conditional distribution given below:

$$Y_k|\mathbf{U}, \mathbf{V} \sim \text{Bernoulli}(\pi_k U_i V_j),$$

where $\log(\pi_k/(1-\pi_k)) = \text{logit}(\pi_k) = \boldsymbol{x}_k^T\boldsymbol{\beta}$, and $\pi_k = \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))$; here $i$ and $j$ are determined for each $k$. That is, $i = i(k)$ and $j = j(k)$. $\mathbf{Y} = (Y_1, \ldots, Y_k, \ldots, Y_N)^T$ is a response vector, with $Y_k$ denoting the $k$th response. $\boldsymbol{\beta}$ is a vector of regression parameters.

The conditional expectation and the conditional variance of the response can be expressed as

$$\text{E}(Y_k|\mathbf{U},\mathbf{V}) = \pi_k U_i V_j, \tag{3.3}$$

$$\text{Var}(Y_k|\mathbf{U},\mathbf{V}) = \pi_k U_i V_j(1-\pi_k U_i V_j) = \pi_k U_i V_j - \pi_k^2 U_i^2 V_j^2. \tag{3.4}$$

## 3.2 Covariance Structure

This section investigates the moment structure of the model based on which we derive regression parameter estimators and random effects predictors in later sections.

### 3.2.1 Derivation

Our assumptions rely only on the first and second moments of random effects, and the derivations of moment structures apply the conditioning technique (Ma and Jørgensen, 2007), illustrated as follows.

According to Assumption 1 and Assumption 2, $\text{E}(U_i) = \text{E}(V_j) = \frac{\sqrt{2}}{2}$, and $U_i$ and $V_j$ are independent, so we have $\text{E}(U_iV_j) = \text{E}(U_i)\text{E}(V_j) = \frac{\sqrt{2}}{2} \times \frac{\sqrt{2}}{2} = \frac{1}{2}$. Thus, the unconditional expectation of $Y_k$ can be calculated by the law of total expectation,

$$\begin{aligned}\text{E}(Y_k) &= \text{E}(\text{E}(Y_k|\mathbf{U},\mathbf{V}))\\ &= \text{E}(\pi_k U_i V_j)\\ &= \pi_k \text{E}(U_i V_j)\\ &= \frac{1}{2}\pi_k.\end{aligned} \tag{3.5}$$

The unconditional covariance of $Y_k$ and $Y_{k'}$ can be calculated by the law of total covariance,

$$\begin{aligned}\mathrm{Cov}(Y_k, Y_{k'}) &= \mathrm{E}(\mathrm{Cov}(Y_k, Y_{k'})|\mathbf{U}, \mathbf{V}) + \mathrm{Cov}(\mathrm{E}(Y_k|\mathbf{U}, \mathbf{V}), \mathrm{E}(Y_{k'}|\mathbf{U}, \mathbf{V})) \\ &= \mathrm{term1} + \mathrm{term2}, \end{aligned} \tag{3.6}$$

where term1 and term2 are considered with two different cases and four different cases, respectively. Here we discuss cases for term1 first, where $k$ denotes the $k$th observation.

**Case 1:** if $k = k'$, then

$$\begin{aligned}\mathrm{E}(\mathrm{Cov}(Y_k, Y_{k'})|\mathbf{U}, \mathbf{V}) &= \mathrm{E}(\mathrm{Var}(Y_k|\mathbf{U}, \mathbf{V})) \\ &= \mathrm{E}(\pi_k U_i V_j - \pi_k^2 U_i^2 V_j^2) \\ &= \mathrm{E}(\pi_k U_i V_j) - \mathrm{E}(\pi_k^2 U_i^2 V_j^2) \\ &= \frac{1}{2}\pi_k - \pi_k^2 \mathrm{E}(U_i^2)\mathrm{E}(V_j^2) \\ &= \frac{1}{2}\pi_k - \pi_k^2 (\mathrm{Var}(U_i) + (\mathrm{E}(U_i))^2)(\mathrm{Var}(V_j) + (\mathrm{E}(V_j))^2) \\ &= \frac{1}{2}\pi_k - \pi_k^2 (\sigma^2 + \frac{1}{2})(\tau^2 + \frac{1}{2}). \end{aligned} \tag{3.7}$$

**Case 2:** if $k \neq k'$, then

We have $\mathrm{Cov}(Y_k, Y_{k'}|\mathbf{U}, \mathbf{V}) = 0$, so $\mathrm{E}(\mathrm{Cov}(Y_k, Y_{k'}|\mathbf{U}, \mathbf{V})) = 0$.

In summary, term1 in (3.6) can be expressed as follows,

$$\begin{aligned}\mathrm{term1} &= \mathrm{E}(\mathrm{Cov}(Y_k, Y_{k'})|\mathbf{U}, \mathbf{V}) \\ &= \begin{cases} \frac{1}{2}\pi_k - \pi_k^2(\sigma^2 + \frac{1}{2})(\tau^2 + \frac{1}{2}), & \text{if } k = k'; \\ 0, & \text{if } k \neq k'. \end{cases} \end{aligned} \tag{3.8}$$

Next, different cases for term2 $= \mathrm{Cov}(\mathrm{E}(Y_k|\mathbf{U}, \mathbf{V}), \mathrm{E}(Y_{k'}|\mathbf{U}, \mathbf{V}))$ in Eq. (3.6) are discussed as follows. As shown in Eq. (3.3), $\mathrm{E}(Y_k|\mathbf{U},\mathbf{V}) = \pi_k U_i V_j$, where $k$ denotes the $k$th observation that corresponds to the specific multiplicative combination of $U_i V_j$; therefore, we have

$$\begin{aligned}\mathrm{Cov}(\mathrm{E}(Y_k|\mathbf{U},\mathbf{V}),\mathrm{E}(Y_{k'}|\mathbf{U},\mathbf{V})) &= \mathrm{Cov}(\pi_k U_i V_j, \pi_{k'} U_{i'} V_{j'}) \\ &= \pi_k \pi_{k'} \mathrm{Cov}(U_i V_j, U_{i'} V_{j'}) \\ &= \pi_k \pi_{k'} (\mathrm{E}(U_i V_j U_{i'} V_{j'}) - \mathrm{E}(U_i V_j)\mathrm{E}(U_{i'} V_{j'})), \end{aligned} \quad (3.9)$$

where $\pi_k$ and $\pi'_k$ are constant; thus, we only need to investigate four different cases for $\mathrm{Cov}(U_i V_j, U_{i'} V_{j'})$, which are expressed as follows:

**Case 1:** if $i = i', j = j'$, then

$$\begin{aligned}\mathrm{Cov}(U_i V_j, U_{i'} V_{j'}) &= \mathrm{E}(U_i V_j U_{i'} V_{j'}) - \mathrm{E}(U_i V_j)\mathrm{E}(U_{i'} V_{j'}) \\ &= \mathrm{E}(U_i^2 V_j^2) - (\mathrm{E}(U_i V_j))^2 \\ &= \mathrm{E}(U_i^2)\mathrm{E}(V_j^2) - (\mathrm{E}(U_i V_j))^2 \\ &= (\mathrm{Var}(U_i) + (\mathrm{E}(U_i))^2)(\mathrm{Var}(V_j) + (\mathrm{E}(V_j))^2) - \frac{1}{2} \times \frac{1}{2} \\ &= (\sigma^2 + \frac{1}{2})(\tau^2 + \frac{1}{2}) - \frac{1}{4} \\ &= \sigma^2 \tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2. \end{aligned} \quad (3.10)$$

**Case 2:** if $i = i', j \neq j'$, then

$$\begin{aligned}\mathrm{Cov}(U_i V_j, U_{i'} V_{j'}) &= \mathrm{E}(U_i V_j U_{i'} V_{j'}) - \mathrm{E}(U_i V_j)\mathrm{E}(U_{i'} V_{j'}) \\ &= \mathrm{E}(U_i^2 V_j V_{j'}) - \mathrm{E}(U_i)\mathrm{E}(V_j)\mathrm{E}(U_i)\mathrm{E}(V_{j'}) \\ &= \mathrm{E}(U_i^2)\mathrm{E}(V_j)\mathrm{E}(V_{j'}) - (\mathrm{E}(U_i))^2\mathrm{E}(V_j)\mathrm{E}(V_{j'}) \\ &= \frac{1}{2}\mathrm{E}(U_i^2) - \frac{1}{2}(\mathrm{E}(U_i))^2 \\ &= \frac{1}{2}\sigma^2. \end{aligned} \quad (3.11)$$

**Case 3:** if $i \neq i', j = j'$, then

$$\begin{aligned}
\mathrm{Cov}(U_iV_j, U_{i'}V_{j'}) &= \mathrm{E}(U_iV_jU_{i'}V_{j'}) - \mathrm{E}(U_iV_j)\mathrm{E}(U_{i'}V_{j'}) \\
&= \mathrm{E}(U_iU_{i'}V_j^2) - \mathrm{E}(U_i)\mathrm{E}(V_j)\mathrm{E}(U_{i'})\mathrm{E}(V_{j'}) \\
&= \mathrm{E}(V_j^2)\mathrm{E}(U_i)\mathrm{E}(U_{i'}) - \mathrm{E}(U_i)\mathrm{E}(U_{i'})(\mathrm{E}(V_j))^2 \\
&= \frac{1}{2}\mathrm{E}(V_j^2) - \frac{1}{2}(\mathrm{E}(V_j))^2 \\
&= \frac{1}{2}\tau^2. \qquad (3.12)
\end{aligned}$$

**Case 4:** if $i \neq i', j \neq j'$, then

$$\begin{aligned}
\mathrm{Cov}(U_iV_j, U_{i'}V_{j'}) &= \mathrm{E}(U_iV_jU_{i'}V_{j'}) - \mathrm{E}(U_iV_j)\mathrm{E}(U_{i'}V_{j'}) \\
&= \mathrm{E}(U_i)\mathrm{E}(V_j)\mathrm{E}(U_{i'})\mathrm{E}(V_{j'}) - \mathrm{E}(U_i)\mathrm{E}(V_j)\mathrm{E}(U_{i'})\mathrm{E}(V_{j'}) \\
&= 0. \qquad (3.13)
\end{aligned}$$

In summary, term2 in (3.6) can be expressed as follows,

$$\begin{aligned}
\text{term2} &= \mathrm{Cov}(\mathrm{E}(Y_k|\mathbf{U},\mathbf{V}), \mathrm{E}(Y_{k'}|\mathbf{U},\mathbf{V})) \\
&= \pi_k\pi_{k'}\mathrm{Cov}(U_iV_j, U_{i'}V_{j'}) \\
&= \begin{cases}
\pi_k\pi_{k'}(\sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2), & \text{if } i = i' \text{ and } j = j'; \\
\pi_k\pi_{k'}\frac{1}{2}\sigma^2, & \text{if } i = i' \text{ and } j \neq j'; \\
\pi_k\pi_{k'}\frac{1}{2}\tau^2, & \text{if } i \neq i' \text{ and } j = j'; \\
0, & \text{if } i \neq i' \text{ and } j \neq j'.
\end{cases} \qquad (3.14)
\end{aligned}$$

For simplicity, if we apply the Kronecker notation, the covariance structure (3.6) can be expressed as follows,

$$
\begin{aligned}
& \operatorname{Cov}(Y_k, Y_{k'}) \\
= {} & \text{term1} + \text{term2} \\
= {} & \delta(k,k') \left( \frac{\pi_k - 2\pi_k^2(\sigma^2 + \frac{1}{2})(\tau^2 + \frac{1}{2})}{2} \right) \\
& + \pi_k \pi_{k'} \left( \delta(i,i')\delta(j,j')\sigma^2\tau^2 + \delta(i,i')\frac{1}{2}\sigma^2 + \delta(j,j')\frac{1}{2}\tau^2 \right),
\end{aligned} \tag{3.15}
$$

where $\delta(k,k') = 1$ if $k = k'$, 0 otherwise; $\delta(i,i') = 1$ if $i = i'$, 0 otherwise; $\delta(j,j') = 1$ if $j = j'$, 0 otherwise. In addition, this Kronecker notation can also be a way to express the covariance structure in coding after certain transformations.

### 3.2.2 Matrix Expression

Now we express the covariance structure of Eq. (3.15) in matrix form, which could ease the derivation of random effects predictors in later sections and is also necessary for the coding.

#### 3.2.2.1 Var(W) in matrix form

Since the multiplicative combinations of random effects **U** and **V** are represented by the **W** vector in Eq. (3.1), the covariance structure $\operatorname{Cov}(U_iV_j, U_{i'}V_{j'})$ shown in Eq. (3.14), is basically the covariance of **W**, which is an $(I \times J) \times (I \times J)$ squared matrix denoted as $\operatorname{Var}(\mathbf{W})$. Here we partition $\operatorname{Var}(\mathbf{W})$ into $I \times I$ blocks of the same size $J \times J$, where main-diagonal blocks are square matrices **A** and all off-diagonal blocks are squared matrices $\mathbf{M}$, as shown in Eq. (3.16).

$$\text{Var}(\mathbf{W}) = \begin{pmatrix} \mathbf{A} & \mathbf{M} & \cdots & \mathbf{M} & \mathbf{M} \\ \mathbf{M} & \mathbf{A} & \cdots & \mathbf{M} & \mathbf{M} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{M} & \mathbf{M} & \cdots & \mathbf{A} & \mathbf{M} \\ \mathbf{M} & \mathbf{M} & \cdots & \mathbf{M} & \mathbf{A} \end{pmatrix}_{(I\times J)\times(I\times J)}, \tag{3.16}$$

where $\mathbf{A}$, as shown in Eq. (3.17), is a $J \times J$ squared matrix in which all diagonal entries are $\sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2$ and all off-diagonal entries are $\frac{1}{2}\sigma^2$,

$$\mathbf{A} = \begin{pmatrix} \sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2 & \frac{1}{2}\sigma^2 & \cdots & \frac{1}{2}\sigma^2 & \frac{1}{2}\sigma^2 \\ \frac{1}{2}\sigma^2 & \sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2 & \cdots & \frac{1}{2}\sigma^2 & \frac{1}{2}\sigma^2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1}{2}\sigma^2 & \frac{1}{2}\sigma^2 & \cdots & \sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2 & \frac{1}{2}\sigma^2 \\ \frac{1}{2}\sigma^2 & \frac{1}{2}\sigma^2 & \cdots & \frac{1}{2}\sigma^2 & \sigma^2\tau^2 + \frac{1}{2}\sigma^2 + \frac{1}{2}\tau^2 \end{pmatrix}_{J\times J}, \tag{3.17}$$

whereas $\mathbf{M}$, as shown in Eq. (3.18), is a $J \times J$ squared matrix in which all diagonal entries are $\frac{1}{2}\tau^2$, and all off-diagonal entries are 0,

$$\mathbf{M} = \begin{pmatrix} \frac{1}{2}\tau^2 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{2}\tau^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \frac{1}{2}\tau^2 & 0 \\ 0 & 0 & \cdots & 0 & \frac{1}{2}\tau^2 \end{pmatrix}_{J\times J}. \tag{3.18}$$

#### 3.2.2.2 Var$(\mathbf{Y})$ in matrix form

Now we express Var$(\mathbf{Y})$ in matrix form based on my derivation of $\text{Cov}(Y_k, Y_k')$. First, the Kronecker notation for term1 in Eq. (3.15) can be expressed as a diagonal matrix in Eq. (3.19)

$$\mathrm{diag}(\mathbf{d}) = \begin{pmatrix} d_1 & & & & \\ & \ddots & & & \\ & & d_k & & \\ & & & \ddots & \\ & & & & d_N \end{pmatrix}_{N\times N}, \tag{3.19}$$

where

$$d_k = \frac{\pi_k - 2\pi_k^2(\sigma^2+\frac{1}{2})(\tau^2+\frac{1}{2})}{2}, \qquad k = 1, ..., N$$

Next, for term2, we have

$$\mathrm{Cov}(\mathrm{E}(Y_k|\mathbf{U},\mathbf{V}), \mathrm{E}(Y_{k'}|\mathbf{U},\mathbf{V})) = \mathrm{Cov}(\pi_k U_i V_j, \pi_{k'} U_{i'} V_{j'}).$$

Then, we can derive term2 with $\mathbf{B}, \mathbf{W}$ included as follows,

$$\mathrm{Var}(\boldsymbol{\pi}\mathbf{U}\mathbf{V}) = \mathrm{Var}(\mathrm{diag}(\boldsymbol{\pi})\mathbf{B}\mathbf{W}) = \mathrm{diag}(\boldsymbol{\pi})\mathbf{B}\mathrm{Var}(\mathbf{W})\mathbf{B}^T\mathrm{diag}(\boldsymbol{\pi}), \tag{3.20}$$

where $\boldsymbol{\pi} = (\pi_1, ..., \pi_k, ..., \pi_N)^T$, and $\mathbf{B}$ is the indication matrix defined in the previous section. Therefore, adding up term1 expressed in Eq. (3.19) and term2 in Eq. (3.20), we can have the second moment structure of $\mathbf{Y}$, i.e., $\mathrm{Var}(\mathbf{Y})$,

$$\mathrm{Var}(\mathbf{Y}) = \mathrm{diag}(\mathbf{d}) + \mathrm{diag}(\boldsymbol{\pi})\mathbf{B}\mathrm{Var}(\mathbf{W})\mathbf{B}^T\mathrm{diag}(\boldsymbol{\pi}). \tag{3.21}$$

## 3.3 The Best Linear Unbiased Predictor of Random Effects

Applying the orthodox best linear unbiased predictor (BLUP) method, this section predictes the random effects $U_i$ and $V_j$ in the following two subsections.

