

Fall 12-18-2020

Multi-level small area estimation based on calibrated hierarchical likelihood approach through bias correction with applications to COVID-19 data

Nirosha Rathnayake
University of Nebraska Medical Center

Tell us how you used this information in this [short survey](#).

Follow this and additional works at: <https://digitalcommons.unmc.edu/etd>

 Part of the [Biostatistics Commons](#), [Multivariate Analysis Commons](#), [Partial Differential Equations Commons](#), [Statistical Methodology Commons](#), [Statistical Models Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Rathnayake, Nirosha, "Multi-level small area estimation based on calibrated hierarchical likelihood approach through bias correction with applications to COVID-19 data" (2020). *Theses & Dissertations*. 489.

<https://digitalcommons.unmc.edu/etd/489>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@UNMC. It has been accepted for inclusion in Theses & Dissertations by an authorized administrator of DigitalCommons@UNMC. For more information, please contact digitalcommons@unmc.edu.

**Multi-Level Small Area Estimation Based on
Calibrated Hierarchical Likelihood Approach
Through Bias Correction with Applications to
COVID-19 Data**

by

Nirosha Rathnayake

A DISSERTATION

Presented to the Faculty of
the University of Nebraska Graduate College
in Partial Fulfilment of the Requirements
for the Degree of Doctor of Philosophy

Biostatistics

Under the Supervision of Professor Hongying (Daisy) Dai

University of Nebraska Medical Center
Omaha, Nebraska

October 06, 2020

Supervisory Committee:

Jane Meza, Ph.D.

Kendra Schmid, Ph.D.

Steven From, Ph.D.

Acknowledgments

Throughout my academic period at the University of Nebraska Medical Center, I had received excellent advice, support, and encouragement from many people. Mostly, I would like to express my boundless gratefulness to my supervisor Dr. Hongying Dai for helping me with formulating the research topic and providing guidance throughout. She has presented me with profound recommendations, wise thinking, and motivating advice during the entire process, which was helpful to realize my potential. I genuinely appreciate her continuous effort and the enormous amount of time devoted to me to succeed in this work.

Additionally, I immensely thankful to Dr. Jane Meza for the thoughtful guidance and encouragement provided since the beginning of my Ph.D. and for providing feedback and suggestions during my dissertation. Also, my special thanks to Dr. Kendra Schmid for the continued advice and input to complete the dissertation successfully. Moreover, I admirably acknowledge Dr. Steven From, who always guided and motivated me for the last ten years, even during my master's thesis. I am so grateful to Dr. From for suggesting me to pursue my Ph.D. in Biostatistics, which indeed was a great opportunity in my life. I would also like to provide my generous gratitude to Dr. Jiangtao Luo for supervising and delivering me valuable thoughts during the first half of my dissertation. It is my pleasure to thank Dr. Gleb Haynatzki for providing me suggestions and advice to make better decisions during my academic period at UNMC.

Moreover, I would not have completed this dissertation without my husband's tremendous support, boundless encouragement, and enormous sacrifice throughout my Ph.D. program. He provided me a pleasant atmosphere at home, taking care of our two children, who also gave great support, cheering me up, and standing by me during tough times. I am genuinely grateful to my beloved husband, Aruna Weerakoon, 8-year-old son Nesh Weerakoon, and 12-year-old daughter Chanu Weerakoon for their unfaltering patience.

I am very thankful to my only brother, Jagath Rathnayake, who realized my strength and became my greatest mentor providing me valued guidance since my childhood. I wouldn't be who I am today without his generous encouragement and unwavering effort until I stepped into a successful pathway. Also, I am incredibly grateful to my parents for their great sacrifice to offer us a path for a successful future, even though it was a challenging journey for them being farmers in Sri Lanka. Our father taught us many valuable lessons, which are incredibly helpful throughout my life. My special thanks to my parents and my two sisters for their unconditional love and support.

I would also like to thank Ms. Mary Morris, especially for being very helpful and supportive in many ways during the past years. Additionally, I acknowledge the Holland Computing Center of the University of Nebraska at Lincoln (HCC-UNL) for providing computing resources to conduct the simulation studies. Finally, but certainly, not least, I am thankful to all the staff members and friends in the Biostatistics department at UNMC for being helpful in many ways.

Abstract

Multi-Level Small Area Estimation Based on Calibrated Hierarchical Likelihood Approach Through Bias Correction with Applications to COVID-19 Data

Nirosha Rathnayake, Ph.D.

University of Nebraska Medical Center, 2020

Supervisor: Hongying (Daisy) Dai, Ph.D.

Small area estimation (SAE) has been widely used in a variety of applications to draw estimates in geographic domains represented as a metropolitan area, district, county, or state. The direct estimation methods provide accurate estimates when the sample size of study participants within each area unit is sufficiently large, but it might not always be realistic to have large sample sizes of study participants when considering small geographical regions. Meanwhile, high dimensional socio-ecological data exist at the community level, which provides an opportunity for model-based estimation to incorporate rich auxiliary information at the individual and area levels. Thus, it is critical to developing advanced statistical modeling to extract accurate information.

Most existing methods of maximum likelihood estimation include complicated and computationally expensive integral approximations. Some require prior assumptions for the unobserved random effects. In this dissertation, we proposed a Calibrated Hierarchical (CH) likelihood approach, which does not involve such integral approximations. This work covered three aims:

Aim 1. We developed a novel modeling approach for SAE via hierarchical generalized linear models based on the CH likelihood with improved parameter estimations through bias correction (CHBC). Unified analysis through the h -likelihood provides flexibility in statistical inferences for unobserved random variables. And it leads to a single algorithm, expressed as a set of interlinked

and augmented generalized linear models, to be used for fitting a broad class of new models with random effects.

Aim 2. We then extended this methodology to the joint modeling of multiple outcome variables through shared random effects and multivariate random effects. The joint modeling approach has the flexibility of extending to multidimensional models using different types of outcomes by considering the association among them.

Aim 3. Extensive simulation studies were carried out to assess the empirical performance of estimation accuracy at varying scenarios. We also used COVID-19 data to study the association between confirmed cases and the number of deaths based on the multivariate joint modeling approach. Joint modeling through shared random effects is illustrated using the Youth Risk Behavior Surveillance System (YRBSS) data to assess the impact of tobacco consumption at the county-level. The asymptotic properties of MHLEs were studied. Last, we developed an R package for SAE modeling for the CHBC approach. The development version of the R package is available at <https://niroshar.github.io/hglmbc2/>.

Table of Contents

Acknowledgments.....	i
Abstract.....	iii
Table of Contents.....	v
List of Tables	ix
List of Figures.....	xi
List of Abbreviations	xiv
List of Notations	xv
Chapter 1. Introduction to Small Area Estimation.....	1
1.1 Background.....	1
1.2 Direct Estimation Method.....	4
1.3 Indirect Estimation Method	5
1.3.1 Area Level Model	6
1.3.2 Basic Unit Level Model.....	8
1.4 Best Linear Unbiased Prediction (BLUP) and Empirical BLUP (EBLUP).....	8
1.5 Model Comparison.....	9
Chapter 2. Advanced Modeling in SAE.....	10
2.1 Generalized Linear Mixed Model (GLMM).....	10
2.1 BLUP and EBULP Estimators.....	13
2.2 ML and REML Estimators.....	15
2.3 Empirical and Hierarchical Bayes Method for Small Area Models.....	17
2.4 Hierarchical Generalized Linear Models for SAE Models	19
2.4.1 Hierarchical Generalized Linear Models	19
2.4.2 H-likelihood and MHLE of (β, u)	20
2.4.3 Penalized Partial Maximum Likelihood Estimation of Dispersion Parameters	22
2.5. Multilevel Small Area Estimation in Survey Research Using GLMM.....	25

2.5.1 Emerging Tobacco Use Among Adolescents	26
2.5.2 Survey Materials and Methods	27
.....	33
2.6 Literature Review on Advanced Modelling Techniques in Small Area Estimation	33
Chapter 3. Small Area Estimation Using Calibrated Hierarchical Likelihood Approach	39
3.1 CH with Bias Correction for Generalized Family Distributions.....	39
3.1.1 Binomial-Normal HGLM	42
3.1.2 Poisson-Normal HGLM.....	43
3.2 Maximum Hierarchical Likelihood Estimates of (β, u)	44
3.3 Bias Correction	47
3.4 The MHLEs of Dispersion Parameters ϑ	49
3.5 Asymptotic Properties of MHLEs	52
3.5.1 Asymptotic Efficiency of β	52
3.5.2 Asymptotic Properties of MHLE of u	53
3.5.3 Wald Confidence Interval	56
Chapter 4. Joint Modeling of Multiple Outcomes in Small Area Estimation	57
4.1 Multivariate Joint Model and Hierarchical (\mathbf{h})-Likelihood	58
4.1.2 Bias Correction in Multivariate Joint Model	64
4.1.3 Bivariate Joint Model and H -Likelihood Function.....	65
4.1.4 MHLE Algorithm – Multivariate Joint Model.....	67
4.2 Joint Modeling Through Shared Random Effects.....	68
4.2.1 Parameter Estimation of Fixed Effects and Random Effects.....	71
4.2.2 Parameter Estimation of Dispersion Parameters.....	75
Chapter 5. Simulation	79
5.1 Simulation – Binary HGLM	79
5.1.1 Data Generation	79

5.1.2 MC Simulation Results	80
5.2 Simulation – Poisson HGLM.....	83
5.2.1 Data Generation	83
5.2.2 Simulation Results	83
5.3 Simulation – Joint Model Through Multivariate Random Effects.....	86
5.3.1 Data Generation	86
5.3.2 Simulation Results	87
Chapter 6. Real Data Analysis	91
6.1 CHBC Approach to Tobacco Smoking Data	91
6.2 CHBC Approach to COVID-19 Data	98
6.2.1 Model Selection	109
6.3 Multivariate Joint Modeling Using COVID-19 Data	110
6.4 Joint Modeling Through Shared Random Effects Using Tobacco Smoking Data	123
Chapter 7. Discussion, Limitations, and Future Work.....	128
7.1 Discussion.....	128
7.2 Limitations	131
7.3 Future Work.....	132
Appendix.....	145
Appendix A. Data Analysis Based on GLMM Using NYTS Tobacco Smoking Data.....	145
Appendix B. Benchmark Analysis Using Binomial HGLM and Poisson HGLM.....	146
Appendix C. Simulation Results for the Multivariate Joint Model Using Count Data.....	149
Appendix D. Benchmark Analysis for Mixed Logistic Model Using YRBSS Data	150
Appendix E. Benchmark Analysis for Poisson Mixed Model Using COVID-19 Data	154
Appendix F. The Elements of \mathcal{S} and \mathcal{J} of the Binomial-Normal HGLM	157
Appendix F.1 Partial Derivatives of Model Parameters for General Class of HGLM Family	157

Appendix F.2 Based on Score Function for Canonical GLM Family.....	160
Appendix G. R Package for CHBC Approach in SAE.....	162

List of Tables

Table No.	Description	Page No.
Table 2.1	<i>Summary Statistics of prevalence in 143 counties by the total, age group, sex, and race of NYTS 2014-2015 respondents.</i>	28
Table 2.2	<i>Model estimates using the FH model for ever-use and current-use of E-Cigarettes.</i>	30
Table 3.1	<i>HGLM models with components of some canonical GLM families.</i>	39
Table 6.1	<i>Model estimates for current-use and ever-use of E-Cigarettes based on the CHBC and GLMM.</i>	92
Table 6.2	<i>Model estimates for current-smoke and ever-smoke based on the CHBC and GLMM.</i>	93
Table 6.3	<i>COVID-19 data by regions in Utah State.</i>	99
Table 6.4	<i>MHLEs for fixed effects using the CHBC method for COVID-19 cases.</i>	105
Table 6.5	<i>Model selection criteria based on CHBC and GLMM.</i>	109
Table 6.6	<i>Maximum hierarchical likelihood estimates (MHLEs) of fixed effects based on the multivariate joint model for COVID-19 cases.</i>	111
Table 6.7	<i>Maximum hierarchical likelihood estimates (MHLEs) of fixed effects based on the multivariate joint model for COVID-19 deaths.</i>	113
Table 6.8	<i>Fixed effect estimates for current-use and ever-use of E-Cigarettes based on the joint model through shared random effects.</i>	124