### 3.3.1 The BLUP of Random Effect $U_i$

#### 3.3.1.1 Derivation

The orthodox BLUP of the random effects $\mathbf{U}$, given the response $\mathbf{Y}$, is defined as

$$\widehat{\mathbf{U}} = \mathrm{E}(\mathbf{U}) + \mathrm{Cov}(\mathbf{U}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})(\mathbf{Y} - \mathrm{E}(\mathbf{Y})). \tag{3.22}$$

In addition, the variance of $\mathbf{U}$ can be calculated with the formula

$$\mathrm{Var}(\widehat{\mathbf{U}}) = \mathrm{Cov}(\mathbf{U}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})\mathrm{Cov}(\mathbf{Y}, \mathbf{U}), \tag{3.23}$$

where $\mathrm{E}(\mathbf{U}) = \left(\frac{\sqrt{2}}{2}, ..., \frac{\sqrt{2}}{2}, ..., \frac{\sqrt{2}}{2}\right)^T$ is an $I \times 1$ vector, and $\mathrm{Var}^{-1}(\mathbf{Y})$ is the inverse of $\mathrm{Var}(\mathbf{Y})$. $\mathrm{Cov}(\mathbf{Y}, \mathbf{U})$ in Eq. (3.23) is the transpose of $\mathrm{Cov}(\mathbf{U}, \mathbf{Y})$. $\mathrm{E}(\mathbf{Y})$ is an $N \times 1$ vector of unconditional expectation of response $\mathbf{Y}$. We have $\boldsymbol{\pi} = (\pi_1, ..., \pi_k, ..., \pi_N)^T$, by Eq. (3.5), so we have

$$\mathrm{E}(\mathbf{Y}) = \frac{1}{2}\boldsymbol{\pi}. \tag{3.24}$$

By the law of total covariance, we have

$$\mathrm{Cov}(\mathbf{U}, \mathbf{Y}) = \mathrm{E}(\mathrm{Cov}(\mathbf{U}, \mathbf{Y}|\mathbf{U}, \mathbf{V})) + \mathrm{Cov}(\mathrm{E}(\mathbf{U}|\mathbf{U}, \mathbf{V}), \mathrm{E}(\mathbf{Y}|\mathbf{U}, \mathbf{V})). \tag{3.25}$$

Since $\mathbf{U}$ is a constant vector when $\mathbf{U}$ and $\mathbf{V}$ are given, we have

$$\text{Cov}(\mathbf{U}, \mathbf{Y}|\mathbf{U}, \mathbf{V}) = 0 \tag{3.26}$$

$$\text{E}(\mathbf{U}|\mathbf{U}, \mathbf{V}) = \mathbf{U} \tag{3.27}$$

$$\text{E}(\mathbf{Y}|\mathbf{U}, \mathbf{V}) = \boldsymbol{\pi}\mathbf{U}\mathbf{V} = \text{diag}(\boldsymbol{\pi})\mathbf{B}\mathbf{W}. \tag{3.28}$$

Therefore,

$$\begin{aligned}\text{Cov}(\mathbf{U}, \mathbf{Y}) &= 0 + \text{Cov}(\mathbf{U}, \text{diag}(\boldsymbol{\pi})\mathbf{B}\mathbf{W}) \\ &= \text{Cov}(\mathbf{U}, \mathbf{W})\mathbf{B}^T\text{diag}(\boldsymbol{\pi}). \end{aligned} \tag{3.29}$$

where $\mathbf{B}^T$ is the transpose of the indication matrix $\mathbf{B}$, and $\text{diag}(\boldsymbol{\pi})$ is a constant matrix that can be factored out. Now we derive the covariance of $\mathbf{U}$ and $\mathbf{W}$, taking two different cases into consideration,

$$\text{Cov}(\mathbf{U}, \mathbf{W}) = \text{Cov}(U_i, U_{i'}V_j). \tag{3.30}$$

**Case 1:** if $i = i'$, then

$$\begin{aligned}\text{Cov}(U_i, U_{i'}V_j) &= \text{E}(\text{Cov}(U_i, U_iV_j|\mathbf{U})) + \text{Cov}(\text{E}(U_i|\mathbf{U}), \text{E}(U_iV_j|\mathbf{U})) \\ &= \text{E}(0) + \text{Cov}(U_i, U_i\text{E}(V_j)) \\ &= \text{E}(V_j)\text{Var}(U_i) \\ &= \frac{\sqrt{2}}{2}\sigma^2. \end{aligned} \tag{3.31}$$

**Case 2:** if $i \neq i'$, then

$$\begin{aligned}
\mathrm{Cov}(U_i, U_{i'}V_j) &= \mathrm{E}(\mathrm{Cov}(U_i, U_{i'}V_j|\mathbf{U})) + \mathrm{Cov}(\mathrm{E}(U_i|\mathbf{U}), \mathrm{E}(U_{i'}V_j|\mathbf{U})) \\
&= \mathrm{E}(0) + \mathrm{Cov}(U_i, U_{i'}\mathrm{E}(V_j)) \\
&= \mathrm{E}(0) + \mathrm{E}(V_j)\mathrm{Cov}(U_i, U_{i'}) \\
&= 0 + 0 \\
&= 0. \qquad (3.32)
\end{aligned}$$

In summary,

$$\mathrm{Cov}(U_i, U_{i'}V_j) = \begin{cases} \frac{\sqrt{2}}{2}\sigma^2, & \text{if } i = i'; \\ 0, & \text{otherwise.} \end{cases} \qquad (3.33)$$

#### 3.3.1.2 Matrix expression

$\mathrm{Cov}(\mathbf{U}, \mathbf{W})$ can be expressed in a matrix of size $I \times (I \times J)$ as

$$\mathrm{Cov}(\mathbf{U}, \mathbf{W}) = \begin{pmatrix} \frac{\sqrt{2}}{2}\sigma^2 & \cdots & \frac{\sqrt{2}}{2}\sigma^2 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \frac{\sqrt{2}}{2}\sigma^2 & \cdots & \frac{\sqrt{2}}{2}\sigma^2 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & \frac{\sqrt{2}}{2}\sigma^2 & \cdots & \frac{\sqrt{2}}{2}\sigma^2 \end{pmatrix}_{I \times (I \times J)}. \qquad (3.34)$$

### 3.3.2 The BLUP of Random Effect $V_j$

#### 3.3.2.1 Derivation

In this section, the derivation for the orthodox BLUP of the random effects $\mathbf{V}$ is pretty the same as that for the random effects $\mathbf{U}$. The orthodox BLUP of $\mathbf{V}$, given the response $\mathbf{Y}$, is defined as

$$\widehat{\mathbf{V}} = \mathrm{E}(\mathbf{V}) + \mathrm{Cov}(\mathbf{V}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})(\mathbf{Y} - \mathrm{E}(\mathbf{Y})). \tag{3.35}$$

In addition, the variance of $\mathbf{V}$ can be calculated with the formula

$$\mathrm{Var}(\widehat{\mathbf{V}}) = \mathrm{Cov}(\mathbf{V}, \mathbf{Y})\mathrm{Var}^{-1}(\mathbf{Y})\mathrm{Cov}(\mathbf{Y}, \mathbf{V}), \tag{3.36}$$

where $\mathrm{E}(\mathbf{V}) = \left(\frac{\sqrt{2}}{2}, ..., \frac{\sqrt{2}}{2}, ..., \frac{\sqrt{2}}{2}\right)^T$ is a $J \times 1$ vector, and $\mathrm{Var}^{-1}(\mathbf{Y})$ is the inverse of $\mathrm{Var}(\mathbf{Y})$. $\mathrm{Cov}(\mathbf{Y}, \mathbf{V})$ in Eq. (3.36) is the transpose of $\mathrm{Cov}(\mathbf{V}, \mathbf{Y})$. $\mathrm{E}(\mathbf{Y}) = \frac{1}{2}\boldsymbol{\pi}$, which is obtained from previous section. Again, by the law of total covariance, we have

$$\mathrm{Cov}(\mathbf{V}, \mathbf{Y}) = \mathrm{E}(\mathrm{Cov}(\mathbf{V}, \mathbf{Y}|\mathbf{U}, \mathbf{V})) + \mathrm{Cov}(\mathrm{E}(\mathbf{V}|\mathbf{U}, \mathbf{V}), \mathrm{E}(\mathbf{Y}|\mathbf{U}, \mathbf{V})). \tag{3.37}$$

Since $\mathbf{V}$ is constant when $\mathbf{U}$ and $\mathbf{V}$ are given, we have

$$\begin{aligned} \mathrm{Cov}(\mathbf{V}, \mathbf{Y}|\mathbf{U}, \mathbf{V}) &= 0 & (3.38)\\ \mathrm{E}(\mathbf{V}|\mathbf{U}, \mathbf{V}) &= \mathbf{V} & (3.39)\\ \mathrm{E}(\mathbf{Y}|\mathbf{U}, \mathbf{V}) &= \boldsymbol{\pi}\mathbf{U}\mathbf{V} = \mathrm{diag}(\boldsymbol{\pi})\mathbf{B}\mathbf{W}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{Cov}(\mathbf{V}, \mathbf{Y}) &= \mathrm{E}(\mathrm{Cov}(\mathbf{V}, \mathbf{Y}|\mathbf{U}, \mathbf{V})) + \mathrm{Cov}(\mathrm{E}(\mathbf{V}|\mathbf{U}, \mathbf{V}), \mathrm{E}(\mathbf{Y}|\mathbf{U}, \mathbf{V}))\\ &= 0 + \mathrm{Cov}(\mathbf{V}, \mathbf{W})\mathbf{B}^T\mathrm{diag}(\boldsymbol{\pi}) & (3.40) \end{aligned}$$

Similarly, the covariance of $\mathbf{V}$ and $\mathbf{W}$ considers two different cases,

**Case 1:** if $j = j'$, then

$$\begin{aligned}
\mathrm{Cov}(V_j, U_iV_{j'}) &= \mathrm{Cov}(V_j, U_iV_j) \\
&= \mathrm{E}(\mathrm{Cov}(V_j, U_iV_j|\mathbf{V})) + \mathrm{Cov}(\mathrm{E}(V_j|\mathbf{V}), \mathrm{E}(U_iV_j|\mathbf{V})) \\
&= \mathrm{E}(0) + \mathrm{Cov}(V_j, V_j\mathrm{E}(U_i)) \\
&= \mathrm{E}(0) + \mathrm{E}(U_i)\mathrm{Var}(V_j) \\
&= \frac{\sqrt{2}}{2}\tau^2. \qquad (3.41)
\end{aligned}$$

**Case 2:** if $j \neq j'$, then

$$\begin{aligned}
&\mathrm{Cov}(V_j, U_iV_{j'}) \\
&= \mathrm{E}(\mathrm{Cov}(V_j, U_iV_{j'}|\mathbf{V})) + \mathrm{Cov}(\mathrm{E}(V_j|\mathbf{V}), \mathrm{E}(U_iV_{j'}|\mathbf{V})) \\
&= \mathrm{E}(0) + \mathrm{Cov}(V_j, V_{j'}\mathrm{E}(U_i)) \\
&= 0. \qquad (3.42)
\end{aligned}$$

In summary,

$$\mathrm{Cov}(V_j, U_iV_{j'}) = \begin{cases} \frac{\sqrt{2}}{2}\tau^2, & \text{if } j = j'; \\ 0, & \text{otherwise.} \end{cases} \qquad (3.43)$$

#### 3.3.2.2 Matrix expression

$\mathrm{Cov}(\mathbf{V}, \mathbf{W})$ is a matrix of size $J \times (I \times J)$, which can be expressed as

$$\mathrm{Cov}(\mathbf{V}, \mathbf{W}) = \begin{pmatrix} \frac{\sqrt{2}}{2}\tau^2 & 0 & \cdots & 0 & \cdots & \frac{\sqrt{2}}{2}\tau^2 & 0 & \cdots & 0 \\ 0 & \frac{\sqrt{2}}{2}\tau^2 & \cdots & 0 & \cdots & 0 & \frac{\sqrt{2}}{2}\tau^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sqrt{2}}{2}\tau^2 & \cdots & 0 & 0 & \cdots & \frac{\sqrt{2}}{2}\tau^2 \end{pmatrix}_{J\times(I\times J)}. \qquad (3.44)$$

## 3.4 Parameter Estimation

### 3.4.1 Estimation of Regression Parameters

Here we use quasi-likelihood estimating equation (quasi-score function), denoted as $\psi(\boldsymbol{\beta})$, to estimate the regression parameter $\boldsymbol{\beta}$. The quasi-likelihood estimating equation for the estimation is

$$\psi(\boldsymbol{\beta}) = \sum_{k=1}^{N} \left[\frac{\partial \mathrm{E}(Y_k)}{\partial \boldsymbol{\beta}}\right] \mathrm{Var}^{-1}(Y_k)(Y_k - \mathrm{E}(Y_k)), \tag{3.45}$$

where $\mathrm{E}(Y_k) = \frac{1}{2}\pi_k$ , $\pi_k = \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))$, so

$$\begin{aligned}
\frac{\partial \mathrm{E}(Y_k)}{\partial \boldsymbol{\beta}} &= \frac{1}{2}\frac{\partial \pi_k}{\partial \boldsymbol{\beta}} \\
&= \frac{1}{2}\frac{\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})\boldsymbol{x}_k^T(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})) - \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})\boldsymbol{x}_k^T}{(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))^2} \\
&= \frac{1}{2}\frac{\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})\boldsymbol{x}_k^T}{(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))^2} \\
&= \frac{1}{2}\pi_k(1-\pi_k)\boldsymbol{x}_k^T.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\psi(\boldsymbol{\beta}) &= \frac{1}{2}\sum_{k=1}^{N} \boldsymbol{x}_k^T \pi_k(1-\pi_k)\mathrm{Var}(Y_k)^{-1}(Y_k - \frac{1}{2}\pi_k) \\
&= \frac{1}{2}\mathbf{X}^T \mathrm{diag}(\boldsymbol{\pi}(1-\boldsymbol{\pi}))\mathrm{Var}^{-1}(\mathbf{Y})(\mathbf{Y} - \frac{1}{2}\boldsymbol{\pi}).
\end{aligned} \tag{3.46}$$

This estimating function $\psi(\boldsymbol{\beta})$ attains the minimal asymptotic covariance for the estimator of $\boldsymbol{\beta}$ (see Crowder (1986, 1987) and Ma (1999)). The roots of equation

$$\psi(\boldsymbol{\beta}) = 0 \tag{3.47}$$

are considered as the estimates of $\boldsymbol{\beta}$. As in Ma and Jørgensen (2007), the roots are consistent and asymptotically normal with asymptotic mean $\boldsymbol{\beta}$ and asymptotic variance $\mathrm{Var}(\boldsymbol{\beta})$, where $\mathrm{Var}(\boldsymbol{\beta}) = -\mathrm{S}^{-1}(\boldsymbol{\beta})$ and sensitivity matrix $\mathrm{S}(\boldsymbol{\beta}) = \mathrm{E}_{\boldsymbol{\beta}}\left(\frac{\partial\psi(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right)$. Here we solve the Eq. (3.47) by applying Newton scoring algorithm, which is a version of Newton algorithm but replaces the second derivatives (the derivatives of the estimating function $\psi(\boldsymbol{\beta})$) with its expectations $\mathrm{S}(\boldsymbol{\beta})$. The Newton scoring algorithm yields the iteration for updating $\boldsymbol{\beta}$

$$\boldsymbol{\beta}^* = \boldsymbol{\beta} - \mathrm{S}^{-1}(\boldsymbol{\beta})\psi(\boldsymbol{\beta}), \tag{3.48}$$

where the explicit expression of the sensitivity matrix is

$$\begin{aligned}\mathrm{S}(\boldsymbol{\beta}) &= -\frac{1}{4}\mathbf{X}^T\mathrm{diag}(\mathrm{E}(\mathbf{Y}))\mathrm{Var}^{-1}(\mathbf{Y})\mathrm{diag}(\mathrm{E}(\mathbf{Y}))\mathbf{X} \\ &= -\frac{1}{4}\mathbf{X}^T\mathrm{diag}(\boldsymbol{\pi}(1-\boldsymbol{\pi}))\mathrm{Var}^{-1}(\mathbf{Y})\mathrm{diag}(\boldsymbol{\pi}(1-\boldsymbol{\pi}))\mathbf{X}.\end{aligned} \tag{3.49}$$

### 3.4.2 Estimation of Dispersion Parameters

In the previous section, we estimate the regression parameters given the dispersion parameters known. In this section, we estimate the dispersion parameters, $\sigma^2$ and $\tau^2$, when they are unknown using the adjusted Pearson estimators (Jrgensen et al., 1996), i.e., the Pearson estimator adjusted by its bias correction. As in Ma (1999), the bias correction, here denoted as $c$, is the difference between the variance of the random effect and the variance of its corresponding random effect predictor, which can be expressed as follows,

$$c(i) = \mathrm{Var}(U_i) - \mathrm{Var}(\widehat{U}_i), \tag{3.50}$$

$$c(j) = \mathrm{Var}(V_j) - \mathrm{Var}(\widehat{V}_j). \tag{3.51}$$

We may thus estimate the disperson parameters by the following adjusted Pearson estimators:

$$
\begin{aligned}
\widehat{\sigma}^2 &= \frac{1}{I}\sum_{i=1}^{I}\left(\widehat{U}_i - \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{I}\sum_{i=1}^{I} c(i) \\
&= \frac{1}{I}\sum_{i=1}^{I}\left(\widehat{U}_i - \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{I}\sum_{i=1}^{I}\Big(\mathrm{Var}(U_i) - \mathrm{Var}(\widehat{U}_i)\Big)
\end{aligned} \tag{3.52}
$$

$$
\begin{aligned}
\widehat{\tau}^2 &= \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{V}_j - \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{J}\sum_{j=1}^{J} c(j) \\
&= \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{V}_j - \frac{\sqrt{2}}{2}\right)^2 + \frac{1}{J}\sum_{j=1}^{J}\Big(\mathrm{Var}(V_j) - \mathrm{Var}(\widehat{V}_j)\Big),
\end{aligned} \tag{3.53}
$$

where on the right side of Eq. (3.52), the first term corresponds to Pearson estimator, and the second term is its corresponding bias correction; the same applies to Eq. (3.53).

## 3.5 Computational Procedure

Algorithm 1 shows the details of the parameter estimation algorithm. It starts with specified initial values and iterates until it meets the convergence criterion and then returns the output. The details are discussed in the following sections.

### 3.5.1 Initialization

Though we could start with initial values without strict limitation in most cases, however, if we choose the initial values relatively near the estimates, the convergence may well be fast and may be more likely to happen in certain cases. Therefore, Algorithm 1 starts with the initial values specified as follows.

**Algorithm 1** Parameter Estimation Algorithm

**Input**: the dataset and indication Matrix $\mathbf{B}$
**Output**: $\widehat{\boldsymbol{\beta}}$, $SE(\widehat{\boldsymbol{\beta}})$, $\widehat{\sigma}^2$, and $\widehat{\tau}^2$

Initialize $\widehat{\mathbf{U}}_0$ by Eq. (3.54), $\widehat{\mathbf{V}}_0$ by Eq. (3.55) ▷ Initialization
  $\widehat{\boldsymbol{\beta}}_0$ by Eq. (3.56) and $\widehat{\boldsymbol{\pi}}_0$ by Eq. (3.57)
  $\widehat{\sigma}_0^2$ by Eq. (3.58) and $\widehat{\tau}_0^2$ by Eq. (3.59)
**While** *diff* $< \epsilon$ by Eq. (3.60) **do**, ▷ Iteration
  $\widehat{\boldsymbol{\beta}}_{old} \leftarrow \widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}_{old}^2 \leftarrow \widehat{\sigma}^2$, $\widehat{\tau}_{old}^2 \leftarrow \widehat{\tau}^2$ ▷ copy old values
  estimate $\mathrm{Var}(\mathbf{W})$ by Eq. (3.16) ▷ equations from derivation part
  estimate $\mathrm{Var}(\mathbf{Y})$ by Eq. (3.21)
  estimate $\psi(\boldsymbol{\beta})$ by Eq. (3.46)
  estimate $S(\boldsymbol{\beta})$ by Eq. (3.49)
  update $\widehat{\boldsymbol{\beta}}$ by Eq. (3.48)
  update $\widehat{\boldsymbol{\pi}}$ by Eq. (3.57)
  estimate $\mathrm{Cov}(\mathbf{U}, \mathbf{Y})$ by Eq. (3.29)
  update $\widehat{\mathbf{U}}$ by Eq. (3.22)
  estimate $\mathrm{Cov}(\mathbf{V}, \mathbf{Y})$ by Eq. (3.40)
  update $\widehat{\mathbf{V}}$ by Eq. (3.35)
  update $\widehat{\sigma}^2$ by Eq. (3.52)
  update $\widehat{\tau}^2$ by Eq. (3.53)
**end**
**return** $\widehat{\boldsymbol{\beta}}$, $SE(\widehat{\boldsymbol{\beta}})$, $\widehat{\sigma}^2$ , and $\widehat{\tau}^2$

#### 3.5.1.1 Initialize random effects

The initial values of random effect predictors can be calculated by

$$\widehat{U}_{i(0)} = \begin{cases} \frac{1}{N}\sum_{k=1}^{N} Y_k, & \text{if } O_{U_i} = 0 \text{ or } 1; \\ O_{U_i}, & \text{otherwise.} \end{cases} \tag{3.54}$$

$$\widehat{V}_{j(0)} = \begin{cases} \frac{1}{N}\sum_{k=1}^{N} Y_k, & \text{if } O_{V_j} = 0 \text{ or } 1; \\ O_{V_j}, & \text{otherwise.} \end{cases} \tag{3.55}$$

where $O_{U_i}$ is given as the average of $Y_k$ values that belong to $U_i$, and $O_{V_j}$ is given as the average of $Y_k$ values that belong to $V_j$.

#### 3.5.1.2 Initialize the regression parameter estimates $\widehat{\boldsymbol{\beta}}_{(0)}$ and the probabilty estimates $\widehat{\boldsymbol{\pi}}_0$

We first fit the data to the logistic regression from which we get the fitted values, called $fit$, and set $\boldsymbol{\pi} = 2 \times fit$. Then we can get the estimates $\widehat{\boldsymbol{\beta}}_{(0)}$ from the following logit equation

$$\text{logit}(\boldsymbol{\pi}) = \log\left(\frac{\boldsymbol{\pi}}{1-\boldsymbol{\pi}}\right) = \mathbf{X}^T\boldsymbol{\beta} \tag{3.56}$$

by fitting the linear regression model to the data with $\text{logit}(\boldsymbol{\pi})$ as the response. In this way, we get our initial values of regression parameters $\widehat{\boldsymbol{\beta}}_{(0)}$, based on which we can obtain the initial values of $\widehat{\boldsymbol{\pi}}_{(0)}$ from the following equation

$$\widehat{\boldsymbol{\pi}}_{(0)} = \frac{\exp(\mathbf{X}^T\widehat{\boldsymbol{\beta}}_{(0)})}{1+\exp(\mathbf{X}^T\widehat{\boldsymbol{\beta}}_{(0)})}. \tag{3.57}$$

#### 3.5.1.3 Initialize the dispersion parameter estimates

The initial values of the dispersion parameter estimates are given as

$$\widehat{\sigma}^2_{(0)} = \frac{1}{I}\sum_{i=1}^{I}\left(\widehat{U}_{i(0)} - \frac{\sqrt{2}}{2}\right)^2 \tag{3.58}$$

$$\widehat{\tau}^2_{(0)} = \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{V}_{j(0)} - \frac{\sqrt{2}}{2}\right)^2. \tag{3.59}$$

### 3.5.2 Iteration

In this iteration part, the algorithm iterates line 3 to line 15 many times until it converges. Specifically, it stops after an iteration in which the absolute change between the old estimates from the previous round and those from the current round is less than a specified value, denoted as *diff* in line 2 Algorithm 1. Mathematically, *diff* can be expressed as

$$diff = |\widehat{\sigma}^2_{old} - \widehat{\sigma}^2| + |\widehat{\tau}^2_{old} - \widehat{\tau}^2| + sum(|\widehat{\beta}_{old} - \widehat{\beta}|) < \epsilon, \tag{3.60}$$

where $\epsilon$ could be a very small positive number, e.g., from $10^{-3}$ to $10^{-7}$, or even smaller.

### 3.5.3 Output

Finally, Algorithm 1 returns the estimates: $\widehat{\boldsymbol{\beta}}$, $SE(\widehat{\boldsymbol{\beta}})$, $\widehat{\sigma}^2$, and $\widehat{\tau}^2$ in line 17.

## 3.6 Simulation Study

### 3.6.1 Simulation Procedure

A simulation study was conducted to evaluate the performance of the proposed model. The "true" parameters given in the simulation are fixed; specifically, we set $\boldsymbol{\beta} = (-1, 3, -1)^T$, $\sigma^2 = 0.04$, and $\tau^2 = 0.03$. There are 500 simulated datasets in total, and each dataset has 3,000 observations. The following are the simulation steps.

**Step 1** Generate 20 random numbers from Beta distribution with mean $\frac{\sqrt{2}}{2}$ and variance $\sigma^2$ and denote these by $U_1, ..., U_i, ..., U_{20}$, which represent the random effects of the first crossed factor and can be expressed as:

$$\mathrm{E}(U_i) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(U_i) = \sigma^2.$$

Similarly, generate 75 random numbers from Beta distribution with mean $\frac{\sqrt{2}}{2}$ and variance $\tau^2$ and denote these by $V_1, ..., V_j, ..., V_{75}$, making sure that $\mathbf{U}$ and $\mathbf{V}$ are partially crossed. They represent the random effects of the second crossed factor and can be expressed as:

$$\mathrm{E}(V_j) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(U_j) = \tau^2.$$

**Step 2** Generate response $Y_1, ..., Y_i, ..., Y_N$, given two partially crossed random effects $U_i$ and $V_j$, following Bernoulli distribution:

$$Y_k | \mathbf{U}, \mathbf{V} \sim \mathrm{Bernoulli}(\pi_k U_i V_j),$$

where $\log(\pi_k/(1-\pi_k)) = \mathrm{logit}(\pi_k) = \boldsymbol{x}_k^T \beta$, and $\pi_k = \exp(\boldsymbol{x}_k^T \boldsymbol{\beta})/(1 + \exp(\boldsymbol{x}_k^T \boldsymbol{\beta}))$.

**Step 3** Repeat step 1 and step 2 for 500 times.

Algorithm 1 discussed in Chapter 3 was called to run for 500 times, and the estimated values of the parameters were recorded.

### 3.6.2 Simulation Performance

Table 3.1: Simulation results for the proposed model parameter estimates

| Parameter | True Value | Estimate | Bias | Estimated SE | Simulated SE |
|---|---|---|---|---|---|
| $\beta_0$ | -1.0000 | -0.9953 | 0.0047 | 0.2462 | 0.2482 |
| $\beta_1$ | 3.0000 | 3.0514 | 0.0514 | 0.5566 | 0.5798 |
| $\beta_2$ | -1.0000 | -1.0201 | -0.0201 | 0.1479 | 0.1460 |
| $\sigma^2$ | 0.04 | 0.0366 | -0.0034 | | |
| $\tau^2$ | 0.03 | 0.0308 | 0.0008 | | |

Table 3.1 represents the simulation results for all regression and dispersion parameters based on the proposed model. It should be noted that 496 of all 500 simulations, i.e., 99.2%, converged, which may well happen because binary fitting could be sensitive to the initial values. For example, we have $\boldsymbol{\pi} = \exp(\mathbf{X}^T\boldsymbol{\beta})/(1+\exp(\mathbf{X}^T\boldsymbol{\beta}))$, where the product of matrices $\mathbf{X}^T$ and $\boldsymbol{\beta}$ could give a large number and thus $\exp(\mathbf{X}^T\boldsymbol{\beta})$ could give the infinity, in the situation of which the R programming warns NaNs. Only convergent simulations are used here, and on average, it took about 23 iterations for those convergent simulations to converge.

Here True Value denotes the given values for the parameters based on which we generated our simulation datasets. Estimate, representing the estimated values, is the average of all 496 estimates. Bias is the difference between the Estimate and its corresponding True Value. Estimated SE, representing the estimated standard error, is the average of 496 estimated standard errors, while Simulated SE, meaning the simulated standard error, is the standard deviation of the estimates over 496 simulations.

Table 3.1 shows that the biases for both regression and dispersion parameters are very small; in addition, Estimated SE and Simulated SE values are pretty close.

Therefore, we could say that the estimation algorithm performs reasonably well.

# Chapter 4

# Analysis of Injury Occurrence Data

This chapter applies the binary logistic model with partially crossed random effects proposed in Chapter 3 to motor vehicle accident data. The main objective of this data study is to explore how covariates influence the occurrence of injuries in accidents, while accounting for random agent and location effects.

## 4.1 Exploratory Data Analysis

The data for our binary model study in this Chapter is called the injury occurrence dataset, which is a confidential insurance data about motor vehicle accidents in 2017. Details about the variable description are summarized in Table 2.2 in Chapter 2. It has a total of 12,195 observations.

Table 4.1 illustrates the sample data of the injury occurrence dataset, where Accident Datetime and Registration Date are in the Datetime format, e.g., "2017-12-29 21:40". Since peak traffic times can be closely related to collisions, we extracted the Hour information from the datetime and classified it into three categories, morning peak (7 am–9 am), evening peak (5 pm–8 pm), and off-peak (the rest of the hours),

Table 4.1: Sample data of injury occurrence data

| Agent | Loc | Injury(Y) | GD | IP | IV | Purpose | RD | VA's | Seats | AD |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | No | No | Comprehensive | 186500 | NC | 2003-10-21 0:00 | 14.08 | 5 | 2017-12-29 21:40 |
| 1 | 3 | No | Yes | Compulsory | 40300 | NC | 2010-12-07 0:00 | 6.00 | 8 | 2017-05-01 11:00 |
| 1 | 15 | No | No | Commercial | 34410 | NC | 2006-08-29 0:00 | 9.58 | 8 | 2017-01-29 17:10 |
| 23 | 21 | No | No | Compulsory | 68000 | C | 2008-03-11 0:00 | 9.67 | 3 | 2017-12-13 13:00 |
| 23 | 21 | No | No | M-Compulsory | 4300 | NC | 2013-08-28 0:00 | 4.08 | 2 | 2017-12-27 14:47 |
| 51 | 10 | Yes | No | Comprehensive | 77800 | NC | 2007-11-05 0:00 | 9.00 | 5 | 2017-05-28 16:10 |

* Loc = Location; GD = Good Driver; IP = Insurance Product; IV = Insured Value; NC = Non-commercial; C = Commercial; RD = Registration Date; VA = Vehicle's Age; AD = Accident Datetime; M-Compulsory = Motorcycle-Compulsory.

where the peak periods were split based on the peak-hour standard from the local government. Specifically, we added one new column called Peak in the data, a categorical variable with three levels: morning, evening, and off. A closer check finds that, among the three levels, evening has the highest injury rate, 0.1279, whereas the injury rates for morning and off are 0.0994 and 0.0998. The injury rate for each level is calculated by dividing the total frequency of injury cases by the total number of motor vehicle accidents at that level. In addition, we extracted Year from Registration Date, which basically reflects the same information as Vehicle's Age, so we only kept Vehicle's Age. Also, we investigated if there is a correlation between Insurance Product and Insured Value, and there is no evidence showing that. The summary statistics for numerical and categorical variables are displayed in Table 4.2 and Table 4.3. Before fitting the model to the data, we standardized Insured Value so that numeric variables, Insured Value, Seats, and Vehicle's Age, are on similar scales.