<i>Table A.1</i>	<i>Ten US states with the highest and lowest estimated prevalence of current-use and ever-use of E-Cigarettes based on GLMM.</i>	145
<i>Table B.1</i>	<i>MHLEs of variance parameter from CHBC and GLMM for mixed logit model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).</i>	146
<i>Table B.2</i>	<i>MHLEs of variance parameter from CHBC and GLMM for Poisson mixed model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).</i>	147
<i>Table C.1</i>	<i>Estimates of variance parameters of random effects.</i>	149
<i>Table D.1</i>	<i>Summary statistics of current-use and ever-use of E-Cigarettes, current-smoke, and ever-smoke by gender, race, and age.</i>	150
<i>Table D.2</i>	<i>Ten US states with the highest and lowest estimated prevalence of current-use and ever-use of E-Cigarettes based on CHBC and GLMM.</i>	151
<i>Table D.3</i>	<i>Ten US states with the highest and lowest estimated prevalence of current-smoke and ever-smoke based on CHBC and GLMM.</i>	152
<i>Table E.1</i>	<i>MLEs for fixed effects and variance parameter using the GLMM for COVID-19 data.</i>	154

List of Figures

Figure	Description	Page No.
Figure 2.1	<i>Observed vs. estimated county prevalence of ever-use and current-use of E-Cigarettes in the US based on 2014-2015 NYTS data.</i>	31
Figure 2.2	<i>County prevalence of ever-use of E-Cigarettes in the US based on 2014-2015 NYTS data.</i>	32
Figure 2.3	<i>County prevalence of current-use of E-Cigarettes in the US based on 2014-2015 NYTS data.</i>	33
Figure 4.1	<i>Joint modelling of multiple outcomes through shared random effects.</i>	70
Figure 5.1	<i>MHLEs of fixed effects from CHBC and GLMM for mixed logit model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).</i>	82
Figure 5.2	<i>MHLEs of fixed effects from CHBC and GLMM for Poisson mixed model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).</i>	84
Figure 5.3	<i>Mean Squared Error (MSE) of the estimated random effect using CHBC for mixed logit and Poisson model.</i>	85
Figure 5.4	<i>MHLEs of fixed effects for \mathbf{y}_1 from multivariate joint model based on the CHBC for Poisson data ($m = 10, 20, 30, 50, n = 10, 30, 100, \text{ and } 500$).</i>	87
Figure 5.5	<i>MHLEs of fixed effects for \mathbf{y}_2 from multivariate joint model based on the CHBC for Poisson data ($m = 10, 20, 30, 50, n = 10, 30, 100, \text{ and } 500$).</i>	88

Figure 5.6	<i>Distribution of the estimated random effects in multivariate joint model.</i>	89
Figure 6.1	<i>Observed and estimated prevalence ever-use and current-use of E-Cigarettes, current-smoke and ever-smoke in the United States based on 2015-2017 YRBSS data.</i>	95
Figure 6.2	<i>(a) Total number of trips by month for trips less than 1 mile, 1-4 miles, 5-9 miles, 10-24 miles, 25-49 miles, 50-99 miles, and 100 miles or above and (b) Population not stay home (%) vs. COVID-19.</i>	100
Figure 6.3	<i>Top ten states with highest number of trips made in each category during March 01, 2020 – July 31, 2020.</i>	102
Figure 6.4	<i>Actual vs. estimated COVID-19 cases based on CHBC and GLMM at county-level and state-level.</i>	106
Figure 6.5	<i>Box plots for estimated random effects based on CHBC and GLMM.</i>	107
Figure 6.6	<i>Predicted density at the county-level based on CHBC and GLMM.</i>	108
Figure 6.7	<i>Estimated vs. actual COVID-19 cases at county-level based on joint CHBC model through multivariate random effects.</i>	116
Figure 6.8	<i>Distribution of estimated random effects based on joint CHBC model through multivariate random effects.</i>	116
Figure 6.9	<i>Estimated vs. actual COVID-19 at county-level based on joint CHBC model through multivariate random effects.</i>	118
Figure 6.10	<i>Estimated vs. actual COVID-19 cases at state-level based on multivariate joint CHBC model.</i>	119

<i>Figure 6.11</i>	<i>Estimated vs. actual COVID-19 deaths at state-level based on multivariate joint CHBC model.</i>	120
<i>Figure 6.12</i>	<i>Predicted density of COVID-19 at county-level based on multivariate joint CHBC model (per 10,000).</i>	120
<i>Figure 6.13</i>	<i>Case rate and death rate based on multivariate joint CHBC model (per 10,000).</i>	121
<i>Figure 6.14</i>	<i>Case rate and death rate for US states based on multivariate joint CHBC model (per 10,000).</i>	122
<i>Figure 6.15</i>	<i>Estimated vs. observed prevalence for current-use and ever-use of E-Cigarettes based on joint model through shared random effects.</i>	125
<i>Figure 6.16</i>	<i>Estimated county-level prevalence of current-use and ever-use of E-Cigarettes in the United States based on 2015-2017 YRBSS data based on the Joint CHBC method.</i>	126
<i>Figure B.1</i>	<i>Distribution of estimated random effects in Poisson model.</i>	148
<i>Figure D.1</i>	<i>County prevalence of ever-use and current-use of E-Cigarettes in the United States based on 2015-2017 YRBSS data.</i>	153
<i>Figure E.1</i>	<i>The estimated COVID-19 based on the CHBC and GLMM.</i>	155
<i>Figure E.2</i>	<i>Variance inflation factor (VIF) and standard error (SE) based on univariate mixed model for COVID-19 cases and deaths.</i>	156

List of Abbreviations

Term	Description
AIC	Akaike information criterion
ACS	American community survey
BIC	Bayesian information criterion
BLUP	Best linear unbiased predictor
CHBC	Calibrated hierarchical likelihood with bias correction
EB	Empirical Bayes
EBLUP	Empirical best linear unbiased predictor
EM	Expectation maximization
FH	Fay-Harriot
GHQ	Gaussian-Hermite quadrature
GLM	Generalized linear model
GLMM	Generalized linear mixed model
HB	Hierarchical Bayes
HGLM	Hierarchical generalized linear model
LMM	Linear mixed model
MCMC	Markov Chain Monte Carlo
MHLE	Maximum hierarchical (h)-likelihood estimate
MLE	Maximum likelihood estimate
MOM	Method of moments
MSE	Mean squares error
RCM	Regression calibration method
REML	Restricted maximum likelihood estimate
RMSE	Root mean squared error
SAE	Small area estimation

List of Notations

Term	Description
det	Determinant of a matrix
\mathbf{G}	Variance-covariance matrix of random effect \mathbf{u}
$g()$	Link function in GLM
h	Hierarchical log-likelihood
\mathbf{J}, \mathbf{I}	Fisher or observed information matrix
ℓ	Log-likelihood of the joint density
ℓ_R	Restricted log-likelihood of the joint density
m	Number of random effects (counties or small areas)
N	Total number of observations
n_i	Sample size of small area i (i.e., county i)
p	Number of parameters
\mathbf{R}	Variance-covariance matrix of residual ϵ
\mathbf{S}	Gradient vector (or score function)
$\mathbf{\Sigma}$	Variance-covariance matrix
T, t	Transpose of a matrix or a vector
\mathbf{u}, \mathbf{U}	Vector of area-specific random effects
\mathbf{V}	Variance-covariance matrix of \mathbf{y}
\mathbf{v}	Random component
\mathbf{X}	Design matrix for explanatory variables with fixed effects
\mathbf{y}	Response vector
\mathbf{Z}	Design matrix of explanatory variables with random effects
ϵ	Residual or the sampling error
$\boldsymbol{\beta}$	Vector of regression coefficients for fixed effects
$\tilde{\boldsymbol{\beta}}$	Weighted least squares estimator of $\boldsymbol{\beta}$
$\hat{\boldsymbol{\beta}}$	(Hierarchal) Maximum likelihood estimator of $\boldsymbol{\beta}$
$\hat{\boldsymbol{\delta}}$	(Hierarchal) Maximum likelihood estimator of $\boldsymbol{\delta}$
$\boldsymbol{\mu}$	Conditional mean of \mathbf{y} given \mathbf{u} in GLM

σ_e^2	Standard deviation of normally distributed sampling error
σ_u^2	Standard deviation of random effect
ϕ	Dispersion parameter in GLM
θ	Canonical (natural) parameter in GLM
$\tilde{\theta}_i$	Best linear unbiased prediction (BLUP) of θ_i
ψ, ϑ	Dispersion parameters

Chapter 1. Introduction to Small Area Estimation

1.1 Background

Small area estimation (SAE) has been widely used in recent decades to draw conclusions (estimates) on small areas, mostly in survey sampling studies. The term “small areas” refers to the small subpopulations of a larger sample, also known as small domains, clusters, small geographical areas, etc. The SAE techniques have been applied in many cases such as public health related studies, financial assessment, education planning, forest inventory studies, agricultural studies, government (employment and payroll) studies, etc. (Zahava Berkowitz et al., 2016a; Z. Berkowitz et al., 2019; Hongying Dai, Delwyn Catley, Kimber P Richter, Kathy Goggin, & Edward F Ellerbeck, 2018; R. E. Fay III & R. A. J. J. o. t. A. S. A. Herriot, 1979; Martuzzi & Elliott, 1996; Mauro, Monleon, Temesgen, & Ford, 2017; Nancy A Rigotti & Sara Kalkhoran, 2017; Yasui, Liu, Benach, & Winget, 2000)

In SAE, many researchers consider linear and generalized linear (nonlinear) mixed modeling approaches by introducing random effects to account for between-area variations. Various types of mixed models are developed to improve the accuracy of the global estimates in small areas (John NK Rao & Molina, 2015). The area (aggregated) level and unit (element) level models are the two main models under this scenario. The area-level models are used when the unit-level data for auxiliary variables are not available, instead aggregated data is available. Some studies are done considering both unit and area-level auxiliary variables (Zahava Berkowitz et al., 2016a; Jiang & Lahiri, 2006; Pfeffermann, 2013; Saei & Chambers, 2003).

The most common area-level model in SAE is the basic Fay Harriot (FH) model, which is an indirect estimation technique (R. E. Fay III & R. A. Herriot, 1979). The FH model is based on the direct survey estimates obtained from unit-level data and area-level auxiliary data. The basic FH model assumes that the sampling variance (σ_e^2) is known. However, this assumption might not

appropriate in many applications in SAE. Thus, many researchers have proposed extensions of the basic FH model to improve the direct survey estimates focusing on unknown sampling variance σ_e^2 , smoothing of σ_e^2 through generalized variance function (GVF) technique, or temporal and spatial correlation effects introduced in the spatial FH model (Dick, 1995; Yong You, 2008; Yong You & Chapman, 2006).

An adequate amount of research based on univariate analysis has been done in many cases, such as public health-related studies, financial assessment, education planning, forest inventory studies, agricultural studies, government (employment and payroll) studies, etc. (Zahava Berkowitz et al., 2016a; Z. Berkowitz et al., 2019; Hongying Dai et al., 2018; R. E. Fay III & R. A. J. J. o. t. A. S. A. Herriot, 1979; Martuzzi & Elliott, 1996; Mauro et al., 2017; Nancy A Rigotti & Sara Kalkhoran, 2017; Torabi & Rao, 2014; Yasui et al., 2000). However, it is essential to consider the association of multiple outcomes that occurs within the same individuals. Without ignoring this correlation between multiple outcomes and conducting univariate analysis for each outcome might not provide accurate estimates. For example, individuals with high obesity might also suffer from some other health conditions like high cholesterol, diabetes, etc. These events could be correlated and share similar auxiliary information. Additionally, it can also depend on the geographical location. Hence, considering the association between such multiple outcomes jointly adds more details to the model, which leads to better estimates. Thus, the multiple outcomes are to be modeled based on the multivariate analysis (Benavent & Morales, 2016; Burgard, Esteban, Morales, & Pérez, 2020; González-Manteiga, Lombardía, Molina, Morales, & Santamaría, 2008; Gueorguieva, 2001; Ubaidillah, Notodiputro, Kurnia, & Mangku, 2019).

Most advanced modeling techniques are based on the traditional likelihood method and the Bayesian approximation methods. Obtaining the MLEs in the likelihood-based method is very challenging when the joint likelihood function does not have a closed-form. In most cases, the marginal log-likelihood function does not have a closed-form; hence it cannot be evaluated analytically. It often involves computationally expensive integral approximations, which becomes

challenging in multivariate analysis. In this article, an alternative approach is considered based on the hierarchical (h)-likelihood to obtain the maximum hierarchical likelihood estimates (MHLEs) that are improved with the bias correction method. The proposed approach does not require any prior assumption for the latent random effects. The recent studies have shown that the h -likelihood is a statistically and numerically efficient algorithm which provides reliably better estimates for all the parameters in the model (Ha, Lee, & Song, 2001; Ha, Noh, & Lee, 2017; Lee, Jang, & Lee, 2011; Youngjo Lee & John A. Nelder, 1996; Lee, Nelder, & Pawitan, 2018; Lee, Ronnegard, & Noh, 2017). It also provides the flexibility of using the conditional distribution $u|y$ as a predictive density for random effect u to make predictions for unobserved random variables. The proposed method obtains the MHLEs through iterative approximation with bias correction of the estimates.