Table 4.2: Descriptive statistics for numerical variables

| Variable | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|
| Insured Value | 2200 | 2214000 | 95165 | 100051.6 |
| Vehicle's Age | 0 | 28.500 | 4.342 | 3.7902 |
| Seats | 1 | 51 | 5.661 | 1.4061 |

Table 4.3: Descriptive statistics for categorical variables

| Variable | Categories | Injury Frequency | |
|---|---|---|---|
| | | Yes | No |
| Good Driver | Yes | 932 | 7781 |
| | No | 356 | 3126 |
| Purpose | Commercial | 20 | 113 |
| | Non-Commercial | 1268 | 10794 |
| Peak | Evening | 326 | 2223 |
| | Morning | 234 | 2120 |
| | Off | 728 | 6564 |
| Insurance Product | Special Commercial Vehicle | 0 | 1 |
| | Telemarketing Commercial | 0 | 2 |
| | Compulsory Motorcycle | 61 | 208 |
| | Compulsory Motor Vehicle | 366 | 2424 |
| | Commercial Motor Vehicle | 25 | 214 |
| | Comprehensive Motor Vehicle | 836 | 8058 |

## 4.2 Model Setting

We apply the proposed model to analyze the likelihood of people getting injured in motor vehicle accidents. We denote the crossed random effects Agent and Location as $\mathbf{U}$ and $\mathbf{V}$, respectively. We set variable Injury, the occurrence of injuries (1=yes, 0=no), as the response $\mathbf{Y}$, where $Y_k$ represents the $k$th observation with a multiplicative combination of $U_i$ and $V_j$. Specifically, our objective is to investigate the relationship between the occurrence of injuries and the covariates, such as Insured Value, while accounting for agent and location random effects. Following the model assumptions in Chapter 3, we specify the assumptions as follows:

**Assumption 1:** The first random effect Agent $U_1, \cdots, U_i, \cdots, U_I$ are positive, independently and identically distributed on $(0, 1)$ with

$$\mathrm{E}(U_i) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(U_i) = \sigma^2.$$

**Assumption 2:** The second random effect Location $V_1, \cdots, V_j, \cdots, V_J$ are positive, independently and identically distributed on $(0,1)$ with

$$\mathrm{E}(V_j) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(V_j) = \tau^2.$$

Let $\mathbf{W}$ represent the multiplicative combination of random effects $\mathbf{U}$ and $\mathbf{V}$, which is a vector with $I \times J$ components,

$$\mathbf{W} = (U_1V_1, ..., U_1V_j, ..., U_1V_J, ..., U_iV_j, ..., U_IV_1, ..., U_IV_j, ..., U_IV_J)^T.$$

In addition, we create an indication matrix $\mathbf{B}$ of size $N \times (I \times J)$,

$$\mathbf{B} = (\mathbf{B}_1, ..., \mathbf{B}_k, ..., \mathbf{B}_N)^T,$$

where row $\mathbf{B}_k$ indicates the multiplicative combination of $U_iV_j$ related to the $k$th observation. Specifically, $\mathbf{B}_k$ is a vector with $(I \times J - 1)$ 0's and a value 1 located at the location of $U_iV_j$ in the vector $\mathbf{W}$.

In our dataset, there are 71 different agents and 127 different locations. However, among the 12,195 observations in our dataset, there are only 1,209 different combinations of Agent and Location. That is, not all $9,017$ $(= 71 \times 127)$ possible combinations of Agent and Location appear in our dataset. Therefore, the random effects, Agent and Location, are partially crossed.

**Assumption 3:** Agent $\mathbf{U} = (U_1, ..., U_i, ..., U_I)$ and Location $\mathbf{V} = (V_1, ..., V_j, ..., V_J)$ are independent with conditional distribution given below,

$$Y_k | \mathbf{U}, \mathbf{V} \sim \mathrm{Bernoulli}(\pi_k U_i V_j),$$

where $\log(\pi_k/(1-\pi_k)) = \mathrm{logit}(\pi_k) = \boldsymbol{x}_k^T\boldsymbol{\beta}$, $\pi_k = \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))$, and $\boldsymbol{\beta}$ is a set of regression parameters.

Here we assume Agent and Location are independent because they are conceptually independent in nature: one is the insurance product seller, and the other is the motor vehicle accident location. As a result, we do not consider concretely how the agent and motor vehicle accidents might be locally related.

## 4.3 Data Analysis

### 4.3.1 Experiment Setting

We mainly use R as our programming language for parameter estimation and simulation studies . However, given there are 12,195 observations in the dataset, we need to calculate the inverse of a large matrix Var($\mathbf{Y}$) of size $12,195 \times 12,195$ in estimating procedures, which could be computationally intensive and time consuming. Therefore, we run Python chunks embedded within the R code for matrix inverse calculation. We implement Algorithm 1 in Chapter 3 to the dataset in R on a platform with Apple M1 chip, 16 GB RAM, and macOS Big Sur operating system. The estimation algorithm iterates around 80–150 times (depending on the number of variables) to converge until the *diff* is less than the specified value $\epsilon = 10^{-4}$, where

$$diff = |\widehat{\sigma}^2_{old} - \widehat{\sigma}^2| + |\widehat{\tau}^2_{old} - \widehat{\tau}^2| + sum(|\widehat{\beta}_{old} - \widehat{\beta}|) < \epsilon.$$

### 4.3.2 Estimation Results

#### 4.3.2.1 Full model

We first fit the injury occurrence dataset to the logistic regression model, GLM (generalized linear model) with logit link, without accounting for agent and location random effects. Then we fit the data to our binary logistic model with partially crossed random effects. The estimation results are presented in Table 4.4 and Table 4.5.

Table 4.4: Parameter estimation results based on GLM

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -2.7620 | 0.1295 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.1703 | 0.1759 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.2865 | 0.0734 | 0.0000 |
| Insurance Product-Commercial Motor Vehicle | 0.0765 | 0.2159 | 0.7230 |
| Good Driver-Yes | -0.0698 | 0.0678 | 0.3032 |
| Purpose-Commercial | 0.3648 | 0.2470 | 0.1398 |
| Insured Value | -0.1270 | 0.0392 | 0.0012 |
| Vehicle's Age | 0.0126 | 0.0084 | 0.1351 |
| Seats | 0.0714 | 0.0201 | 0.0004 |
| Peak-evening | 0.2593 | 0.0714 | 0.0003 |
| Peak-morning | 0.0042 | 0.0796 | 0.9576 |

Table 4.5: Parameter estimation results based on the proposed binary model-full model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -2.0059 | 0.1797 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.4575 | 0.2611 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.3432 | 0.0841 | 0.0000 |
| Insurance Product-Commercial Motor Vehicle | 0.0907 | 0.2398 | 0.7052 |
| Good Driver-Yes | -0.1011 | 0.0763 | 0.1854 |
| Purpose-Commercial | 0.4044 | 0.3067 | 0.1873 |
| Insured Value | -0.1013 | 0.0400 | 0.0114 |
| Vehicle's Age | 0.0217 | 0.0097 | 0.0260 |
| Seats | 0.0908 | 0.0281 | 0.0012 |
| Peak-evening | 0.3135 | 0.0818 | 0.0001 |
| Peak-morning | 0.0381 | 0.0876 | 0.6633 |
| $\sigma^2$ | 0.0042 | | |
| $\tau^2$ | 0.0361 | | |

We observe that the effects of some covariates are statistically insignificant at the significance level of 0.05 in both Table 4.4 and Table 4.5: Insurance Product-Commercial Motor Vehicle, Good Driver-Yes, Purpose-commercial, and Peak-morning.

We cannot directly compare the magnitude of the estimates obtained from these two models because for the same covariate, the true values of the regression parameter in the two models are intrinsically different. As we know, $\pi_k = \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1 + \exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))$ in logistic regression, whereas the marginal probability $p_k = \frac{1}{2}\pi_k$ from our derivation (refer to Eq. (4.1)).

In terms of statistical significance, the effects of all covariates in both tables are generally consistent at the significance level of 0.05, except for Vehicle's Age. Vehicle's Age is only significant in estimation results in our proposed model when random agent and location effects are considered.

#### 4.3.2.2 Final model

As shown in the full model in Table 4.5, Insurance Product-Compulsory Motorcycle, Insurance Product-Compulsory Motor Vehicle, and Insurance Product-Commercial Motor Vehicle are three levels of the categorical variable Insurance Product. Insurance Product has six levels in total: Compulsory Motorcycle Insurance, Compulsory Motor Vehicle Insurance, Commercial Motor Vehicle Insurance, Comprehensive Motor Vehicle Insurance, Special Commercial Vehicle Insurance, and Telemarketing Commercial Insurance. The last two levels were deleted from the data because they only include one and two observations, respectively. Generally, a categorical variable with $n$ levels is transformed into $n - 1$ indicator variables each with two levels (STHDA, 2018). The default option in R is to use the level of the variable that first appears in the dataset as a reference, and we need to interpret the remaining levels relative to it. Here, we set Comprehensive Motor Vehicle Insurance as the reference level because most of the motor vehicles have this type of insurance. Similarly, the

Table 4.6: Parameter estimation results-final model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -1.9892 | 0.1763 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.4642 | 0.2596 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.3433 | 0.0835 | 0.0000 |
| Insured Value | -0.1024 | 0.0399 | 0.0103 |
| Vehicle's Age | 0.0208 | 0.0097 | 0.0313 |
| Seats | 0.0874 | 0.0278 | 0.0017 |
| Peak-evening | 0.3049 | 0.0788 | 0.0001 |
| $\sigma^2$ | 0.0038 | | |
| $\tau^2$ | 0.0369 | | |

categorical variable Peak has three levels – evening, morning, and off– and we set the dummy variable, off, as the reference level because off-peak hours represent the highest proportion of all the hours.

We reduce those insignificant covariates by backward stepwise selection (or backward elimination). It starts with the full model, and the least significant (highest p-value) covariate is removed from the model one after the other. The elimination process is terminated when all the insignificant covariates are removed, yielding a final model with only significant covariates.

Regarding estimates for Insurance Product-Commercial Motor Vehicle and Peak-morning, they are not statistically different from their reference levels, Comprehensive Motor Vehicle Insurance and Peak-off, respectively, because of the large p-values of 0.7052 and 0.6633 as shown in Table 4.5. Therefore, Commercial Motor Vehicle Insurance can be merged into Comprehensive Motor Vehicle Insurance and morning peak hours can be merged into off-peak hours during the elimination process, yielding new reference levels for Insurance Product and Peak, respectively. The estimation results of the final model are summarized in Table 4.6, where, for example, Peak now consists of two levels instead, Peak-evening and Peak-non-evening.

#### 4.3.2.3 Interpretation of parameter estimates

Based on our model assumption 3, we have $Y_k|\mathbf{U},\mathbf{V} \sim \text{Bernoulli}(\pi_k U_i V_j)$, where $\log(\pi_k/(1-\pi_k)) = \boldsymbol{x}_k^T\boldsymbol{\beta}$, and $\pi_k = \exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta}))$. In addition, based on our derivation in Eq. (3.5), the unconditional expectation $\text{E}(Y_k) = \frac{1}{2}\pi_k$. Therefore, the marginal probability of the occurrence of injuries in motor vehicle accidents can be expressed as

$$p_k = \frac{1}{2}\pi_k = \frac{1}{2}\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})/(1+\exp(\boldsymbol{x}_k^T\boldsymbol{\beta})), \tag{4.1}$$

which denotes the average risk that can be explained by the covariates.

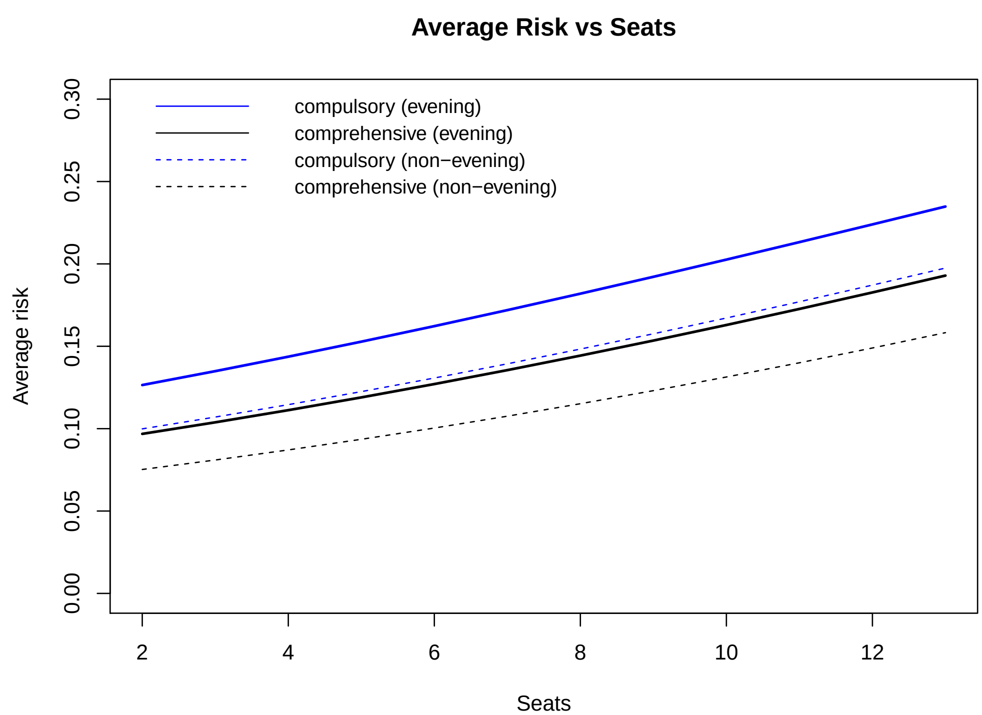


Figure 4.1: Average risk of injury vs Seats

The regression parameter estimates (the estimated coefficients) for covariates in the final model in Table 4.6 are interpreted in terms of their relationship to the average risk of the occurrence of injuries in motor vehicle accidents. Take covariates Seats

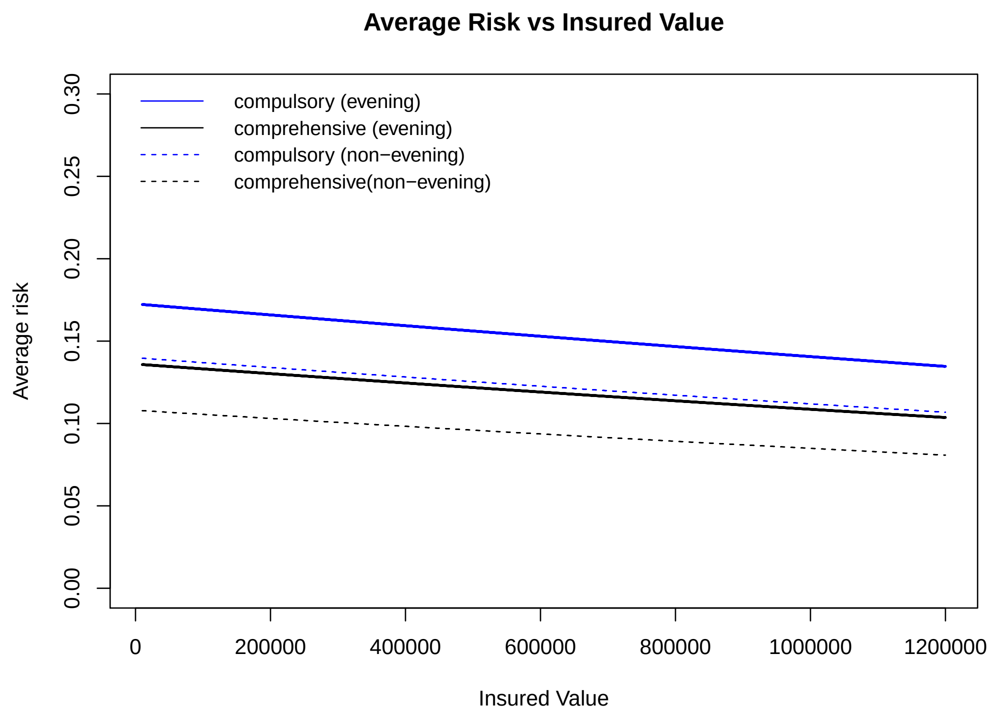


Figure 4.2: Average risk of injury vs Insured Value

and Insured Value for example, Fig. 4.1 and Fig. 4.2 display how the average risk changes according to the number of seats and insured value, respectively. Insured Value and Vehicle's Age are held constant in Fig. 4.1, whereas Seats and Vehicle's Age are held constant in Fig. 4.2.

In both figures, blue and black lines denote compulsory motor vehicle insurance and comprehensive motor vehicle insurance (reference level), respectively, whereas solid and dashed lines represent evening-peak hours and non-evening-peak hours (reference level), respectively. The figures demonstrate that, when holding all other covariates constant, the number of seats correlates positively with injury, suggesting that a greater number of seats is associated with a higher average risk. In contrast, a higher insured value is associated with a lower risk.

As shown in both figures, the solid blue lines at the top denote that when a vehicle has compulsory motor vehicle insurance and driving occurs in evening peaks, the average risk of injuries in a motor vehicle accident is higher than that of the other situations listed here. In contrast, the dotted black lines at the bottom of the figures indicate that when a vehicle has comprehensive motor vehicle insurance and driving occurs in non-evening-peak hours, it is less likely to be associated with human injuries from motor vehicle accidents.

The compulsory motorcycle insurance is not illustrated in the comparison in Fig. 4.1 and Fig. 4.2 since motorcycles only have either 1 or 2 seats in the given data. We observe that the estimated value for Insurance Product-Compulsory Motorcycle, 1.4642, is much greater than that of Insurance Product-Compulsory Motor Vehicle, 0.3433. It indicates that, compared to a change to compulsory motor vehicle insurance, changing to compulsory motorcycle insurance from the reference level comprehensive motor vehicle insurance results in a higher increase in the average risk.

The estimated value of the dispersion parameter $\sigma^2$ ($\widehat{\sigma}^2 = 0.0038$) shows the dispersion for heterogeneity among agents, whereas $\tau^2$ ($\widehat{\tau}^2 = 0.0369$) represents the dispersion for heterogeneity among locations, after the covariates effects are explained. It indicates that the variation across locations is larger than that among agents.

#### 4.3.2.4 Interpretation of random effects

Based on the predicted results of the crossed random effects Agent and Location, two scatterplots are plotted as shown in Fig. 4.3 and Fig. 4.4, where the two dashed black lines represent the predicted values of two standard deviations away from the mean of $U_i$ and $V_j$ , both of which are $\frac{\sqrt{2}}{2}$, as indicated by the orange dotted centerline.

The predicted agent random effects are displayed in Fig. 4.3, which demonstrate agent-specific risks relative to the average risk of the occurence of injuries that can be explained by the covariates. Generally, if the predicted random effect is higher than $\frac{\sqrt{2}}{2}$, the agent is associated with a higher relative risk than average. Conversely,

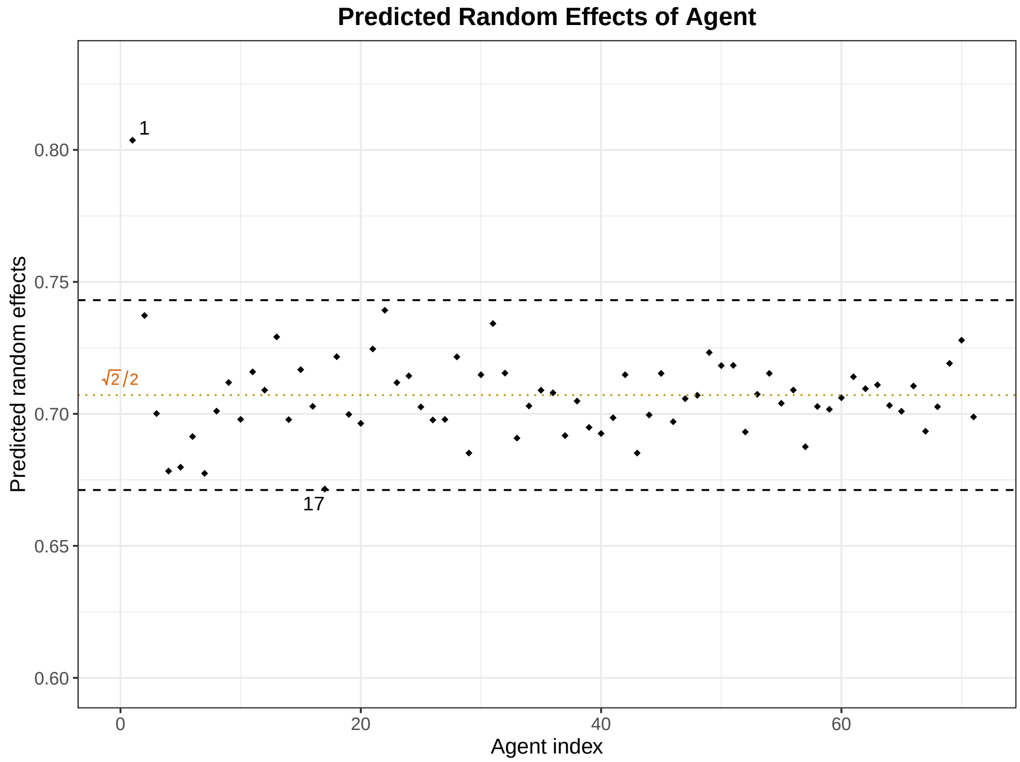


Figure 4.3: Predicted random effects of agent

when the predicted random effect is lower than $\frac{\sqrt{2}}{2}$, the agent is associated with a lower relative risk. It shows that there is only one data point that is not within two standard deviations of the centerline. Specifically, agent 1 has the largest predicted random effect, and agent 17 has the lowest. Generally, after the covariates effects are explained, we could say that purchasing insurance from agent 1 is associated with a higher agent-specific relative risk of injury than average, whereas a lower agent-specific relative risk is associated with insurance purchases from agent 17. It may be the case that some agents tend to select specific type of clients who are more prone to injuries due to motor vehicle accidents. For example, certain agents might prefer those young drivers who need to pay relatively higher insurance premiums.

The predicted location random effects are shown in Fig. 4.4, which show location-specific relative risks to the average risk that can be explained by the covariates. Generally, if the predicted random effect is higher than $\frac{\sqrt{2}}{2}$, the location has a higher

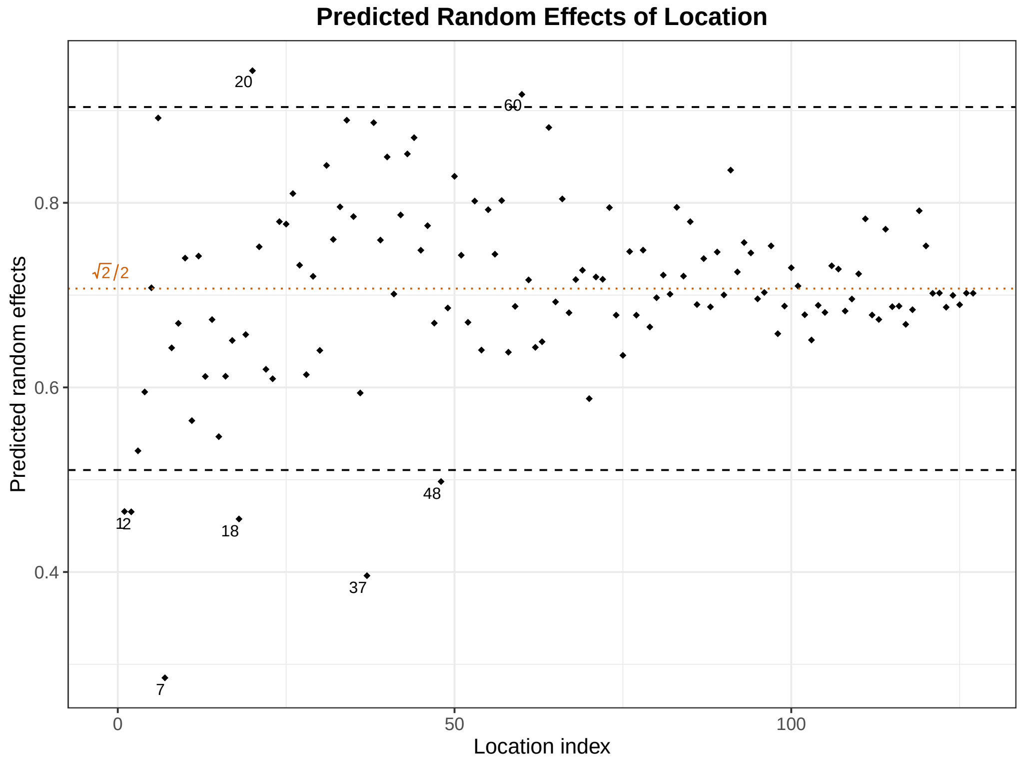


Figure 4.4: Predicted random effects of location

relative risk than average, whereas if the predicted random effect is lower than $\frac{\sqrt{2}}{2}$, the location has a lower relative risk. There are 6 data points – 1, 2, 7, 18, 37 and 48 – two standard deviations below the mean, two data points – 20 and 60 – above the mean, and the rest are all within the two standard deviations of the mean. Specifically, locations 7 and 37 have the lowest predicted values. Therefore, we could say that locations 20 and 60 would have higher and locations 7 and 37, particularly, would have lower location-specific relative risks of injury than average.

Corresponding to the scatter plot in Fig. 4.4, Fig. 4.5 is a map plot of the random effects of Location with spatial coordinates, where *na* indicates no data for the two gray locations. The darker the blue, the higher the value. The labelled locations are those with predicted random effects two standard derivations away from the mean. The likelihood of whether a motor vehicle accident would involve human injury is determined by various factors associated with the complication of local conditions.

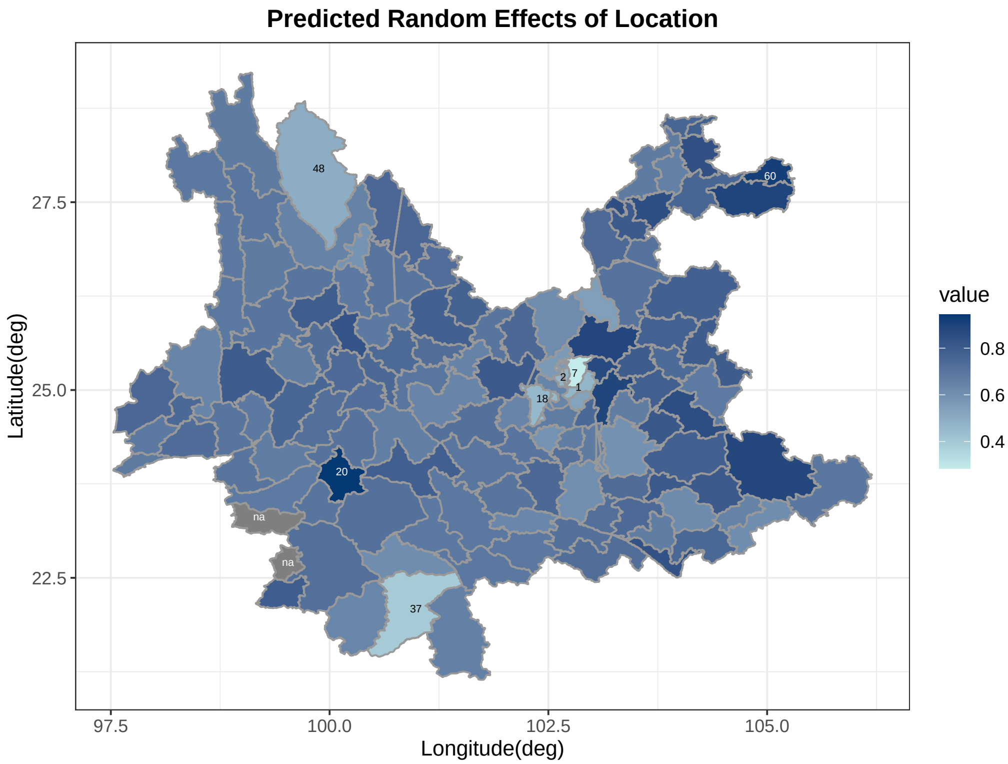


Figure 4.5: Predicted random effects of location with spatial coordinates

These factors include residents's socioeconomic status, age, and gender; driver's education; driving distractions; road infrastructure safety; vehicle safety; etc. (World-Health-Organization, 2021). Given that multiple factors are involved in the explanation of the injuries, the following example provides only one possible interpretation. For example, locations 20 and 60 have a higher relative risk of injury than average in motor vehicle accidents possibly because the former has two high-traffic national highways and an airport, and the latter is within a mountain area with rolling roads. In contrast, locations 1, 2, 7 and 18 have a lower relative risk of injury possibly because these locations are highly developed with infrastructures that contain well-designed junctions and better overall signage. Additionally, location 48 is sparsely populated, and location 37 is a tourist area with likely mature traffic circulations. Therefore, traffic collisions in locations 1, 2, 7, 18, 37, and 48 have a lower relative risk than average.

# Chapter 5

# Analysis of Injury Severity Data

This chapter first introduces our baseline-category logit model with partially crossed random effects for multinomial analysis. Then, it applies the model to analyze the injury severity dataset. Finally, it illustrates and interprets the estimation results.

## 5.1 Baseline-Category Logit Models with Partially Crossed Random Effects

In this section, we introduce our baseline-category logit model with partially crossed random effects, which is essentially an extension of the binary logistic model with partially crossed random effects proposed in Chapter 3.

### 5.1.1 Review of Baseline-Category Logit Models

Assume that the response $\mathbf{Y}$ has $R(1, \ldots, r, \ldots, R)$ response categories in total. Let $\boldsymbol{\pi}_r(\mathbf{X}) = P(\mathbf{Y} = r|\mathbf{X})$ at a fixed dataset; for observations at that setting, the number of observations out of $N$ total observations at the different $R$ categories is considered as a multinomial variate with probabilities $\boldsymbol{\pi}_1(\mathbf{X}), \ldots, \boldsymbol{\pi}_R(\mathbf{X})$, and $\sum_{r=1}^{R} \boldsymbol{\pi}_r(\mathbf{X}) = 1$ (Agresti, 2003).

Assume $r^*$ is the baseline category, which can be the last one or the most common one, then for each $r \in R$, the corresponding baseline logit can be written as,

$$\log\left(\frac{\boldsymbol{\pi}_r}{\boldsymbol{\pi}_{r^*}}\right) = \mathbf{X}^T\boldsymbol{\beta}_r, r \neq r^*, \tag{5.1}$$

where each response category is paired with the baseline category $r^*$. Thereby, we have $R-1$ baseline-category logits in total.

### 5.1.2 Baseline-Category Logit Models with Partially Crossed Random Effects

Incorporating partially crossed random effects, we set $r^*$ as the baseline category and get $R-1$ baseline-category logit models with partially crossed random effects. Each of the $R-1$ models is essentially the binary logistic model with partially crossed random effects proposed in Chapter 3.

Specifically, each of the $R-1$ models is analyzed on the $r$th subset of the entire data that only includes observations belonging to $r$th and the baseline $r^*$ categories. Each model has the following assumptions:

**Assumption 1:** The first random effects $U_1, \cdots, U_i, \cdots, U_I$ are positive, independently and identically distributed on $(0, 1)$ with

$$\mathrm{E}(U_i) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(U_i) = \sigma^2.$$

**Assumption 2:** The second random effects $V_1, ..., V_j, ..., V_J$ are positive, independently and identically distributed on $(0, 1)$ with

$$\mathrm{E}(V_j) = \frac{\sqrt{2}}{2} \text{ and } \mathrm{Var}(V_j) = \tau^2.$$

**Assumption 3:** For observations in the dataset, $\mathbf{U} = (U_1, ..., U_i, ..., U_I)$ and $\mathbf{V} =$

$(V_1, ..., V_j, ..., V_J)$ are independent with conditional distribution given below,

$$Y_{rk}|\mathbf{U}, \mathbf{V} \sim \text{Bernoulli}(\pi_{rk} U_i V_j),$$

where $\log(\pi_{rk}/\pi_{r^*k}) = \boldsymbol{x}_{rk}^T \boldsymbol{\beta}_r$, and $\pi_{rk} = \exp(\boldsymbol{x}_{rk}^T \boldsymbol{\beta}_r)/(1 + \exp(\boldsymbol{x}_{rk}^T \boldsymbol{\beta}_r))$.
$\mathbf{Y} = (Y_{r1}, \ldots, Y_{rk}, \ldots, Y_{rN})$ is a response vector, with $Y_{rk}$ denoting the $k$th response in the $r$th subset including only observations belonging to the $r$th and the baseline $r^*$th categories. $\boldsymbol{\beta}_r$ =($\boldsymbol{\beta}_{r0}$, $\boldsymbol{\beta}_{r1}$, . . ., $\boldsymbol{\beta}_{rp}$) is a vector of regression parameters. Here $i$ and $j$ are determined for each $rk$. That is, $i = i(rk)$ and $j = j(rk)$.