This work is organized as follows. Chapter 1 covers the introduction to SAE, including the different types of SAE techniques and model selection criteria. Section 1.2 briefly states the direct estimation technique in SAE, which is also known as the Horvitz-Thompson estimator. Section 1.3 covers the indirect estimation methods, namely, the most common area-level FH model and the basin unit-level model. Next, in section 1.4, BLUP and EBLUP of the basic area-level model are given. Model comparison using AIC and BIC are presented in section 1.5. In this paper, we will be using small areas and clusters interchangeably.

Chapter 2 discusses the advanced modeling techniques in SAE. Sections 2.1 and 2.2 covers GLMM with BLUP, EBLUP, ML, and REML estimates. Empirical and Hierarchical Bayes methods for small area models are discussed in section 2.3. Section 2.4 explains the HGLMs, parameter estimation of fixed effects, random effects, and dispersion parameters using the hierarchical likelihood. Next, section 2.5 includes an application based on GLMM discussing the multilevel small area estimation in survey research. Section 2.6 consists of the literature review on advanced modeling techniques and the limitations of existing methods describing the proposed method.

Chapter 3 describes the proposed Calibrated h -likelihood approach with bias correction (CHBC) that can be extended to the canonical GLM family distributions. The method illustrates using Binomial-Normal HGLM and Poisson-Normal HGLM. Section 3.2 discusses the maximum hierarchical likelihood estimates (MHLEs) of fixed effects and random effects using the CHBC method. The bias correction procedure is discussed in section 3.3, and the MHLEs of dispersion parameters are given in section 3.4. The subsections in 3.5 cover the asymptotic properties of MHLEs, and the Wald confidence interval of MHLES.

In chapter 4, the proposed CHBC method is extended to joint modeling of multiple outcomes based on two ways; 1) through multivariate random effects, explained in detail in subsections in 4.1, and 2) through shared random effects discussed briefly in section 4.2. Next, Chapter 5 contains the simulations studies performed to assess the empirical performance of the proposed CHBC approach in univariate and multivariate modeling in SAE using Binary HGLM in section 5.1, Poisson HGLM in 5.2, and Poisson multivariate HGLM in 5.3. In Chapter 6, we applied the CHBC method for each Binomial HGLM and Poisson HGLM in the univariate model and multivariate model. Section 6.1 discusses the CHBC method using a tobacco smoking data set from the Youth Risk Behavior Surveillance System (YRBSS) for univariate cases, and the data set is applied to joint modeling in section 6.4, discussing the importance of joint modeling in multiple outcomes. Next, a publicly available COVID-19 data set is used in the univariate case in section 6.2 and the multivariate model in section 6.3, discussing the importance of modeling jointly of the correlated outcomes. Finally, Chapter 7 covers the discussion and future areas of research.

1.2 Direct Estimation Method

Direct estimation technique or design based small area estimation provides precise estimates for small areas with sufficiently large sample sizes. Still, this method often cannot provide accurate estimates for areas with small sample sizes or no sampled units. The direct estimator of area means

\widehat{Y}_i when sampling without replacement is obtained using Horvitz-Thompson (H-T) estimator given as

$$\widehat{Y}_i = \frac{1}{n_i} \sum_{j \in s_i} \frac{Y_{ij}}{\pi_{ij}} = \frac{1}{n_i} \sum_{j \in s_i} w_{ij} Y_{ij},$$

where $i = 1, \dots, m$ small areas, Y_{ij} is j^{th} measurement from the sample set s_i in area i , n_i is the sample size for area i , $\pi_{ij} = 1/w_{ij}$ is the probability of selecting j^{th} unit from area i , and w_{ij} is sampling weight.

The direct estimation method is not appropriate in many problems, it provides large standard errors due to small sample size areas, and also it cannot be used to make conditional inferences unlike in frequentist and Bayesian approaches. Hence, indirect estimation or model-based methods are more prevalent in most research problems (R. E. Fay III & R. A. Herriot, 1979; John NK Rao & Molina, 2015).

1.3 Indirect Estimation Method

The direct estimate will provide inadmissibly significant standard errors with small areas as it does not increase the effective sample size in small areas. However, the indirect estimation method takes care of this issue by increasing the effective sample size, hence will reduce the standard error. The indirect estimation technique will bring information from related areas and/or time periods by linking them to the estimation process through a model; thus, it increases the sample size. Some of the indirect estimators are synthetic estimators, composite estimators, and the James-Stein estimators, which do not consider between area variation. The synthetic estimator is a direct estimate that is obtained based on several small areas in a large area, however, assuming that similar characteristics for both small areas and the large area, and used as an indirect estimator for the large area (Gonzalez & Hoza, 1978). The most common indirect estimation technique in SAE, the Fay-Herriot (FH) model developed by Robert E. Fay III and Roger A. Herriot in 1979, improves the

quality of direct estimates by borrowing strengths from the related areas, considering the between area variation. Hence it increases the accuracy of the estimates (John NK Rao & Molina, 2015).

1.3.1 Area Level Model

Area-level models are being used when the unit-level auxiliary data is not available. The most common area-level model is the Fay-Herriot (FH) model, which is also known as the linking model. Fay and Herriot developed the FH model by combining two parts: direct area-level estimates and the synthetic estimates obtained from the linearly related auxiliary model. The basic FH model is a particular case of the linear mixed model. Suppose that the population is divided into m areas (domains) with sample sizes n_1, \dots, n_i where $i = 1, \dots, m$, then the direct estimate for the variable of interest y_i for area i , \hat{y}_i is given as

$$\hat{y}_i = \theta_i + e_i, \quad e_i \sim N(0, \sigma_e^2) \quad i = 1, \dots, m,$$

where $\theta_i = \bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, $i = 1, \dots, m$ and the total sample size, $N = \sum_{i=1}^m n_i$.

The synthetic estimates or θ_i is related to area specific auxiliary data $\mathbf{X}_i = (X_1, \dots, X_p)$ through a linear model

$$\theta_i = \mathbf{X}_i^T \boldsymbol{\beta} + Z_i u_i, \quad i = 1, \dots, m \quad (1.1)$$

where p is the number of covariates and Z_i 's are known positive constants.

Fay Herriot (FH) Model

The FH model estimates balance the bias and precision of the results borrowing the strength of available information. The linking model assumes that the area-level true mean θ_i for the area, i is linearly related to area-level auxiliary variables (\mathbf{X}_i). The area-level means can be represented as a linear combination of fixed and random effects,

$$\theta_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i \quad (1.2)$$

These two parts build the basic area-level model (FH model) as follows

$$\hat{y}_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i + e_i, \quad (1.3)$$

$$i = 1, \dots, m, \quad u_i \sim N(0, \sigma_u^2), \quad e_i \sim N(0, \sigma_e^2),$$

where \mathbf{X}_i is a vector of covariates for area i , $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, u_i is area specific random effect for area i , $u_i \sim N(0, \sigma_u^2)$, assumed to be independent and identically distributed, with unknown σ_u^2 , and e_i is the sampling error for area i , $e_i \sim N(0, \sigma_e^2)$ with known σ_e^2 , $\mathbf{X}_i^T \boldsymbol{\beta}$ are the fixed effects. The estimates of $\boldsymbol{\beta}$ and σ_u^2 , denoted by $\widehat{\boldsymbol{\beta}}$ and $\widehat{\sigma_u^2}$, can be obtained using the method of moments (MOM), maximum likelihood (ML), or restricted ML (REML) techniques. However, for known σ_u^2 and unknown $\boldsymbol{\beta}$, the best-unbiased predictor for small area mean i , ($\eta_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i$) is given by the Best Linear Unbiased Prediction (BLUP) (Henderson, 1950).

The BLUP estimators depend on the variance-covariance of random effects σ_u^2 which can be estimated using ML or REML methods. If both $\boldsymbol{\beta}$ and σ_u^2 in the FH model are unknown, the estimate for the small area means is obtained replacing $\boldsymbol{\beta}$ by their estimators $\widehat{\boldsymbol{\beta}}$ and σ_u^2 by $\widehat{\sigma_u^2}$, that are called the Empirical Best Linear Unbiased Prediction (EBLUP) estimators. Similarly, empirical and hierarchical Bayes estimation methods can also provide accurate estimates for small area means (Ghosh & Rao, 1994; Molina & Marhuenda, 2015; John NK Rao & Molina, 2015; Saei & Chambers, 2003; Torabi & Rao, 2014).

Many studies have obtained the unit-level and area-level EBLUPs in SAE applications and have shown that the unit-level estimates are more accurate than the area-level estimates (Jiang & Lahiri, 2006; Mauro et al., 2017; Yun & Lee, 2004; Zhang et al., 2014). Similarly, the root means squared errors (RMSEs) of the area-level estimates are higher than that of the unit-level estimates. Furthermore, some studies have shown that RMSEs of the direct estimates are comparably larger than RMSEs of EBLUPs estimates (Mauro et al., 2017). The extensions of the standard FH model, such as the spatial FH model, deals when the auxiliary information is correlated, which is quite often in real applications. Some applications are included in section 2.6 under the literature review.

1.3.2 Basic Unit Level Model

Fuller and Battese (1973) first introduced the nested error linear regression, which is also known as the basic unit-level model (Battese & Fuller, 1981; Fuller & Battese, 1973). It assumes that the auxiliary data are available for each element j in each small area i , and also the variable of interest y_{ij} is related through the nested error linear regression model as follows

$$y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij},$$

$$i = 1, \dots, m, \quad j = 1, \dots, n_i, \quad u_i \sim N(0, \sigma_u^2), \quad e_{ij} \sim N(0, \sigma_e^2),$$

where $\mathbf{X}_{ij}^T = (X_{ij1}, \dots, X_{ijp})^T$ is auxiliary variables $k = 1, \dots, p$ for each small area i , $u_i \sim N(0, \sigma_u^2)$ is area-level random effects, and $e_{ij} \sim N(0, \sigma_e^2)$ is residual and independent of u_i .

1.4 Best Linear Unbiased Prediction (BLUP) and Empirical BLUP (EBLUP)

First, consider the basic area-level model described above, from the equation (1.3),

$$\hat{y}_i = \mathbf{X}_i^T \boldsymbol{\beta} + u_i + e_i, \quad i = 1, \dots, m, \quad u_i \sim N(0, \sigma_u^2), \quad e_i \sim N(0, \sigma_e^2).$$

Given that σ_u^2 and σ_e^2 are known, the Best Linear Unbiased Prediction (BLUP) (Henderson, 1950) of θ_i is

$$\tilde{\theta}_i = E(\eta|y) = \mathbf{X}_i^T \tilde{\boldsymbol{\beta}} + E(u_i|y) = \mathbf{X}_i^T \tilde{\boldsymbol{\beta}} + \tilde{u}_i,$$

where $\tilde{u}_i = \gamma_i(\hat{y}_i - \mathbf{X}_i^T \tilde{\boldsymbol{\beta}})$ is the predicted random effect for area i and $\gamma_i = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$,

$\gamma_i \in (0,1)$. The weighted least squares estimator of $\boldsymbol{\beta}$ is

$$\tilde{\boldsymbol{\beta}} = \left[\sum_{i=1}^m \frac{1}{\sigma_u^2 + \sigma_e^2} \mathbf{X}_i \mathbf{X}_i^T \right]^{-1} \sum_{i=1}^m \frac{1}{\sigma_u^2 + \sigma_e^2} \mathbf{X}_i \hat{y}_i.$$

When σ_u^2 is unknown, the EBLUP can be calculated by replacing σ_u^2 with a consistent estimator $\hat{\sigma}_u^2$. The EBLUP can be expressed as follows,

$$\tilde{\theta}_i^{EBLUP} = \mathbf{X}_i^T \tilde{\boldsymbol{\beta}} + \tilde{u}_i.$$

Substituting $\tilde{u}_i = \hat{\gamma}_i(\hat{y}_i - \mathbf{X}_i^T \tilde{\boldsymbol{\beta}})$, we have

$$\tilde{\theta}_i^{EBLUP} = \mathbf{X}_i^T \tilde{\boldsymbol{\beta}} + \hat{\gamma}_i (\hat{y}_i - \mathbf{X}_i^T \tilde{\boldsymbol{\beta}}) = \hat{\gamma}_i \hat{y}_i + (1 - \hat{\gamma}_i) \mathbf{X}_i^T \hat{\boldsymbol{\beta}}.$$

Both BLUP and EBLUP can be considered as a combination of direct estimators, \hat{y}_i , and regression-synthetic estimators, $\mathbf{X}_i^T \hat{\boldsymbol{\beta}}$. When the measurement error, σ_e^2 , is small as compared with the small area random effects, σ_u^2 , the direct estimator becomes reliable. Thus, the BLUP and EBLUP estimates stay closer to the direct estimator. In contrast, when the direct estimator is unreliable, the BLUP and EBLUP estimates get closer to the regression-synthetic estimator.

1.5 Model Comparison

The model parameters can be obtained using maximum likelihood (ML) and restricted maximum likelihood (REML) techniques. Hence, the model comparison and the selection are typically made based on standard goodness-of-fit measures, including the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). AIC selects the model that produces the closest distribution to the exact distribution through an asymptotic approximation of Kullback-Leibler information distance (KL divergence). BIC is very similar to AIC, except BIC is obtained from the Bayesian model comparison through large sample asymptotic approximation of the marginal likelihood (Akaike, 1973; Schwarz, 1978).

$$AIC = -2\ell(\mathbf{A}, \boldsymbol{\beta}) + 2(p + 1),$$

$$BIC = -2\ell(\mathbf{A}, \boldsymbol{\beta}) + (p + 1) \log N,$$

where p is the number of estimated parameters, $\ell(\mathbf{A}, \boldsymbol{\beta})$ is log-likelihood, N is the number of observations.

Both AIC and BIC proximity measures are a combination of goodness of fit and model complexity part in terms of the number of parameters and number of observations. It provides how much information is lost when we approximate one distribution with another distribution, hence lower AIC or BIC indicates a better model fit.

Chapter 2. Advanced Modeling in SAE

2.1 Generalized Linear Mixed Model (GLMM)

Linear Mixed Model (LMM) is a powerful and flexible methodology, which is an extension of linear models that allows fixed effects to explain the effect from auxiliary information and random effects to account for the between area variation simultaneously in SAE. The LMM models a variety of data types, including clustered data, repeated measures, multilevel/hierarchical data, and spatial data, which are linearly related to the mean of the normally distributed outcome variable of interest. However, in many situations, the observations (outcome variable of interest) are not normally distributed. Thus, the Generalized Linear Models (GLM) and GLMs with random effects (namely, GLMM) can be used to accommodate a broad class of distributions, including both continuous and categorical observations. In this scenario, the mean of the response variable is possibly not linearly related to the auxiliary information, i.e., $g(\mu_j) = g(E[y_j]) = \sum_{p=1}^p \beta_p X_{jp}, j = 1, \dots, n$ in the GLMs and $g(E[y_{ij}|u_i]) = \mathbf{X}_{ij}\boldsymbol{\beta} + Z_{ij}u_i, i = 1, \dots, m, j = 1, \dots, n_i, \boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T, \mathbf{X}_{ij} = (x_{ij1}, \dots, x_{ijp})$ in the GLMMs, where m is the number of random effects (similarly, the number of domains or small areas), and n_i is the sample size of i^{th} domain. The “mixed” in GLMM refers to the presence of fixed effects and random effects in the linear predictor $\mathbf{X}_{ij}\boldsymbol{\beta} + Z_{ij}u_i$. Here, we discuss the advanced modeling techniques in SAE, which includes both fixed and random effect components in the model, such as LMMs, GLMMs, and Bayesian modeling approaches.

In GLMMs, the mean outcome is linearly related to auxiliary data through a link function $g(\cdot)$ depending on the observation type (binary, continuous, or count), and the variance is a function of the mean. The GLMM models a variety of exponential family distributions such as Gaussian, binomial, Poisson, exponential, multinomial, etc. (McCulloch & Searle, 2004; Schwarz,

1978). The basic univariate unit-level models in small area estimation are the special cases of general linear mixed models (Datta & Ghosh, 1991). First, consider an LMM,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\epsilon}, \quad (2.1)$$

where $\boldsymbol{\epsilon}$ and \mathbf{u} are mutually independent with $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{R})$, $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$, the variance-covariance matrices, \mathbf{R} (dimension: $N \times N$), and \mathbf{G} ($m \times m$). The variance-covariance matrix \mathbf{R} is also known as the conditional covariance matrix of $\mathbf{y} | \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$. The dimensions of the other matrices and vectors are \mathbf{X} ($N \times p$), $\boldsymbol{\beta}$ ($p \times 1$), \mathbf{Z} ($N \times m$), \mathbf{u} ($m \times 1$), and $\boldsymbol{\epsilon}$ ($N \times 1$) respectively, where \mathbf{X} and \mathbf{Z} are known design matrices, \mathbf{y} ($N \times 1$) is a vector of outcome measures. The response vector \mathbf{y} is a linear combination of normally distributed random variables; thus, the marginal distribution has the form of $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R})$, where $\mathbf{V} = \mathbf{V}(\boldsymbol{\delta}) = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$, and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$ are variance parameters that covariance matrices \mathbf{G} and \mathbf{R} depend.

Now, consider binary outcome data, where the fixed and random effects in GLMM are linearly related to the mean of the outcome through the logit link, while count data is related through the log link. In recent years, GLMM has been widely considered in different ways with small area estimations (Jiang & Lahiri, 2006; John NK Rao & Molina, 2015). Suppose that the variable of interest in small area estimation is a binary outcome, then the logit of the probability of success for small areas i can be written as

$$\text{logit}(p_{ij}) = \text{logit}(P(y_{ij} | u_i = 1)) = \mathbf{X}_{ij}^T \boldsymbol{\beta} + u_i + e_{ij}, \quad (2.3)$$

where $i = 1, \dots, m$, $j = 1, \dots, n_i$, $u_i \sim N(0, \sigma_u^2)$, $e_{ij} \sim N(0, \sigma_e^2)$, p_{ij} is the probability of $y_{ij} = 1$ for the element j in area i , u_i is the random effect for area i , e_{ij} is the residual of element j in area i . The direct calculation shows that

$$\frac{p_{ij}}{1 - p_{ij}} = \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta} + u_i),$$

$$p_{ij} = \frac{\exp(\mathbf{X}_{ij}^T \boldsymbol{\beta} + u_i)}{1 + \exp(\mathbf{X}_{ij}^T \boldsymbol{\beta} + u_i)}.$$

As stated above, the GLMM considers both fixed and random effects to model continuous and discrete observations $y_j, j = 1, \dots, N$, which are neither independent nor normally distributed. Unlike in linear mixed models where $\mathbf{y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{R})$, the GLMM assumes that conditional-response $\mathbf{y}|\mathbf{u}$ is not normally distributed, only $\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$. In such cases, the estimation of the marginal distribution of \mathbf{y} is challenging, and the likelihood function does not have a closed-form, leading to computationally heavy evaluations,

$$f(\mathbf{y}) = \int f(\mathbf{y}, \mathbf{u}) d\mathbf{u} = \int \dots \int f(\mathbf{y}|\mathbf{u}) f(\mathbf{u}) du_1 \dots du_m \quad (2.2)$$

Existing methods to overcome this problem include the use of numerical methods to approximate the integrals of likelihood functions, which can be done using Gaussian-Hermite Quadrature (GHQ), and through Bayesian-based approaches. The GHQ approximates the integral of a polynomial (and non-polynomial) function by summing up the rectangular areas of the polynomial function (Crouch & Spiegelman, 1990). The approximated likelihood accuracy depends on the number of quadrature points (nodes); however, increasing quadrature points increases the complexity, and hence it will be computationally intensive.

Additionally, likelihood approximation through the Bayesian method is based on optimizing the conditional expectation. Dempster et al. (1977) proposed maximizing $\ell(\boldsymbol{\beta}; \mathbf{y})$ is similar to maximizing the conditional expectation as described below. Consider the log-likelihood function

$$\ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}, \mathbf{u}|\mathbf{y}) = \int L(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}, \mathbf{u}) d\mathbf{u} = \ell(\boldsymbol{\beta}; \mathbf{y}) + \log f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}),$$

taking expectation and the partial derivative with respect to $\boldsymbol{\beta}$

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\beta}} E[\ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}, \mathbf{u})|\mathbf{y}] &= \frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\beta}; \mathbf{y}) + E\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})\right] \\ E\left[\frac{\partial}{\partial \boldsymbol{\beta}} \log f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})\right] &= \int \frac{\partial}{\partial \boldsymbol{\beta}} \log f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}) f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}) d\mathbf{u} \end{aligned}$$

$$\begin{aligned}
&= \int \frac{\partial \log f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})}{\partial f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})} \frac{\partial f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})}{\partial \boldsymbol{\beta}} f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}) du \\
&= \int \frac{\partial f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})/\partial \boldsymbol{\beta}}{f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y})} f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}) du \\
&= \frac{1}{\partial \boldsymbol{\beta}} \int \partial f_{\boldsymbol{\beta}}(\mathbf{u}|\mathbf{y}) du = 0.
\end{aligned}$$

Now, the log-likelihood can be obtained through the EM algorithm, as described in Dempster et al. (1977). The E-step of the EM algorithm can be evaluated using the Monte Carlo EM algorithms, Gibbs sampling, and the GHQ (Crouch & Spiegelman, 1990; A. P. Dempster, Laird, & Rubin, 1977; Gelfand & Smith, 1990; Jiang & Lahiri, 2006; McCullagh, 2018; Vaida, Meng, & Xu, 2004). The M-step involves maximizing the $E[\ell(\boldsymbol{\beta}, \mathbf{u}; \mathbf{y}, \mathbf{u})|\mathbf{y}]$ until it converges to the log likelihood function $\ell(\boldsymbol{\beta}; \mathbf{y})$.

2.1 BLUP and EBULP Estimators

The best linear unbiased prediction (BLUP) minimizes the mean squared error among the linear unbiased estimators. It does not depend on the normality of the random effects but depends on the variance components. As stated in section 1.4, these BLUP parameters are estimated through the method of moments (MOM), ML, or REML (Hartley & Rao, 1967; Henderson, 1953; Patterson & Thompson, 1971). When the estimated variance components replace the BLUP estimator, it is referred to as the empirical BLUP (EBLUP) (Harville, 1991). Consider a linear combination in a form $\mu = \mathbf{l}^T \boldsymbol{\beta} + \mathbf{m}^T \mathbf{u}$, with regression parameters $\boldsymbol{\beta}$, random effects \mathbf{u} , and constant vectors \mathbf{l}, \mathbf{m} . A linear estimator $\hat{\mu}$ of $\mu = \mathbf{a}^T \mathbf{y} + b$ is unbiased if $E(\hat{\mu}) = E(\mu)$ (John NK Rao & Molina, 2015).

For a given $\boldsymbol{\delta}$, the BLUP estimator of $\mu = \mathbf{l}^T \boldsymbol{\beta} + \mathbf{m}^T \mathbf{u}$ can be derived as (Henderson, 1950)

$$\tilde{\mu} = \mathbf{l}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \tilde{\mathbf{u}} = \mathbf{l}^T \tilde{\boldsymbol{\beta}} + \mathbf{m}^T \mathbf{GZ}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \quad (2.4)$$

where

$$\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \quad (2.5)$$

is the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$, and

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}}(\boldsymbol{\delta}) = \mathbf{G}\mathbf{Z}^T\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}). \quad (2.6)$$

Consider the GLMM to illustrate the BLUP and EBLUP of $\boldsymbol{\beta}$ and \mathbf{u} . First, assume that \mathbf{u} and $\boldsymbol{\epsilon}$ follow multivariate normal distributions, then the joint density of \mathbf{y} and \mathbf{u} is

$$\begin{aligned} f(\mathbf{y}, \mathbf{u}) &= f(\mathbf{y}|\mathbf{u}) \times f(\mathbf{u}) \\ &= (2\pi)^{-\frac{N}{2}}(\det(\mathbf{R}))^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})\right] \\ &\quad \times (2\pi)^{-\frac{h}{2}}(\det(\mathbf{G}))^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}\right] \\ &\propto \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) - \frac{1}{2}\mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}\right]. \end{aligned}$$

Assume the variance parameters $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$ are known, maximizing the joint likelihood of \mathbf{y} and \mathbf{u} with respect to $\boldsymbol{\beta}$ and \mathbf{u} is equivalent to maximizing the joint log-likelihood

$$\ell(\boldsymbol{\beta}, \mathbf{u}) = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) - \frac{1}{2}\mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}. \quad (2.7)$$

Since \mathbf{u} is unobservable, $\ell(\boldsymbol{\beta}, \mathbf{u})$ can be considered as a penalized likelihood with a penalty term $\frac{1}{2}\mathbf{u}^T\mathbf{G}^{-1}\mathbf{u}$ added to the traditional log-likelihood function. The penalized log-likelihood function is conditioning on \mathbf{u} as fixed (known).

In (2.7), by setting the partial derivative of $\ell(\boldsymbol{\beta}, \mathbf{u})$ with respect to $\boldsymbol{\beta}$ and \mathbf{u} to be zero, we have the following mixed model equations for both fixed effects and random effects

$$\begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^T\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{u}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}^T\mathbf{R}^{-1}\mathbf{y} \end{bmatrix}. \quad (2.8)$$

The solution to (2.8) and BLUP estimators of $\boldsymbol{\beta}$ and \mathbf{u} are identical, i.e. $\boldsymbol{\beta}^* = \tilde{\boldsymbol{\beta}}$ and $\mathbf{u}^* = \tilde{\mathbf{u}}$. Thus the BLUP estimators can also be considered as joint maximum likelihood estimators.