## 5.2 Data Description

The injury severity data for the model study here is generally the same as the injury occurrence data in Chapter 4. The only difference is that the category of yes in binary response Injury (yes or no injuries) in the injury occurrrence data is further divided into three severity levels– death, serious injuries and minor injuries– in the injury severity data. The category of no injuries in the two datasets stays the same. The four categories of the response variable Severity are denoted as Death, Serious, Minor, and No. There are 12,195 observations in the dataset. The details of the variables are summarized in Table 2.2 in Chapter 2.
Table 5.1 shows the sample of the injury severity dataset. Of 12,195 observations, 10,907 fall in the No (no injuries) category, 800 in the Minor (minor injuries) category, 469 in the Serious (serious injuries) category, and only 19 in the Death category. Same as those in Chapter 4, Insured Value was standardized so that all the numeric variables Insured Value, Vehicle's Age and Seats are on similar scales.

Table 5.1: Sample data of injury severity data

| Agent | Loc | Severity(Y) | IP | Purpose | GD | IV | VA | Seats | Peak |
|---|---|---|---|---|---|---|---|---|---|
| 3 | 4 | Serious | Comprehensive | NC | No | 74000 | 3.08 | 5 | off |
| 21 | 26 | Minor | Comprehensive | C | Yes | 83300 | 0.92 | 5 | off |
| 43 | 8 | No | Commercial | NC | No | 224500 | 11.00 | 5 | morning |
| 1 | 16 | No | Compulsory | C | No | 78000 | 6.92 | 3 | evening |
| 21 | 43 | Death | Comprehensive | NC | No | 54800 | 4.00 | 5 | off |
| 4 | 63 | No | M-Compulsory | NC | No | 10300 | 0.00 | 2 | evening |
| 21 | 40 | Minor | Comprehensive | NC | No | 40000 | 4.58 | 7 | morning |

* Loc = Location; IP = Insurance Product; NC = Non-commercial; C = Commercial; GD = Good Driver; IV = Insured Value; VA = Vehicle's Age; M-Compulsory = Motorcycle-Compulsory.

## 5.3 Multinomial Analysis of Injury Severity Data

In this section, considering the random effects Agent and Location, we use our proposed baseline-category logit model with partially crossed random effects to investigate how certain insurance and vehicle information could be related to injury severity in motor vehicle accidents in 2017, while accounting for random effects.

First, we set the category No as the baseline category, as it is the most common one, i.e., with the largest number among the four categories. Next, we pair each of the remaing three categories with the baseline category and subset three different subsets from the whole dataset. Specifically, observations in Minor and No categories form a subset, called minor injury subset ($D_{MN}$), where we set Minor as 1 and No as 0. Similarly, we have the other two datasets, serious injury subset ($D_{SN}$), and death case subset ($D_{DN}$).

Applying the binary logisitic model with partially crossed random effects to each of the three datasets $D_{MN}$, $D_{SN}$, and $D_{DN}$, we can obtain the corresponding parameter estimation for minor injuries, serious injuries, and death cases.

## 5.4 Analysis Results

We conduct the baseline-category logit model with partially crossed random effects model to the motor vehicle dataset with the same experimental setting as in Chapter 4. We use R as the main programming language, with Python embedded within the R session for matrix inverse calculation. Algorithm 1, discussed in Chapter 3, is used to estimate the regression and dispersion parameters.

### 5.4.1 Minor Injury Subset

#### 5.4.1.1 Interpretation of parameter estimates

Table 5.2: Estimation results based on the proposed baseline-category model (minor injuries)-full model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -2.3728 | 0.1923 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.2096 | 0.2691 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.2800 | 0.0989 | 0.0046 |
| Insurance Product-Commercial Motor Vehicle | 0.2743 | 0.2690 | 0.3078 |
| Good Driver-Yes | -0.0526 | 0.0902 | 0.5602 |
| Purpose-Commercial | -0.1707 | 0.4064 | 0.6744 |
| Insured Value | -0.0904 | 0.0473 | 0.0562 |
| Vehicle's Age | 0.0185 | 0.0114 | 0.1051 |
| Seats | 0.0541 | 0.0293 | 0.0646 |
| Peak-evening | 0.3517 | 0.0950 | 0.0002 |
| Peak-morning | 0.0900 | 0.1036 | 0.3846 |
| $\sigma^2$ | 0.0047 | | |
| $\tau^2$ | 0.0444 | | |

We first pair category Minor with the baseline category No, aiming to investigate how covariates affect minor injuries due to motor vehicle accidents, while accounting for agent and location random effects. The parameter estimates are illustrated in Table 5.2. Insurance Product-Compulsory Motorcycle, Insurance Product-Compulsory Motor Vehicle, and Peak-evening are statistically significant at the significance level of 0.05. As mentioned in Chapter 4, the reference category for Insurance Product is Comprehensive Motor Vehicle Insurance and for Peak is off. By applying backward stepwise selection and merging the insignificant levels to their corresponding reference levels, we have the final reduced model. The estimation results are displayed in Table 5.3.

Table 5.3: Estimation results based on the proposed baseline-category model (minor injuries)-final model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -1.9946 | 0.0802 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.0151 | 0.2399 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.3317 | 0.0921 | 0.0003 |
| Insured Value | -0.0997 | 0.0478 | 0.0372 |
| Peak-evening | 0.3312 | 0.0916 | 0.0003 |
| $\sigma^2$ | 0.0058 | | |
| $\tau^2$ | 0.0448 | | |

The positive sign of estimates for Insurance Product-Compulsory Motorcycle and Insurance Product-Compulsory Motor Vehicle indicates that, all else being equal, motor vehicles with compulsory motorcycle insurance and compulsory motor vehicle insurance have a relatively higher probability of minor injuries due to traffic collisions in comparison to those with comprehensive motor vehicle insurance. Similarly, the positive sign of Peak-evening indicates that, in comparison to non-evening-peak hours, evening-peak hours are more likely to be associated with minor injuries. It makes sense that people could be tired during evening peak hours and behave more recklessly on the road after a day's work, and thus it is more likely that minor injuries occur.

Similar to the estimates interpretation in Chapter 4, Fig. 5.1 illustrates how the average risk of minor injuries varies according to Insured Value. As shown in the figure, the solid blue line at the top denotes when a vehicle has compulsory motor vehicle insurance and driving occurs in evening peaks, it is more likely to be involved in minor injuries in comparison to the other situations listed here. The comprehensive (evening) line and the compulsory (non-evening) line overlap, forming the middle black line in the figure, because the estimates for Insurance Product-Compulsory Motor Vehicle insurance dummy and Peak-evening dummy are almost the same, 0.3317 and 0.3312, respectively.

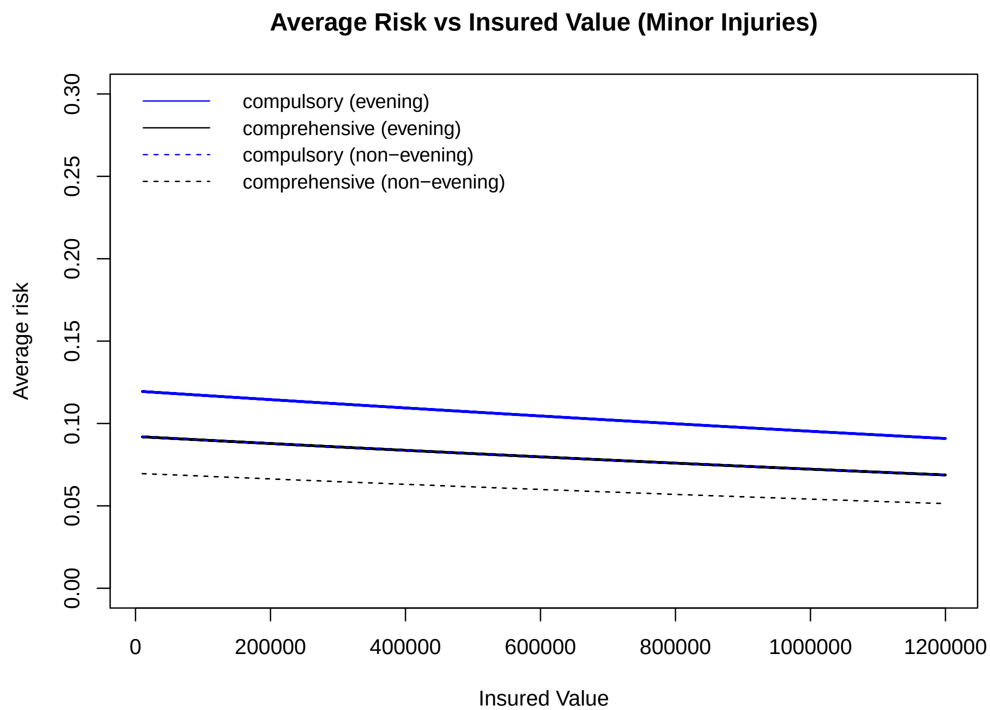


Figure 5.1: Average risk of minor injuries vs Insured Value

#### 5.4.1.2 Interpretation of random effects

Predicted random effects for Agent and Location are discussed together in this part, which are displayed in scatterplots in Fig. 5.2 and Fig. 5.3, respectively. Also, there is one map plot in Fig. 5.4 corresponding to the scatterplot for Location.

Like the scatterplots of predicted random effects in Chapter 4, the dotted centerline represents the mean of the crossed random effects Agent and Location, $\frac{\sqrt{2}}{2}$, with two dashed lines on both sides, denoting the predicted values of two standard deviations away from the mean.

First, the predicted agent random effects are shown in Fig. 5.2, which demonstrate agent-specific risks relative to the average risk of minor injuries that can be explained by the covariates. Fig. 5.2 shows that agents 1 and 21 have the two largest predicted random effects, whereas agent 17 has the smallest. The results suggest that, holding all covariates constant, there would be a higher agent-specific relative risk of minor

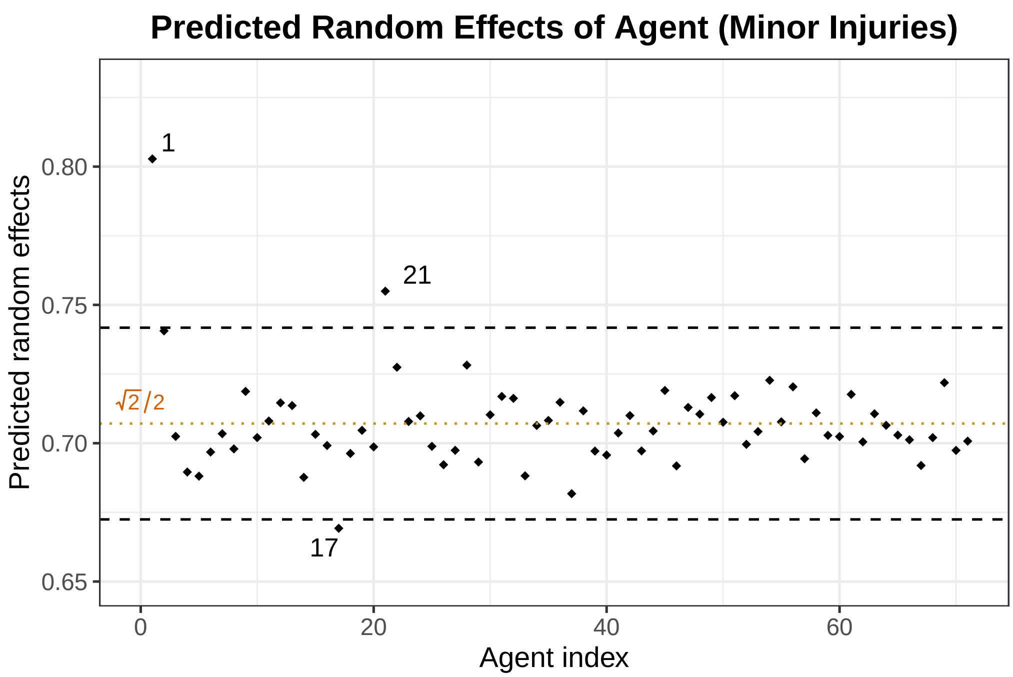


Figure 5.2: Predicted random effects of agent (minor injuries)

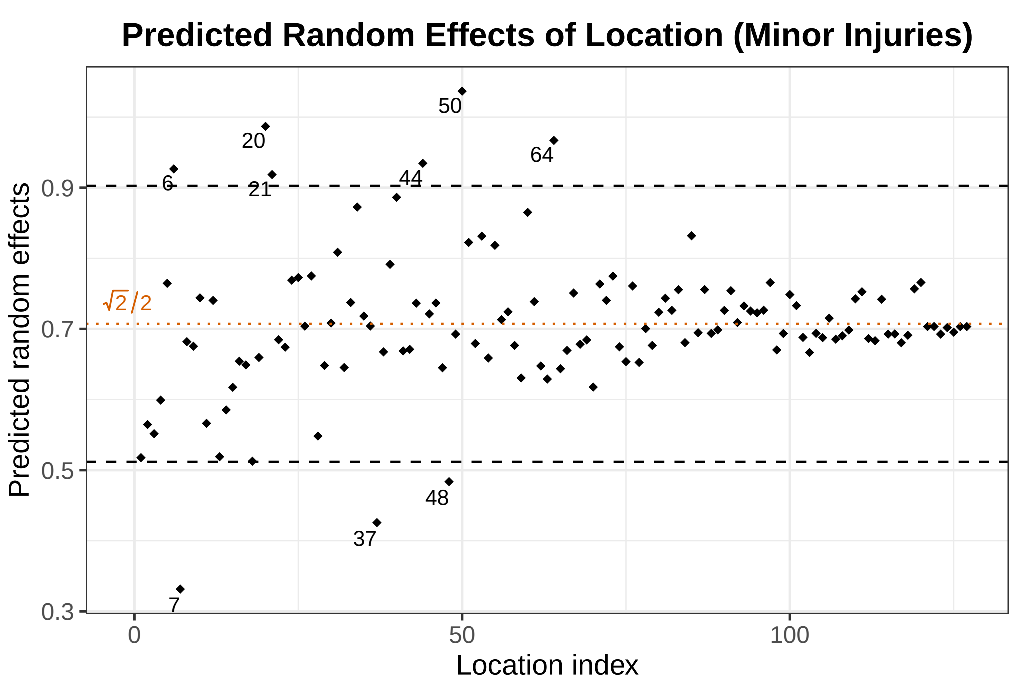


Figure 5.3: Predicted random effects of location (minor injuries)

injuries than average when the insurance is purchased from agent 1 or agent 21, whereas the relative risk is lower when the insurance is purchased from agent 17.

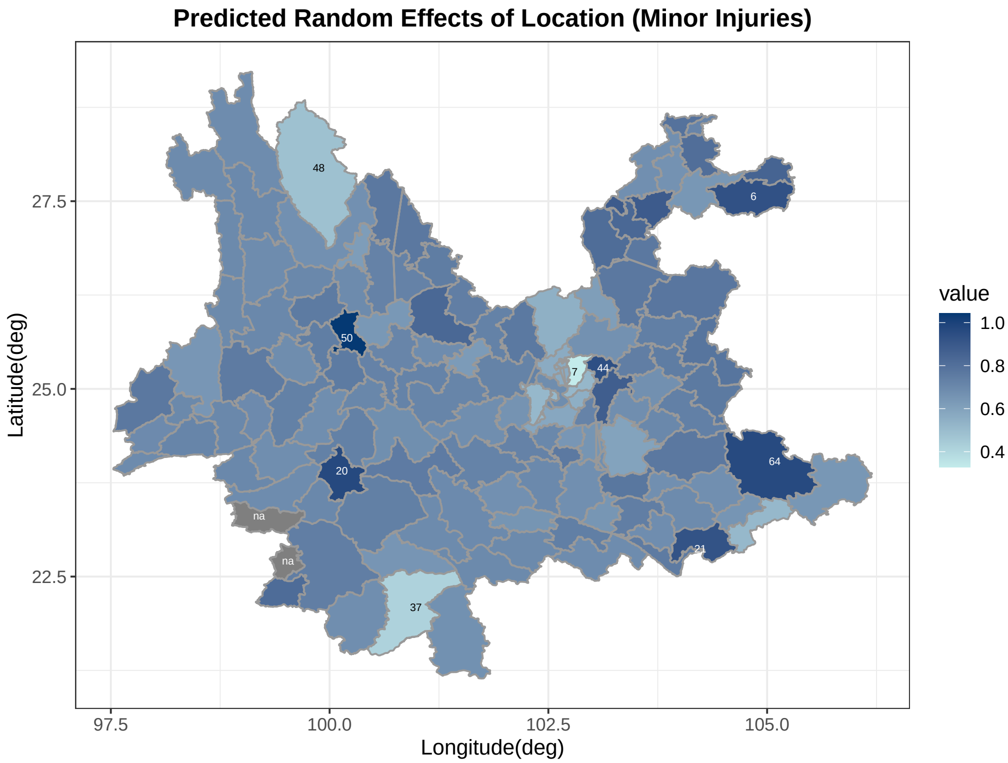


Figure 5.4: Predicted random effects of location with spatial coordinates (minor injuries)

Second, Fig. 5.3 and Fig. 5.4 suggest that, locations 6, 20, 21, 44, 50, and 64 have the largest predicted random effects, which indicate that these locations have a higher location-specific relative risk of minor injuries than the average risk that can be explained by the covariates. In contrast, as indicated by their small predicted random effects, the relative risk is lower than average for locations 7, 37, and 48.