When the variance parameters $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$ are unknown, the empirical BLUP (EBLUP) estimator of \mathbf{u} , $\hat{\mathbf{u}} = t(\hat{\boldsymbol{\delta}}, \mathbf{y}) = t(\hat{\boldsymbol{\delta}})$ is obtained by replacing the unknown variance parameters $\boldsymbol{\delta} = (\delta_1, \dots, \delta_q)^T$ by an estimator $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}(\mathbf{y})$.

2.2 ML and REML Estimators

Under the general linear mixed model, the maximum likelihood (ML) and restricted maximum likelihood (REML) estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are obtained by maximizing the log-likelihood function. Under the normality assumption, the log-likelihood is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\delta}) = -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2} \log \det(\mathbf{V}) + c, \quad (2.9)$$

where $c = -\frac{N}{2} \log(2\pi)$ is a constant, \mathbf{V} is the variance-covariance matrix of $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$.

Taking the partial derivative of $\ell(\boldsymbol{\beta}, \boldsymbol{\delta})$ with respect to $\boldsymbol{\beta}$, we have

$$\frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} - \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} \boldsymbol{\beta}.$$

Thus, the MLE of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}.$$

Taking the partial derivative of $\ell(\boldsymbol{\beta}, \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$, we have

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \delta_j} &= -\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \frac{\partial \mathbf{V}^{-1}}{\partial \delta_j}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2} \frac{\partial \log \det(\mathbf{V})}{\partial \delta_j} \\ &= \frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2} \text{trace} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \right), \end{aligned}$$

Note from Rao and Molina (2015, pp.102-103) (John NK Rao & Molina, 2015; Searle, Casella, & McCulloch, 2009)

$$\begin{aligned} \frac{\partial \mathbf{V}^{-1}}{\partial \delta_j} &= -\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \mathbf{V}^{-1}, \\ \frac{\partial |\mathbf{V}|}{\partial \delta_j} &= \text{trace} \left(\text{adj}(\mathbf{V}) \frac{\partial \mathbf{V}}{\partial \delta_j} \right), \end{aligned}$$

$$\frac{\partial \log \det(\mathbf{V})}{\partial \delta_j} = \text{trace} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \right).$$

The expected second derivative of $-\ell(\boldsymbol{\beta}, \boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ is given by

$$\begin{aligned} \mathbf{I}_{jk}(\boldsymbol{\delta}) &= E \left[-\frac{\partial^2 \ell(\boldsymbol{\beta}, \boldsymbol{\delta})}{\partial \delta_j \partial \delta_k} \right] \\ &= E \left[-\frac{\partial \left(\frac{1}{2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - \frac{1}{2} \text{trace} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \right) \right)}{\partial \delta_k} \right] \\ &= E \left[-\frac{\partial \left(-\frac{1}{2} \text{trace} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \right) \right)}{\partial \delta_k} \right] \\ &= \frac{1}{2} \text{trace} \left(\mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_j} \mathbf{V}^{-1} \frac{\partial \mathbf{V}}{\partial \delta_k} \right), \end{aligned}$$

where $j = 1, \dots, q$ and $k = 1, \dots, q$ are the elements of the Fisher Information matrix, $-\mathbf{I}_{jk}(\boldsymbol{\delta})$.

The MLEs $\hat{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ can be obtained iteratively by using the Newton Raphson algorithm based on the first and second order partial derivatives of the log-likelihood function with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$,

$$\boldsymbol{\delta}^{(k+1)} = \boldsymbol{\delta}^{(k)} + [\mathbf{I}(\boldsymbol{\delta}^{(k)})]^{-1} \frac{\partial}{\partial \boldsymbol{\delta}} \ell(\boldsymbol{\beta}, \boldsymbol{\delta}) \Big|_{\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}^{(k)}), \hat{\boldsymbol{\delta}} = \boldsymbol{\delta}^{(k)}},$$

where $k = 0, 1, \dots$, is the number of iterations.

The ML estimators of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ at convergence are $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\delta}})$ respectively, where $\hat{\boldsymbol{\delta}} = \boldsymbol{\delta}^{(k)}$ and $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta}^{(k)})$ are values of $\boldsymbol{\delta}$ and $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\boldsymbol{\delta})$ at the k^{th} iteration. The asymptotic covariance matrix of ML estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\delta}}$ is $\text{diag}[\text{var}(\hat{\boldsymbol{\beta}}), \text{var}(\hat{\boldsymbol{\delta}})]$, where

$$\begin{aligned} \text{var}(\hat{\boldsymbol{\beta}}) &= \text{var}[(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}] \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{V} ((\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1})^T \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{V} \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}, \end{aligned}$$

and $var(\hat{\boldsymbol{\delta}}) = \mathbf{I}^{-1}(\boldsymbol{\delta})$.

The ML method's major drawback is the loss of degrees of freedom of one parameter when estimating the other parameter. The REML takes care of this issue by transforming data $\mathbf{y}^* = \mathbf{A}^T \mathbf{y}$, where \mathbf{A} is any $N \times (N - p)$ full-rank matrix orthogonal to \mathbf{X} . The REML estimators are obtained through the restricted log-likelihood of the joint density of \mathbf{y}^* expressed as a function of $\boldsymbol{\delta}$

$$\ell_R(\boldsymbol{\delta}) = -\frac{1}{2} \log \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X} - \frac{1}{2} \mathbf{y}^T \mathbf{P} \mathbf{y} - \frac{1}{2} \log |\mathbf{V}| + c,$$

where $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1} \mathbf{X} (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1}$.

REML estimators $\hat{\boldsymbol{\delta}}_{RE}$ of $\boldsymbol{\delta}$ and $\hat{\boldsymbol{\beta}}_{RE} = \tilde{\boldsymbol{\beta}}(\hat{\boldsymbol{\delta}}_{RE})$ of $\boldsymbol{\beta}$ also are obtained iteratively using the Newton Raphson algorithm. The covariance matrices are asymptotically equal in both ML and REML estimators of $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ for fixed p , $var(\hat{\boldsymbol{\delta}}) \approx var(\hat{\boldsymbol{\delta}}_{RE})$ and $var(\hat{\boldsymbol{\beta}}) \approx var(\hat{\boldsymbol{\beta}}_{RE})$.

2.3 Empirical and Hierarchical Bayes Method for Small Area Models

Bayes techniques are based on the basic Bayes theorem using marginal (prior distribution $f(\boldsymbol{\theta})$) and conditional probability density functions (posterior distribution $f(\mathbf{y}|\boldsymbol{\theta})$). The marginal likelihood of \mathbf{y} , $f(\mathbf{y}) = \int f(\mathbf{y}, \boldsymbol{\theta}) d\boldsymbol{\theta} = \int f(\mathbf{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}$, which involves multiple integrals, becomes challenging with complicated posterior distributions and also with multivariate prior distributions. In such situations, parameters are obtained via simulation approaches, such as Markov Chain Monte Carlo (MCMC) approaches Metropolis-Hastings and Gibbs sampling. The Hierarchical Bayes (HB) method first simulates $\boldsymbol{\theta}$, then simulates \mathbf{y} given $\boldsymbol{\theta}$. The Empirical Bayes (EB) method is based on the convenient prior distribution for $f(\boldsymbol{\theta}|\boldsymbol{\beta})$ known as the conjugate prior, where $\boldsymbol{\beta}$ is a hyperparameter which is estimated by frequentist methods (Martuzzi & Elliott, 1996; John NK Rao & Molina, 2015; Yasui et al., 2000).

The Empirical Bayes approach in SAE assumes that the parameter of interest $\boldsymbol{\theta} = \boldsymbol{\mu}$, the population means for the small area i in the linking model of FH model (equation (1.1)), has some prior distribution $f(\boldsymbol{\theta}|\boldsymbol{\beta})$ where $\boldsymbol{\beta}$ is an unknown parameter. First, the posterior distribution of $\boldsymbol{\theta}$,

given data, is obtained assuming that β is known, then β is estimated using the marginal distribution of the data (Farrell, MacGibbon, & Tomberlin, 1997; R. E. Fay III & R. A. J. J. o. t. A. S. A. Herriot, 1979; Ghosh & Rao, 1994; Morris, 1983).

The Hierarchical Bayes model assumes that the parameter of interest θ is from a prior distribution with some unknown parameters (β), and again this unknown parameter has another prior distribution with unknown parameters. Therefore, in SAE, the primary area model under the hierarchical Bayesian framework can be expressed as a two-stage hierarchical model, which is also known as a conditionally independent hierarchical model. Two models have been discussed based on unknown sampling variance σ_e^2 and known sampling variance (unbiased estimate, s_e^2) in the basic FH model (Yong You & Chapman, 2006). The authors showed that the results obtained from the proposed HB method for the model with unknown variance perform well regardless of the sample size, by using two survey data sets, corn-soybean and milk data from the U.S. Department of Agriculture and U.S. Bureau of Labor Statistics, respectively.

The basic FH model (equation (1.1)) assumes that the direct estimate y_i and sampling variance σ_e^2 of $e_i \sim N(0, \sigma_e^2)$ are known and obtained from auxiliary data. This assumption could be too strong and lead to biased results, especially for the areas with small sample sizes. You and Chapman (2006) introduced a Hierarchical Bayes (HB) model using the Gibbs sampling technique to estimate unknown σ_e^2 from an unbiased estimator s_e^2 assuming that the s_e^2 is independent of the direct estimator y_i , and $(n_i - 1)s_e^2 \sim \sigma_e^2 \chi_{n_i-1}^2$, where n_i is the sample size for area i (Yong You & Chapman, 2006). Some applications of empirical and hierarchical Bayes methods in SAE are discussed in detail under the literature review in section 2.6 (Ghosh, Natarajan, Stroud, & Carlin, 1998; Hobza & Morales, 2016).

2.4 Hierarchical Generalized Linear Models for SAE Models

2.4.1 Hierarchical Generalized Linear Models

A hierarchical generalized linear model (HGLMs) uses a generalization of Henderson's joint likelihood to allow components of random effects in the linear predictors of generalized linear mixed models to have a conjugate distribution from the exponential family. For example, the distribution of random effect u_i for area i in Poisson-Gamma HGLM (or Poisson conjugate HGLM) is Poisson with mean λ , and the distribution of $\mathbf{y}|u_i$ is gamma with a canonical log link function. Some other examples of HGLMs are Binomial-Beta, Gamma-Inverse Gamma, and Inverse Gaussian-Gamma HGLMs. Normal-Normal or Normal conjugate HGLM is a specific case of HGLM with identity link function, where u_i is normally distributed, also known as the Linear Mixed Model (LMM) (Youngjo Lee & John A Nelder, 1996; Lee, Nelder, & Pawitan, 2006).

Lee and Nelder (1996) originally defined the HGLM as a GLM family for the response variable \mathbf{y} given a random effect \mathbf{u} , satisfying

$$E(\mathbf{y}|\mathbf{u}) = \boldsymbol{\mu} \text{ and } \text{var}(\mathbf{y}|\mathbf{u}) = \boldsymbol{\phi}V(\boldsymbol{\mu}),$$

with the conditional log-likelihood of \mathbf{y} given \mathbf{u}

$$\ell(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{y}|\mathbf{u}) = \frac{\sum\{\mathbf{y}\boldsymbol{\theta} - b(\boldsymbol{\theta})\}}{\boldsymbol{\phi}} + c(\mathbf{y}, \boldsymbol{\phi}), \quad (2.10)$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\mu})$ is the canonical (natural) parameter, $\boldsymbol{\mu}$ is the conditional mean of \mathbf{y} given \mathbf{u} , $\boldsymbol{\phi}$ is the dispersion parameter. The linear predictor takes the form of

$$\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v},$$

where $g(\cdot)$ is the link function, $\boldsymbol{\beta}$ s are the regression coefficients of fixed effects, and random component $\mathbf{v} = \mathbf{v}(\mathbf{u})$ is a monotonic function of random effects \mathbf{u} . The random effects \mathbf{u} is extended to conjugate distributions from the GLM family with parameters α . This will be a key advantage of HGLMs compared to GLMs.

The HGLM approach has been widely used in modeling binary and count data, frailty modeling for survival data, repeated measures data, and survey data in both univariate and multivariate cases (Ha et al., 2001; Molenberghs, Verbeke, Demétrio, & Vieira, 2010). Some examples of HGLMs are listed below.

Normal-Normal HGLM

A Normal-Normal HGLM is a GLM with $\mathbf{y}|\mathbf{u} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}, \sigma^2)$, $\mathbf{u} \sim N(0, \sigma_u^2)$ and the identity link function ($g(\boldsymbol{\mu}) = \boldsymbol{\mu}$), hence $\mathbf{v} = \mathbf{u}$, and the log-likelihood of $\mathbf{y}|\mathbf{u}$

$$\begin{aligned} \ell(\boldsymbol{\theta}, \phi; \mathbf{y}|\mathbf{u}) &= -\frac{1}{2\sigma^2}(\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}))^2 - \log 2\pi\sigma^2 \\ &= \left\{ \frac{\mathbf{y}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}) - (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v})^2/2}{\sigma^2} - \left(\frac{\mathbf{y}}{2\sigma^2} + \log 2\pi\sigma^2 \right) \right\}, \end{aligned}$$

has the form of (2.10), where $Var(\mathbf{y}|\mathbf{u}) = \phi = \sigma^2$, $\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}$, $b(\boldsymbol{\theta}) = (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v})^2$, $V(\boldsymbol{\mu}) = 1$, $\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{v}$, $\mathbf{v} = \mathbf{u}$, $\mathbf{u} \sim N(0, \sigma_u^2)$, and $c(\mathbf{y}, \phi) = -(\mathbf{y}/2\sigma^2 + \log 2\pi\sigma^2)$. Here, the random component \mathbf{v} has a normal distribution with the identity link function.

Poisson-Gamma HGLM

Poisson-Gamma HGLM is an extension of GLM with the random effect (\mathbf{u}) has a Gamma distribution, and $\mathbf{y}|\mathbf{u} \sim Poisson(\mathbf{X}\boldsymbol{\beta} + \mathbf{v})$. The log-likelihood of $\mathbf{y}|\mathbf{u}$

$$\ell(\boldsymbol{\theta}, \phi; \mathbf{y}|\mathbf{v}) = \mathbf{y} \log(\mathbf{X}\boldsymbol{\beta} + \mathbf{v}) - (\mathbf{X}\boldsymbol{\beta} + \mathbf{v}) - \log \mathbf{y}!$$

where $\boldsymbol{\theta} = \log(\mathbf{X}\boldsymbol{\beta} + \mathbf{v})$, $b(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$, $\phi = 1$, and $\mathbf{v} = \log \mathbf{u}$. The random component \mathbf{v} has the log-gamma distribution with link function being the log link and \mathbf{u} being gamma distribution in Poisson-Gamma HGLM. The expected value of $\mathbf{y}|\mathbf{u}$, $E(\mathbf{y}|\mathbf{u}) = b'(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\beta} + \log \mathbf{u}$.

2.4.2 H-likelihood and MHLE of $(\boldsymbol{\beta}, \mathbf{u})$

The h -likelihood is the logarithm of the joint density function of \mathbf{y} and $\mathbf{v} (= v(\mathbf{u}))$, equivalently, it is the joint density function of \mathbf{y} and \mathbf{u} , which needs to be the form of

$$h = \log f_{\boldsymbol{\beta}, \phi}(\mathbf{y}|\mathbf{u}) + \log f_{\alpha}(\mathbf{u}) = \ell(\boldsymbol{\theta}, \phi; \mathbf{y}|\mathbf{u}) + \ell(\alpha|\mathbf{u}),$$

where $\ell(\boldsymbol{\theta}, \phi; \mathbf{y}|\mathbf{u})$ is the logarithm of the density function of $\mathbf{y}|\mathbf{u}$, $\ell(\alpha|\mathbf{u})$ is the logarithm of the density function of \mathbf{u} , $\boldsymbol{\beta}$ are fixed effects, and (ϕ, α) are the dispersion parameters (Lee et al., 2018). The classical approach of estimating the BLUP of random effects \mathbf{u} and the MLE of fixed effects $\boldsymbol{\beta}$ is obtained by maximizing the joint density function, is straightforward in a standard linear mixed model with $\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 \mathbf{I})$ and $\mathbf{u} \sim N(0, \sigma_u^2)$ (Henderson, Kempthorne, Searle, & Von Krosigk, 1959). When the joint log-likelihood does not have a closed-form, it is very challenging to estimate BLUP or EBLUP. In such situations, h -likelihood plays an important role in model inference. Lee and Nelder (1996) derived the hierarchical(h) maximum likelihood estimates (HMLEs) using the score equations $\partial h / \partial \boldsymbol{\beta} = 0$, and $\partial h / \partial \mathbf{u} = 0$ for given ϕ and α . Unlike in GLM, the random component \mathbf{v} , or equivalently, the random effect \mathbf{u} in HGLM is not assumed to be normally distributed; instead, it is estimated by the properties of data through a prior distribution. However, the HMLEs are often obtained via numerical approximation methods due to intractable integrals in the log-likelihood function.

The inferences about $\boldsymbol{\beta}$, \mathbf{u} , and the dispersion parameters $\vartheta = (\phi, \alpha)$ in HGLMs involve three likelihoods, the h -likelihood, and two adjusted profile likelihoods (marginal likelihood and restricted likelihood). Inference about $\boldsymbol{\beta}$ is based on the marginal likelihood $L = \log \int \exp(h) d\mathbf{u}$, inference about \mathbf{u} is based on the h -likelihood, and the inference about the dispersion parameters is based on restricted likelihood, respectively.

Let $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$ be the model parameters. As stated above, given the dispersion parameters $\vartheta = (\phi, \alpha)$, the maximum hierarchical (h)-likelihood estimators (MHLE) $\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$ for $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$ are obtained by solving the score function $\partial h / \partial \boldsymbol{\tau} = 0$. If the solution does not have a closed-form, we can use Newton-Raphson methods to generate an iterative procedure that uses the gradient vector and the observed information matrix to approximate the points that maximize a likelihood

function. Start at initial values $(\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}^{(0)}, \mathbf{u} = \widehat{\mathbf{u}}^{(0)})$, the approximate maximums are updated iteratively using

$$\begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(k+1)} \\ \widehat{\mathbf{u}}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \widehat{\boldsymbol{\beta}}^{(k)} \\ \widehat{\mathbf{u}}^{(k)} \end{pmatrix} + (\mathbf{J}^{-1} \boldsymbol{\mathcal{S}}(\boldsymbol{\tau}))|_{(\boldsymbol{\beta}, \mathbf{u}) = (\widehat{\boldsymbol{\beta}}^{(k)}, \widehat{\mathbf{u}}^{(k)})}, \quad (2.11)$$

where $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$,

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}) = \begin{pmatrix} \frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \mathbf{u}} \end{pmatrix},$$

is the score function,

$$\mathbf{J} = \begin{pmatrix} -\left[\frac{\partial^2 h}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} \right]_{p \times p} & -\left[\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} \right]_{p \times m} \\ -\left[\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} \right]_{p \times p} & -\left[\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \mathbf{u}} \right]_{m \times m} \end{pmatrix},$$

is the asymptotic covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$. The variance-covariance matrix will be estimated by MHLE of ϑ of adjusted h -likelihood through an iterative procedure, which is described in detail in chapter 3.

2.4.3 Penalized Partial Maximum Likelihood Estimation of Dispersion Parameters

In general, the variance components and dispersion parameters are estimated from the REML approach by maximizing the REML likelihood of $f_{\boldsymbol{\theta}}(\mathbf{y}|\widehat{\boldsymbol{\beta}})$, which is straightforward in a linear mixed model. In most cases, marginal and joint likelihoods do not have a closed-form; hence it requires some approximation technique to compute the likelihoods. In such situations, the likelihood function is approximated via integral approximation methods, such as EM algorithm, Gibbs sampling, MCMC type algorithms, Laplace approximation, and h -likelihood approximation through the Laplace approximation (Laplace, 1986; Lee et al., 2006). Many researchers have shown that the Laplace approximation is computationally efficient and with less biased compared to Gibbs

sampling and MCMC algorithms with high dimensional integration approximations (Breslow & Clayton, 1993; Youngjo Lee & John A Nelder, 1996; Noh & Lee, 2007). Lee and Nelder (2006) introduced an alternative approach to obtain dispersion parameters using h -likelihood through the Laplace approximation, which is known as the adjusted h -likelihood function (Lee et al., 2006).

The Laplace approximation is used to approximate an integral of the form $\int_{x_0}^{x_1} e^{cf(x)} dx$, where $f(x)$ is a twice differentiable function, and c is a constant. Suppose that $f(x)$ is a continuous, n times differentiable function, $\lim_{x \rightarrow x_0} f(x) = f(x_0)$, and $f''(x_0) < 0$. From, Taylor series expansion

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2}f''(x_0)(x - x_0)^2 + \dots + \frac{1}{n!}f^n(x_0)((x - x_0)^n),$$

which can be expressed as

$$f(x) \approx f(x_0) - \frac{1}{2}|f''(x_0)|(x - x_0)^2, \quad (2.12)$$

where $\lim_{x \rightarrow x_0} f'(x) = 0$. The equation (2.12) can be simplified as

$$\int_{x_0}^{x_1} e^{f(x)} dx \approx e^{f(x_0)} \int_{x_0}^{x_1} e^{-\frac{1}{2}|f''(x_0)|(x-x_0)^2} dx.$$

The above integral has the form of an arbitrary Gaussian integral function. Since $|f''(x_0)| > 0$,

$$\int_{\mathbb{R}} e^{-\frac{1}{2}n|f''(x_0)|(x-x_0)^2} dx = \lim_{n \rightarrow \infty} \left(\frac{2\pi}{n|f''(x_0)|} \right)^{1/2}.$$

Now, the Laplace approximation to $\int_{x_0}^{x_1} e^{nf(x)} dx$ can be written as

$$\int_{x_0}^{x_1} e^{nf(x)} dx \approx e^{nf(x_0)} \left(\frac{2\pi}{n|f''(x_0)|} \right)^{\frac{1}{2}},$$

which can be shown using the lower and upper bounds from Taylor's theorem as

$$\lim_{n \rightarrow \infty} \frac{\int_{x_0}^{x_1} e^{nf(x)} dx}{e^{nf(x_0)} \left(\frac{2\pi}{n(-f''(x_0))} \right)^{\frac{1}{2}}} = 1,$$

where $f(x)$ has a global maximum at $x = x_0$, then $f''(x_0) < 0$. This expression can be easily generalized to first-order Laplace approximation as

$$\int e^{f(x)} dx \approx e^{f(x_0)} \left\{ \left| -\frac{1}{2\pi} f''(x) \right|^{-\frac{1}{2}} \right\} \Bigg|_{x=x_0},$$

where x_0 is a global maximum of some function $f(x)$. This technique is used defining an adjusted h -likelihood h_A where it is cumbersome to approximate the integrals when obtaining the REML and marginal likelihoods (Barndorff-Nielsen, 1983).

Taking the log transformation of Laplace approximation

$$\begin{aligned} \log \int \exp f(x) dx &\approx \log \left\{ \left| -\frac{1}{2\pi} \frac{\partial^2 f(x)}{\partial x^2} \right|^{-\frac{1}{2}} \exp f(x) \right\} \Bigg|_{x=x_0}, \\ &= f(x) \Big|_{x=x_0} - \frac{1}{2} \left\{ \log \det \left(\frac{\mathcal{J}}{2\pi} \right) \right\} \Bigg|_{x=x_0}, \\ \int f(x) dx &= f(x) \Big|_{x=x_0} + \frac{1}{2} \{ \log \det(2\pi \mathcal{J}^{-1}) \} \Big|_{x=x_0}, \end{aligned}$$

where $\mathcal{J} = \partial^2 f(x) / \partial x^2$. This concept was used proposing the adjusted h -likelihood h_A to approximate the REML log-likelihood $f_\theta(\mathbf{y} | \hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$ (Lee et al., 2006)

$$\begin{aligned} h_A &= h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2} \log \{ \det(2\pi \mathcal{J}^{-1}) \} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}, \\ &= h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2} \log \{ (2\pi)^{p+m} \det(\mathcal{J}^{-1}) \} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}, \\ &= h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2} (p+m) \log 2\pi - \frac{1}{2} \log \det(\mathcal{J}) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}, \quad (2.13) \end{aligned}$$

which is also known as the penalized partial likelihood function (PPL). Now, the MLE of ϑ is obtained by maximizing the adjusted profile h -likelihood h_A with respect to ϑ . Since \mathcal{J} is invertible, we have the score function $\partial h_A / \partial \vartheta$ as

$$\begin{aligned} \frac{\partial h_A}{\partial \vartheta} &= \frac{\partial h}{\partial \vartheta} \Big|_{\beta=\hat{\beta}, u=\hat{u}} - \frac{\partial}{\partial \vartheta} \left(\frac{1}{2} \log\{\det(\mathcal{J})\} \Big|_{\beta=\hat{\beta}, u=\hat{u}} \right), \\ &= \frac{\partial h}{\partial \vartheta} \Big|_{\beta=\hat{\beta}, u=\hat{u}} - \frac{1}{2} \text{trace} \left(\mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \vartheta} \right) \Big|_{\beta=\hat{\beta}, u=\hat{u}}. \end{aligned} \quad (2.14)$$

The MHLE of ϑ is obtained by solving $\partial h_A / \partial \vartheta = 0$ iteratively updating $\hat{\beta}$ and \hat{u} which are obtained through the Newton Raphson method.