Similarly as what discussed in Chapter 4, whether there would be minor human injuries in a motor vehicle accident is determined by many local conditions. Because multiple factors are involved in the occurrence of minor injuries, the following examples provide only some possible interpretations. For example, locations 6, 21 and 64 have a higher relative risk of minor injuries probably because they are impoverished areas, with GDP per capita ranking very low in the province. These locations may lack well-designed traffic circulations and contain confusing intersections with relatively low traffic speed, leading to the higher relative risk of minor injuries. In addition, as shown in the map plot in Fig. 5.4, locations 20, 44 and 50 also have a higher relative risk of minor injuries than average. It might be due to the heavy traffic surrounding the highways and airport in location 20 and impoverished traffic infrastructures in locations 44 and 50.

As discussed in Chapter 4, locations 7, 37 and 48 have a lower risk of injury possibly because location 7 is a highly developed city with well-maintained traffic infrastructures, location 37 is a tourist area with likely mature traffic circulations, and location 48 is sparsely populated. These local conditions may also explain the relatively lower probability of minor injuries associated with these locations. Since the minor dataset and the injury occurrence dataset share the same no-injury subset, and minor injuries in the former dataset are part of the injury cases in the latter dataset, therefore, locations with a lower relative risk of injury tend to have a lower relative risk of minor injuries.

### 5.4.2 Serious Injury Subset

#### 5.4.2.1 Interpretation of parameter estimates

Next, we pair Serious with the baseline category No, and the estimation results are shown in Table 5.4. Insurance Product-Compulsory Motorcycle, Insurance Product-Compulsory Motor Vehicle, Purpose-Commercial, Insured Value, and Seats are statistically significant at the p-value of 0.05.

By applying backward stepwise selection and merging the insignificant levels into their corresponding reference levels, we have the final reduced model. The estimation results are displayed in Table 5.5. Here Insured Value is kept in the final model partly because the estimated coefficient has a p-value of 0.0587, which is only slightly higher than 0.05. In addition, if Insured Value is removed from the final model, the

Table 5.4: Estimation results based on the proposed baseline-category model (serious injuries)-full model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -3.1187 | 0.2330 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.3004 | 0.3145 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.3989 | 0.1153 | 0.0005 |
| Insurance Product-Commercial Motor Vehicle | -0.2027 | 0.3796 | 0.5933 |
| Good Driver-Yes | -0.1608 | 0.1089 | 0.1399 |
| Purpose-Commercial | 1.0052 | 0.3430 | 0.0034 |
| Insured Value | -0.1227 | 0.0613 | 0.0454 |
| Vehicle's Age | 0.0184 | 0.0136 | 0.1761 |
| Seats | 0.1113 | 0.0355 | 0.0017 |
| Peak-evening | 0.1946 | 0.1142 | 0.0886 |
| Peak-morning | -0.0276 | 0.1256 | 0.8263 |
| $\sigma^2$ | 0.0078 | | |
| $\tau^2$ | 0.0695 | | |

algorithm cannot converge. It is possible because the category of serious injuries only has a relatively small number of observations (469 observations) compared to its reference level no injuries, which has 10,907 observations.

Table 5.5: Estimation results based on the proposed baseline-category model (serious injuries)-final model

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -3.0224 | 0.2202 | 0.0000 |
| Insurance Product-Compulsory Motorcycle | 1.3466 | 0.3122 | 0.0000 |
| Insurance Product-Compulsory Motor Vehicle | 0.4583 | 0.1070 | 0.0000 |
| Purpose-Commercial | 0.9942 | 0.3435 | 0.0038 |
| Insured Value | -0.1156 | 0.0612 | 0.0587 |
| Seats | 0.1028 | 0.0351 | 0.0034 |
| $\sigma^2$ | 0.0080 | | |
| $\tau^2$ | 0.0697 | | |

Specifically, the positive sign of the estimate for Purpose-Commercial indicates that vehicles for commercial purposes are more likely to be related with serious injuries in comparison to the non-commercial used ones. Based on estimation results from Table 5.5, Fig. 5.5 and Fig. 5.6 demonstrate how the average risk of major injuries changes according to Seats and Insured Value. In both figures, blue and black lines denote compulsory motor vehicle insurance and comprehensive motor vehicle insurance (reference level), respectively. In contrast, solid and dashed lines represent commercial and non-commercial vehicles, respectively. The solid blue lines at the top in both Fig. 5.5 and Fig. 5.6 denote that when a commercial-usage vehicle has compulsory vehicle insurance, it tends to have a higher average risk of major injuries in comparison to the other situations listed here. Consistent with what is discussed in previous sections, the number of seats in a vehicle and insured value are positively and negatively correlated with serious injuries, respectively.

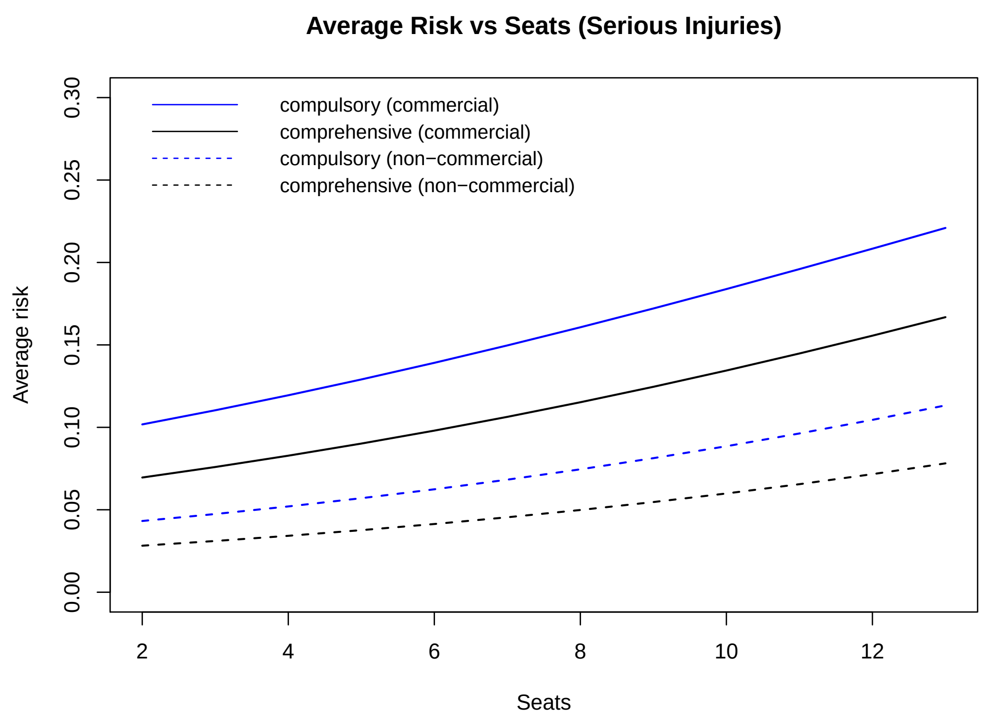


Figure 5.5: Average risk of serious injuries vs Seats

#### 5.4.2.2 Interpretation of random effects

Scatter plots in Fig. 5.7 and Fig. 5.8 display predicted random effects of Agent and Location for the serious subset. Additionally, Fig. 5.9 shows a corresponding map with spatial coordinates. The darker the blue, the higher the predicted values.

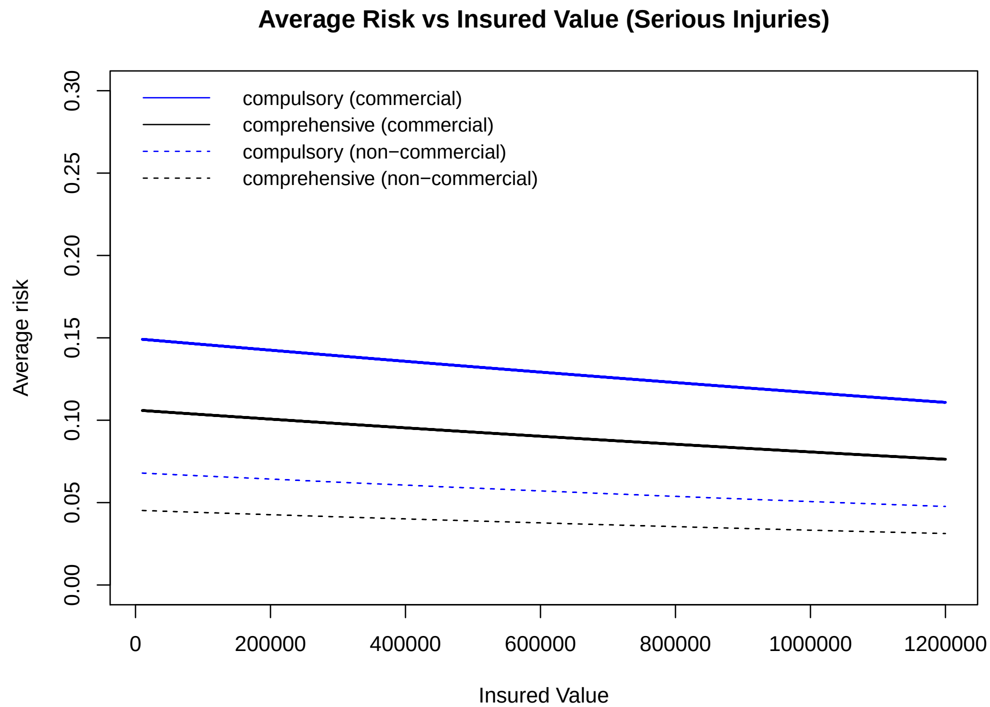


Figure 5.6: Average risk of serious injuries vs Insured Value

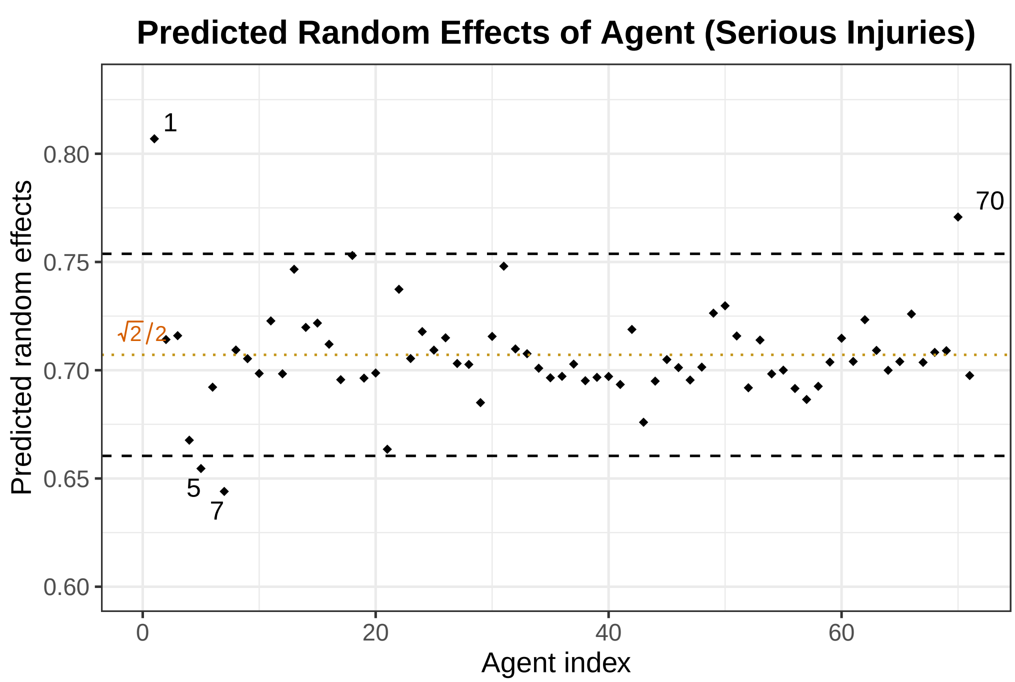


Figure 5.7: Predicted random effects of agent (serious injuries)

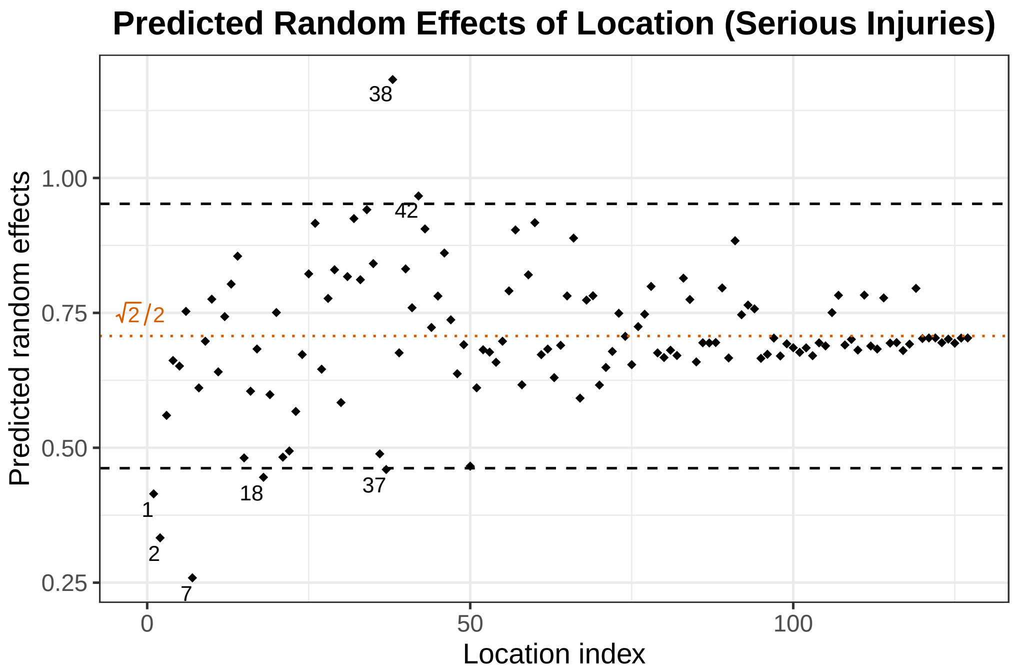


Figure 5.8: Predicted random effects of location (serious injuries)

Again, as shown in Fig. 5.7, holding all covariates constant, vehicles with insurance from agents 1 and 70 are associated with a higher relative risk of serious injuries, and agents 5 and 7 are associated with a lower risk. Similarly, Fig. 5.8 suggests that collisions in locations 38 and 42, particularly 38, have a higher relative risk of serious injuries due to motor vehicle accidents, whereas locations 1, 2, 7, 18 and 37, particularly 7, are less likely so. It is noticeable that several location points are close to the orange centerline in Fig. 5.8, indicating low or zero variance, which is due to the average number of motor vehicle accidents in these locations being around 2. In addition, we observe that the predicted value for location 38 is greater than 1. It is because that the BLUP approach adopted in the model to predict the random effects does not have the constraints that can remove those extreme values, so it occasionally happens that the predicted values are beyond the range $(0, 1)$.