2.5. Multilevel Small Area Estimation in Survey Research Using GLMM

Small area estimation (SAE) is a statistical technique mostly used in survey research to obtain statistically reliable estimates for smaller areas such as zip code areas, counties, or sub-county areas (Rahman & Harding, 2016; John NK Rao & Molina, 2015). Existing studies of SAEs that were developed using adult data from the Behavioral Risk Factor Surveillance System (BRFSS) have been focusing on the estimation of adult smoking prevalence (Dwyer-Lindgren et al., 2014). In this section, we applied the GLMM approach to estimate youth electronic cigarette use prevalence at the county level using the combined 2014 and 2015 National Youth Tobacco Survey (NYTS).

We performed a multilevel SAE model to incorporate individual-level tobacco use behaviors and area-level (county and state) ecological characteristics to electronic use prevalence among youth at the community level (i.e., county). Our SAE model is based on the analytical framework in conjunction with the mixed modeling of random effects and post-stratification simulation at the county level. This analytical design is considered superior to individual only or ecological only studies, which could yield mixed results as they miss the other important component in the analysis (Wills & Soneji, 2018).

2.5.1 Emerging Tobacco Use Among Adolescents

Tobacco use is the leading cause of preventable death in the world (Centers for Disease Control and Prevention, 2013; U.S. Department of Health and Human Services, 2014), with the vast majority of tobacco use beginning during adolescence. In 2017, an estimated 3 million middle and high school students were current tobacco product users (T. W. Wang et al., 2018). In the United States, 9 out of 10 current smokers started smoking before age 18 (U.S. Department of Health and Human Services, 2012). Nicotine exposure and tobacco use at an early age can harm brain development, is associated with lower rates of smoking cessation and increased risk of addiction and other substance use, including marijuana (Taioli & Wynder, 1991; U.S. Department of Health and Human Services, 2012, 2014).

The tobacco use landscape of youth has substantially changed in recent years, with more adolescents using e-cigarettes and other emerging tobacco products (Jamal et al., 2017). Among U.S. high school students, the current use of electronic cigarettes (e-cigarettes) has outpaced the use of traditional cigarettes (Jamal et al., 2017; U.S. Department of Health and Human Services, 2016). E-cigarettes contain varying levels of nicotine, many potentially toxic substances, and different flavors, which might serve as a gateway for future cigarette use (Barrington-Trimis et al., 2016; Cameron et al., 2014; Dinakar & O'Connor, 2016; Goniewicz, Hajek, & McRobbie, 2014; Jensen, Luo, Pankow, Strongin, & Peyton, 2015; Leventhal et al., 2015; Primack, Soneji, Stoolmiller, Fine, & Sargent, 2015; Soneji et al., 2017; Wills et al., 2016). Cigars currently ranked third among the most commonly used tobacco products (Jamal et al., 2017), and cigar smoke is possibly more toxic than cigarette smoke (Institute, 1998; U.S. Department of Health and Human Services, 2012). Flavorings are frequently added to cigars to enhance their appeal to youth (Kostygina, Glantz, & Ling, 2016).

A growing body of literature has evaluated emerging youth tobacco use at the national level and identify patterns and socio-economic factors associated with youth substance use (H. Dai,

2018; H. Dai, 2019; H. Dai, D. Catley, K. P. Richter, K. Goggin, & E. F. Ellerbeck, 2018). However, no study has evaluated the prevalence of emerging tobacco use among adolescents. Although enormous progress has been made in reducing tobacco use in the US, this progress has not been equally distributed across populations with large disparities in tobacco use persisting across groups defined by race/ethnicity, education level, income level, region, and other factors (National Cancer Institute, 2017; N. A. Rigotti & S. Kalkhoran, 2017). Thus, assessing tobacco-related health disparities using small area modeling can provide information for policymakers and stakeholders to develop interventions in curbing emerging tobacco use among adolescents.

2.5.2 Survey Materials and Methods

We used data from the National Youth Tobacco Survey (NYTS). The NYTS is an FDA-approved science-based survey approach to study public health issues with tobacco use, which includes data on the long term, intermediate and short term relevant to tobacco usage. It was developed to provide data necessary to support the design, implementation, and evaluation of state and national tobacco prevention and control programs (Brian King, 2014; MacDonald, Starr, Schooley, Yee, & Klimowski, 2001). The NYTS includes tobacco-related data tobacco use (e.g., cigarettes, e-cigarettes, cigars, tobacco pipes, bidis, etc.), exposure to secondhand smoke, smoking cessation, etc. representing all middle school and high school students in the 50 US states and the District of Columbia. The NYTS employed a stratified, three-stage cluster sample design to produce a nationally representative sample to represent the middle school and high school students in the US.

The sampling was done without replacement and oversampling the non-Hispanic African American students. The students within each school were selected from the primary sampling units (PSUs) considered as a county, or a group of small counties, or part of a very large county; the secondary sampling units (SSUs) are defined as schools or linked schools within each selected PSU. The participation in the NYTS was voluntary at both the school and student levels, and the

survey was completed via pencil and paper using a self-administered, scannable questionnaire booklet. Schools used either passive or active permission forms to fulfill requirements of the No Child Left Behind Act, giving options to the parents to opt-out their child's participation.

We used 2014-2015 NYTS youth tobacco use data downloaded from the Centers for Disease Control and Prevention (CDC). There were 207 participants from 258 schools (school participation rate of 80.2%) in 2014, and 185 participants from 255 schools (school participation rate of 72.6%) in 2015. A total of 22,007 out of 24,084 (82 US counties) in 2014 and 17,711 out of 20,259 (80 US counties) students in 2015 had completed the questionnaires. More details about the survey methodology are given in Hu 2016 (Hu, 2016).

Table 2.1: Summary statistics of prevalence in 162 counties by the total, age group, sex, and race of NYTS 2014-2015 respondents.

Characteristic	No. of Respondents	Prevalence E-cig Ever (%)	No. of Respondents E-cig current	Prevalence E-cig current (%)
Total	38683	22.74	38835	10.08
Missing	1125		883	
Age				
10-14	19846	13.10	19897	5.42
15-19	18837	32.90	18938	14.98
Sex				
Male	19553	25.01	19654	11.73
Female	18969	20.39	19019	8.36
Race				
White	18189	23.61	16208	11.06
Black	5781	17.14	5832	5.69
Hispanic	10598	26.36	10651	11.60
Others	2357	17.14	2375	8.00

Initially, the e-cigarettes were introduced to reduce tobacco use in so doing to decrease the health issues. However, e-cigarette usage among youth has been increased due to many reasons, such as flavors of cigarettes, being less harmful than other tobacco, etc. (Ambrose et al., 2015; Hongying Dai, 2019; H. Dai, 2019; Hongying Dai & Hao, 2016, 2018; Villanti, Richardson,

Vallone, & Rath, 2013). Our primary focus is the prevalence of current and ever use of E-cigarettes, 1. ever use of E-cigarettes (yes or no), 2. current use of E-cigarettes (yes or no). The original NYTS data had 39,718 unit (individual or student)-level observations from 162 counties in 41 states in the US.

We extracted poverty data for 3220 counties in 50 US states, the District of Columbia, and US territories from the American Community Survey (ACS) for the period 2011-2015 US Census. We considered four demographic groups for the race, white, African American, Hispanic, and others, including “Asian alone”, “American Indian and Alaska Native alone”, “Native Hawaiian and other Pacific Islander alone”, and “multiple races”. County wise subpopulation for age x sex x race cross-tabulated data for 16 demographic categories were extracted from 2015 5-year API Census data using tidycensus R package (K. Walker, 2018). The fully adjusted sampling weight is taken adjusting the sampling weight, which is the inverse of the probability of selection to alleviate the effect from non-respondents.

2.5.3 GLMM Model Specification

We consider the GLMM to obtain EBLUPs for county-wise prevalence as follows

$$\text{logit}(P_{ijkc}) = \text{logit}(P(y_{ijkc} = 1|u)) = \alpha_i + \beta_j + \gamma_k + x'_c\eta + \mu_c + e_{ijkc}, \quad (2.15)$$

where α_i , β_j , and γ_k are regression coefficients of age group $i = 1, 2$ ($1 = 10 - 14$, $2 = 15 - 19$ years), sex $j = 1, 2$ ($1 = \text{male}$, $2 = \text{female}$), and race $k = 1, \dots, 4$ ($1 = \text{white}$, $2 = \text{black}$, $3 = \text{hispanic}$, $4 = \text{others}$) respectively. x_c is a vector of county level covariates, η is corresponding regression coefficients for county level covariates, μ_c is a vector of county-level random effects, and e_{ijkc} is the residual. We only considered the county level poverty rate for x_c and obtained two models to obtain EBLUPs for current-use and ever-use of E-cigarettes using the FH method. The poverty rates were extracted from the 2011-2015 US Census. We got county-level fixed effects and random effects using PROC GLIMMIX in SAS (Schabenberger, 2005). Table 2.2 gives the fixed

effect estimates, age, gender, race, and the poverty rate for the ever-use and current use of the E-Cigarettes using the FH model.

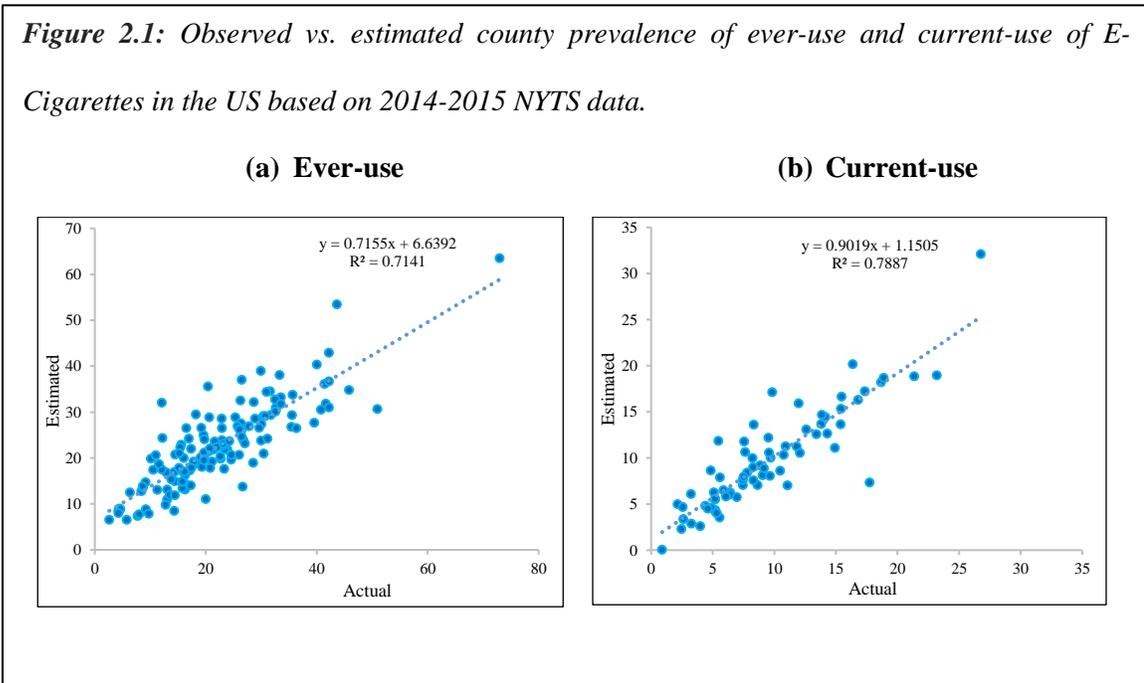
Table 2.2: Model estimates using the FH model for ever-use and current-use of E-Cigarettes.