For those possible interpretations for the likelihood of serious injuries regarding the map plot in Fig. 5.9, we find that the mountainous and mountainous alpine areas account for 87.5% of the total area of location 38. In addition, as stated in a gov-

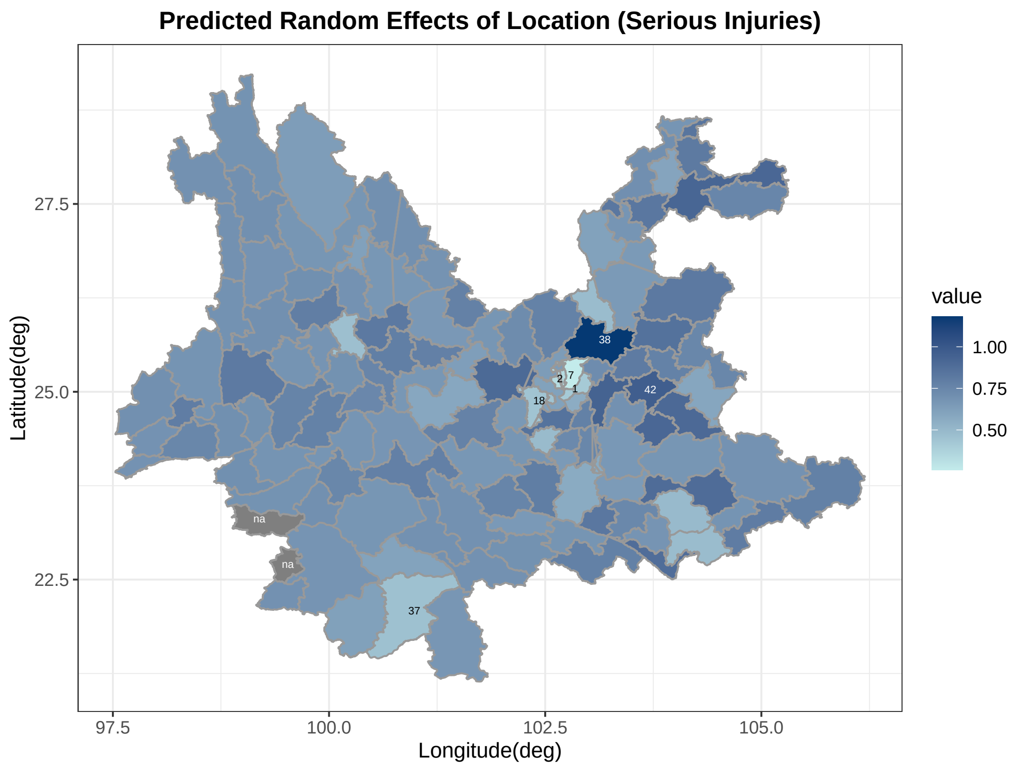


Figure 5.9: Predicted random effects of location with spatial coordinates (serious injuries)

ernment report (Luliang-County, 2021), one typical phenomenon of location 42 is that large and medium-sized freight vehicles go through the city's center. Therefore, these traffic conditions might contribute to the higher relative risk of serious injuries in locations 38 and 42.

Regarding the lower location-specific relative risk of serious injuries, locations 1, 2, 7, 18 and 37 also have a lower relative risk of injury, as discussed in Chapter 4. Since serious injury cases are only part of the injury cases, possible reasons for a lower relative risk of injury could also explain the lower relative risk of serious injuries. That is, locations 1, 2, 7, and 18 are highly developed areas with well-established traffic infrastructures, and location 37 is a tourist area with likely well-established traffic circulations.

### 5.4.3 Death Case Subset

Table 5.6: Estimation results based on the proposed baseline-category model (death cases)

| Parameter | Estimate | SE | P-value |
|---|---|---|---|
| Intercept | -5.5387 | 0.3219 | 0.0000 |
| Insured Value | -0.5276 | 0.4330 | 0.2230 |
| Peak-evening | 0.2079 | 0.4974 | 0.6759 |
| Peak-morning | -0.6705 | 0.7145 | 0.3481 |
| $\sigma^2$ | 0.0238 | | |
| $\tau^2$ | 0.1925 | | |

The death data only include observations from Death and No categories. Although 10,907 observations fall in the No baseline category, only 19 are in the Death category. We know that 19 observations are insufficient for a model with ten variables, as shown in Table 5.4, including those dummy variables. Furthermore, a close look at the 19 observations shows very weak correlations among specific covariates and the death cases. Therefore, taking the crossed random effects Agent and Location into consideration, we tentatively choose two variables– Insured Value and Peak–to fit the baseline-category model, the estimation results shown in Table 5.6. As expected, those two variables are not significant, indicating that they are not statistically significantly different from 0. Therefore, we do not provide any further discussion over the death case data.

### 5.4.4 Comparison analysis

Table 5.7: Comparison of the estimation results based on the proposed baseline-category models

| Parameter | Minor | | | Serious | | |
|---|---|---|---|---|---|---|
| | Estimate | SE | P-value | Estimate | SE | P-value |
| Intercept | -1.9946 | 0.0802 | 0.0000 | -3.0224 | 0.2202 | 0.0000 |
| IPC Motorcycle | 1.0151 | 0.2399 | 0.0000 | 1.3466 | 0.3122 | 0.0000 |
| IPC Motor Vehicle | 0.3317 | 0.0921 | 0.0003 | 0.4583 | 0.1070 | 0.0000 |
| Purpose-C | —— | —— | —— | 0.9942 | 0.3435 | 0.0038 |
| Insured Value | -0.0997 | 0.0478 | 0.0372 | -0.1156 | 0.0612 | 0.0587 |
| Seats | —— | —— | —— | 0.1028 | 0.0351 | 0.0034 |
| Peak-evening | 0.3312 | 0.0916 | 0.0003 | —— | —— | —— |

IPC = Insurance Product-Compulsory; Purpose-C = Purpose-Commercial.

Table 5.7 illustrates the estimation results for the final models for minor injury and serious injury subsets, where — in the table denotes those variables not in the final models. All variables in the final models are significant at the level of 0.05, except for Insured Value for serious injuries with a p-value of 0.0587. Strictly speaking, the similarity between minor and serious injuries in terms of significance only lies in the two dummy variables, Insurance Product-Compulsory Motorcycle and Compulsory Motor Vehicle. The estimates for these two dummies for serious injuries are slightly higher than those for minor injuries.

In terms of difference, the estimate for Purpose-Commercial dummy for serious injuries is significant but not for minor injuries. It is understandable that commercial usage vehicles could be used to transport goods or passengers to other cities or for other types of delivery, possibly including cumbersome loads. Therefore, to some extent, serious injuries could be more likely to happen in these cases due to traffic collisions. Similarly, the estimate for Seats is only significant for serious injuries, which suggests that the more seats a vehicle has, the more likely serious injuries are to occur, holding all the other covariates constant.

In contrast, the estimate for Peak-evening dummy is significant for minor injuries but not for serious injuries. One common scenario in peak evening hours is that people tend to feel tired and impatient when driving during these times, but people do not necessarily drive at high speeds because of the speed limit in the city, traic jams, low visibility, etc. This may explain that during peak evening hours, there is a higher chance of minor injuries but relatively fewer serious injuries.

# Chapter 6

# Discussion

## 6.1 Conclusion

In this thesis, by adopting the orthodox best linear unbiased predictor (BLUP) approach, we have proposed a binary logistic model with partially crossed random effects and its extension for multinomial analysis, namely, the baseline-category logit model with partially crossed random effects.

The proposed models are characterized with some advantages as follows. First, our models are robust against misspecification of random effects distribution since our assumptions on random effects are only based on the first two moments. Second, they are, in general, computationally efficient since they do not need to perform high-dimensional integration. Third, we have developed the estimation algorithm, which performs well since our simulation studies have demonstrated that the bias of the parameter estimators is pretty small, and the estimated and simulated standard errors are very close.

We applied our models to the insurance data about motor vehicle accidents. The data analysis results can help understand how covariates and random effects influence the outcomes. For example, it indicates that there is a statistically significant relationship between the occurrence of injuries and covariates, such as different insurance products, insured value, number of seats in a motor vehicle, and age of a motor vehicle, but not for the covariates, such as the purpose of a motor vehicle (commercial or non-commercial) and good driver (yes/not). As a result, the insurance company might need to adjust insurance premiums based on different criteria. Furthermore, after the effects of the covariates are incorporated in the model, the data analysis shows that specific locations have a higher probability of injury in motor vehicle accidents. Therefore, for these locations with high traffic injury risks, there might be ways to help people increase their awareness of the surroundings to reduce the chance of injury. Similarly, motor vehicles with insurance purchased from specific agents are more prone to motor vehicle accident injuries. Therefore, it may well be beneficial to investigate the reasons behind it.

## 6.2 Further Study

There are several research directions that we may extend to in the future.

First, our work is for cross-classified data with two crossed factors. Thus, we could consider extending the models for data with more complicated data structures, such as data with both nested and crossed random effects. For example, Stitch Fix is an online personal styling service that the company sends a sample of clothing items to clients, and clients may purchase any of those items and return the others (Ghosh et al., 2021). Our binary response is whether a client considers an item a top-rated fit (yes or no). Those fixed effects are client fit profile, client dress size, client chest size, clothing material, etc. For example, assume we have N= 20,000 ratings from 500 clients on 2000 items and 80 salespersons who are only responsible for specific clothing items. Then clients $U_i$ and items $V_j$ can be treated as crossed random effects, and items $V_j$ are nested within salespersons $S_t$. One possible way of considering the model is to add one more level; for crossed structure, each response $Y$ is determined by crossed random effects $U_i$ and $V_j$, and now is by $U_i$, $V_j$ and $S_t$ instead. However, the covariance matrix could be much more complicated, and the computational cost can be more challenging.

In addition, we know that one driver may well be involved in traffic collisions more than once over the years. For example, the dataset for our model study is about motor vehicle accidents in 2017, where around 12% of vehicles are involved in multiple accidents. Therefore, another direction we may extend our work is to model longitudinal data with crossed random effects.

Furthermore, future work needs to be done to reduce the convergence time. For example, when the dataset is relatively large, our current algorithm would be slow to invert huge matrices.

# Bibliography


Agresti, A. (2003). *Categorical data analysis*, Volume 482. John Wiley & Sons.

Bakbergenuly, I. and E. Kulinskaya (2018). Meta-analysis of binary outcomes via generalized linear mixed models: a simulation study. *BMC medical research methodology 18*(1), 1–18.

Bates, D. M. (2010). *lme4: Mixed-effects modeling with R*. Springer New York.

Bologa, C. G., V. S. Pankratz, M. L. Unruh, M. E. Roumelioti, V. Shah, S. K. Shaffi, S. Arzhan, J. Cook, and C. Argyropoulos (2021). High performance implementation of the hierarchical likelihood for generalized linear mixed models: an application to estimate the potassium reference range in massive electronic health records datasets. *BMC Medical Research Methodology 21*(1), 1–24.

Breslow, N. E. and D. G. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association 88*(421), 9–25.

Browne, W. J., H. Goldstein, and J. Rasbash (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling 1*(2), 103–124.

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

Crowder, M. (1986). On consistency and inconsistency of estimating equations. *Econometric Theory 2*(3), 305–330.

Crowder, M. (1987). On linear and quadratic estimating functions. *Biometrika 74*(3), 591–597.

Cui, J. (2019). Tweedie generalized linear models with crossed random effects. Master's thesis, University of New Brunswick.

Davies, O. L. and P. L. Goldsmith (1972). Statistical methods in research and production.

Dean, C. B., M. D. Ugarte, and A. F. Militino (2004). Penalized quasi-likelihood with spatially correlated data. *Computational statistics & data analysis 45*(2), 235–248.

Dohoo, I. R., W. Martin, and H. E. Stryhn (2003). *Veterinary epidemiologic research*.

Forcina, A. and L. Franconi (1988). Regression analysis with the beta-binomial distribution. *Rivista di Statistica Applicata 21*(1).

Ghosh, S., T. Hastie, and A. B. Owen (2021). Scalable logistic regression with crossed random effects. *arXiv preprint arXiv:2105.13747*.

Ha, I. D. and Y. Lee (2005). Comparison of hierarchical likelihood versus orthodox best linear unbiased predictor approaches for frailty models. *Biometrika 92*(3), 717–723.

Jeon, M., F. Rijmen, and S. Rabe-Hesketh (2017). A variational maximization–maximization algorithm for generalized linear mixed models with crossed random effects. *Psychometrika 82*(3), 693–716.

Jrgensen, B., S. Lundbye-Christensen, X.-K. Song, and L. Sun (1996). A longitudinal study of emergency room visits and air pollution for prince george, british columbia. *Statistics in medicine 15*(7-9), 823–836.

Kim, S., Y. Jeong, and S. Hong (2021). The impact of ignoring a crossed factor in cross-classified multilevel modeling. *Frontiers in Psychology 12*, 567.

Leckie, G. (2019). Cross-classified multilevel models. *arXiv preprint arXiv:1907.02569*.

Lee, Y. and J. A. Nelder (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological) 58*(4), 619–656.

Lee, Y., L. Rönnegård, and M. Noh (2017). *Data analysis using hierarchical generalized linear models with R.* Chapman and Hall/CRC.

Lindstrom, M. J. and D. M. Bates (1988). Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association 83*(404), 1014–1022.

Liu, Q. and D. A. Pierce (1994). A note on Gauss–Hermite quadrature. *Biometrika 81*(3), 624–629.

Luliang-County (2021). Announcement on the implementation of restricted traffic diversion of freight vehicles in the central urban area of luliang county. http://www.luliang.gov.cn/pub/description/24535.html.

Luo, W. and O. Kwok (2009). The impacts of misspecifying cross-classified random effects models. *Multivariate Behavioral Research 44*, 182–212.

Ma, R. (1999). *An orthodox BLUP approach to generalized linear mixed models.* Ph. D. thesis, University of British Columbia.

Ma, R. and B. Jørgensen (2007). Nested generalized linear mixed models: an orthodox best linear unbiased predictor approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(4), 625–641.

McCullagh, P. and J. Nelder (1989). *Generalized Linear Models*. Chapman and Hall, London.

McCulloch, C. E. and S. R. Searle (2001). *Generalized, linear and mixed models*. Wiley.

Meyers, J. L. and S. N. Beretvas (2006). The impact of inappropriate modeling of cross-classified data structures. *Multivariate behavioral research 41*(4), 473–497.

Najera-Zuloaga, J., D.-J. Lee, and I. Arostegui (2018). Comparison of beta-binomial regression model approaches to analyze health-related quality of life data. *Statistical methods in medical research 27*(10), 2989–3009.

Pawitan, Y. (2001). *In all likelihood: statistical modelling and inference using likelihood*. Oxford University Press.

Pinheiro, J. and D. Bates (2000). *Mixed-effects models in S and S-PLUS*. Springer science & business media.

Pinheiro, J. C. and D. M. Bates (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of computational and Graphical Statistics 4*(1), 12–35.

Pinheiro, J. C. and E. C. Chao (2006). Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics 15*(1), 58–81.

Rich, J. L. (2018). Comparison of generalized linear mixed model estimation methods. Master's thesis, Montana State University.

Rodríguez, G. and N. Goldman (1995). An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 158*(1), 73–89.

STHDA (2018). Regression with categorical variables: Dummy coding essentials in R. http://www.sthda.com/english/articles/40-regression-analysis/163-regression-with-categorical-variables-dummy-coding-essentials-in-r/.

Tierney, L. and J. B. Kadane (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association 81*(393), 82–86.

Tuerlinckx, F., F. Rijmen, G. Verbeke, and P. De Boeck (2006). Statistical inference in generalized linear mixed models: A review. *British Journal of Mathematical and Statistical Psychology 59*(2), 225–255.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*(4), 791–795.

World-Health-Organization (2021). Road traffic injuries. https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries.

Ye, F. and L. Daniel (2017). The impact of inappropriate modeling of cross-classified data structures on random-slope models. *Journal of Modern Applied Statistical Methods 16*(2), 25.

Zeger, S. L. and M. R. Karim (1991). Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American statistical association 86*(413), 79–86.

# Vita

**Candidate's full name:**

Zizhe Zhang

**University attended:**

Master of Science, November 2021, University of New Brunswick, Fredericton, NB, Canada

Master of Digital Experience Innovation, October 2012, University of Waterloo, Waterloo, ON, Canada

Master of Arts, March 2009, Shanghai Jiao Tong University, Shanghai, China

Bachelor of Arts, June 2004, Shanghai Ocean University, Shanghai, China

**Publications:**

None

**Conference Presentations:**

None