Solutions for Fixed Effects								
	E-Cigarette Ever				E-Cigarette Current			
Effect	Estimate	SE	DF	Pr > t	Estimate	SE	DF	Pr > t
Intercept	-1.510	0.165	129.8	<.0001	-2.426	0.266	119.7	<.0001
Age								
10-14	-1.251	0.001	36791	<.0001	-1.232	0.001	36947	<.0001
15-19	0	.	.	.	0	.	.	.
Gender								
Male	0.212	0.001	36791	<.0001	0.340	0.001	36947	<.0001
Female	0	.	.	.	0	.	.	.
Race								
White	0.313	0.002	36791	<.0001	0.233	0.003	36947	<.0001
African American	-0.062	0.002	36791	<.0001	-0.400	0.003	36947	<.0001
Hispanic	0.530	0.002	36791	<.0001	0.383	0.003	36947	<.0001
Others	0	.	.	.	0	.	.	.
Poverty (%)	1.884	1.017	129.5	0.0662	0.411	1.643	119.6	0.8030

The estimates in Table 2.2 were used to make predictions for prevalence in all the counties in the US using RStudio (Team, 2015). The p-value of the poverty rate is not significant at the 0.05 significance level, indicating that it is not associated with both ever-use and current-use of E-Cigarettes. The odds ratios $e^{0.53} = 1.70$, and $e^{0.38} = 1.45$ are highest among the Hispanic youth compared to other races after adjusting for age, gender, and the poverty rate for both ever-use and current-use. White also has a higher odds ratio than African Americans compared to other races. It shows that the age group 10-14 has a lower odds ratio for both ever-use and current-use of E-Cigarettes compared to the age group 15-19.

Based on scatter plots 2.1(a) and 2.1(b) for estimated and observed prevalence, it shows that reliable estimates for both ever-use and current use of E-Cigarettes. Overall, most counties have a prevalence of less than 60% for ever-use and less than 30% for current-use except for one

county. Tables A.1 and A.2 show the estimated and actual prevalence for ever-use and current-use of E-cigarettes for 162 US counties using the FH model, respectively.



2.5.4 County Level Prediction

As stated above, we used the estimated random effects to predict random effects for unknown counties via the nearest neighboring approach (NNA). It is reasonable to obtain random effects using the NNA approach since the neighboring counties might share similar characteristics. The unknown random effect for county i , is replaced with the random effect of the closest county (Zahava Berkowitz et al., 2016b) based on the Euclidian distance of the centroid of each

$$\tilde{\mu}_{c_i} = \hat{\mu}_{c_j}, s. t. \min \text{dist}(c_i, c_k), k = 1, \dots, m - 1,$$

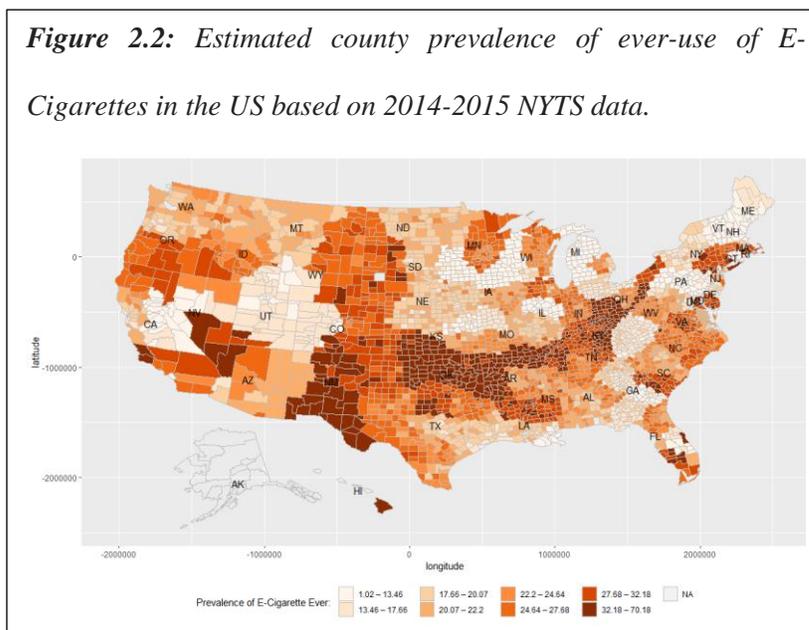
where $\hat{\mu}_{c_j}$ is the estimated random effect of county c_j , and c_j is the closest to county c_i . The estimated county-level random effects and the model parameters were used to predict the individual level prevalence for E-cigarette current usage and E-cigarette ever usage using the equation (2.15).

$$\tilde{P}_{ijkc}(y_{ijkc} = 1|u) = \frac{\exp(\alpha_i + \beta_j + \gamma_k + x'_c\eta + \mu_c)}{1 + \exp(\alpha_i + \beta_j + \gamma_k + x'_c\eta + \mu_c)}$$

The county-level prevalence was determined using

$$\tilde{P}(y_c = 1|u) = \frac{\sum_i \sum_j \sum_k \tilde{P}_{ijkc} \times \text{Pop}_c}{\text{Pop}_{ijkc}},$$

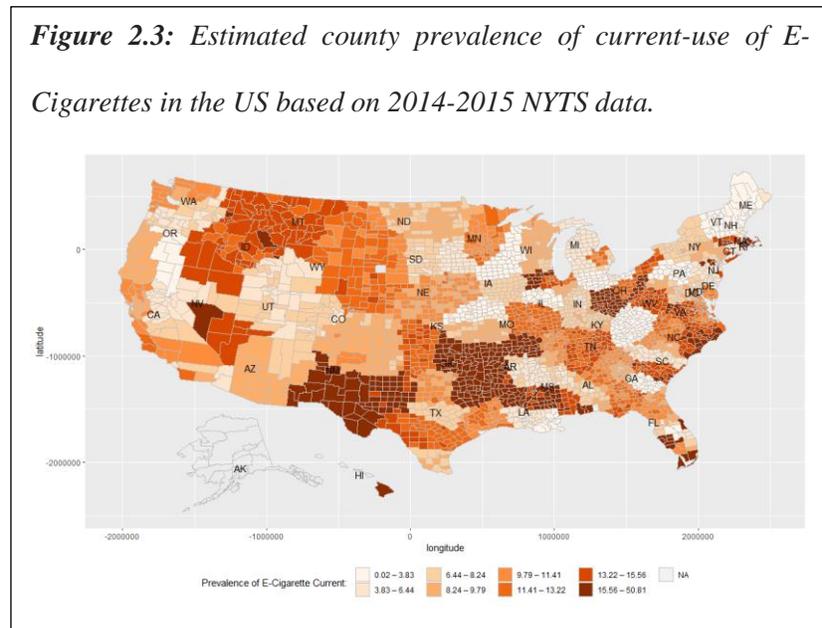
where $\text{Pop}_c = \sum_i \sum_j \sum_k \text{Pop}_{ijkc}$ is the total population for county c .



Based on Figure 2.2, most counties have a prevalence between 21-30% for ever-use of E-Cigarettes. Next, the prevalence group of 11-20% is more prevalent compared to other prevalence groups. Overall, the states Hawaii, Oklahoma, New Mexico, Arkansas, Kentucky, Rhode Island, Indiana, Ohio, Tennessee, and Kansas fall in the top ten US states with large ever-use of E-Cigarettes prevalence compared to other US states.

Figure 2.3, compared with Figure 2.2, implies that similar areas that have a higher prevalence for ever-use also have a higher prevalence for current-use, which indicates that modeling these two outcomes through joint modeling might be crucial to study the association

among them. We proposed the joint modeling approach for such variables of interest, which is discussed in detail in chapter 4.



2.6 Literature Review on Advanced Modelling Techniques in Small Area Estimation

Many researchers have adopted a Hierarchical Generalized Linear modeling (HGLM) approach in the past to make more accurate inferences in SAE due to its' flexibility of analyzing continuous and discrete data, as well as allowing random effect \mathbf{u} to have any conjugate distribution from the GLM family. In recent years, the Empirical Bayes (EB) and Hierarchical Bayes (HB) related GLM methods are also proposed. Furthermore, the robustness of these estimators against the estimators from different estimation methods is extensively discussed (Datta & Ghosh, 1991; Farrell et al., 1997; Gómez-Rubio, Best, Richardson, Li, & Clarke, 2010; MacGibbon & Tomberlin, 1987; Martuzzi & Elliott, 1996; Stroud, 1994; Yasui et al., 2000). The EB approach assumes that the population data are a sample from a super population, which can be represented by an empirical Bayes model. This assumption clearly states the issues of repeated survey analysis where the

population of inference might be different across the period of the survey conducted. Such situations have been handled via time series methodology in SAE, expanding the empirical Bayes method, and borrowing information from previous surveys as well as from other areas (Jon NK Rao & Yu, 1994).

Most research studies are conducted based on cross-sectional data as well as the repeated measures data. Rao and Yu (1994) proposed an extension of the basic area-level FH model considering autocorrelated random effects and sampling error for SAE using time series and cross-sectional data. The two-stage estimator of the small area mean for the current period is obtained under two scenarios; 1. for known autocorrelation, the BLUP is first obtained and then replaced with their consistent estimators, 2. for unknown autocorrelation, three methods have been proposed to estimate autocorrelation via the simulation approach (Jon NK Rao & Yu, 1994).

Hierarchical Bayes (HB) methodology in SAE has been widely considered in different scenarios (Ghosh et al., 1998; Stroud, 1994). Ghosh et al. (1998) provided a theorem covering binary, count, multi-category, and spatial data based on the Hierarchical Bayes GLMM and implemented via the MCMC approximation method, in particular through the Gibbs sampling technique using two real datasets. One dataset with a multi-category outcome variable representing any negative impact of experiencing health hazards exposure in the workplace (yes, no, not exposed, and not applicable or not stated) is considered along with other demographic information including age, sex, and region in 15 regions in Canada on a 1991 sample. The authors compared the estimates from the simulation approach with the estimates from the multi-category logistic regression model, describing that the standard errors of the HB estimates are much lower than the sample proportions.

The second data set relates to cancer mortality rates due to lung cancer in 115 counties in Missouri during 1972-1981. The outcome variable of interest, the log relative risk, was modeled linearly based on auxiliary information for gender (male, female), age (45-54, 55-64, 65-74, and 75 and older), and gender-age interaction. The HB model is fitted via the Gibbs sampling approach

with and without spatial correlation. They tried a few models starting from the simplest model (dropping the interaction term) to the complex model (increasing the number of predictors) and showed that the model with five fixed effects offers the best fit. Furthermore, the authors claimed that the spatial model provides reasonably accurate results than the non-spatial model (Ghosh et al., 1998).

Many researchers have proposed empirical Bayes SAE approaches with the assumption of population data is a sample from the larger super population (A. Dempster & Tomberlin, 1980; R. E. Fay III & R. A. Herriot, 1979; MacGibbon & Tomberlin, 1987; Morris, 1983; Yasui et al., 2000). The EB methodology for estimation of small area proportions or binary responses is first proposed by Dempster et al. (1980). The technique was illustrated incorporating random effects and nested random effects to estimate census undercount of small groups of the population. The US Census Bureau has conducted matching studies in each census since 1950, collecting data from external sources such as administrative records close to the census date. It is then compared with the appropriate census group, and the missing proportion in census records is considered as undercount for relevant subgroups. The probability of an individual being in a census is estimated as a function of demographic characteristics such as age, gender, and race categories using a logistic model introducing a local area effect. The disadvantage of this model is the ability to include data only within the local area and only obtaining estimates and interactions for regional areas. The authors extended the model considering the prior distributions for random effects of primary sampling units (PSU), secondary sampling units (SSU) within PSUs, and tertiary sampling units within SSUs are assumed to be normal with zero mean and different variances. The corresponding EB logistic model is adopted to estimate census undercount from an approximate posterior distribution of undercounting (A. Dempster & Tomberlin, 1980).

Furthermore, MacGibbon et al. (1989) considered the empirical Bayes SAE technique proposed by Dempster et al. (1980) to estimate small area proportions, including random effects and nested random effects. The EB estimates are obtained through an MC simulation study and

compared with the estimated proportions based on classical unbiased estimates, model-based estimates (synthetic estimates), and empirical Bayes estimates using a labor force participation data from 15 PSUs. The auxiliary data were generated with identical distributions, uniformly distributed age, and Bernoulli distributed gender variable with proportion 0.5 for each group. The model-based or the synthetic estimate was obtained based on the fixed effects logistic model considering only local area-level data. However, the synthetic estimators are biased due to a lack of capturing the between area effects. This is solved from the empirical Bayes logistic model where random effects are assumed to have a multivariate normal distribution with mean zero and the variance σ_u^2 . The model first assumed the variance component of the random effects σ_u^2 , then estimate the proportions or the posterior distribution of random effects from the simulated data set. It is then used to obtain the maximum likelihood function of σ_u^2 , and estimate MLE $\widehat{\sigma}_u^2$ via the EM algorithm. Next, $\widehat{\sigma}_u^2$ is used to estimate the posterior variances (MacGibbon & Tomberlin, 1987).

The binomial logit modeling in SAE has been considered through various estimation approaches to estimate fixed, random effect estimates using maximum likelihood estimation methods, Bayesian inferential methods, and methods of moments (MOM), etc. Hobza and Morales (2016) adopted a unit-level binomial logit mixed model to estimate the empirical best prediction of weighted sums of probabilities via MCMC simulation (Hobza & Morales, 2016). The random effects are assumed to be independent and identically distributed $\mathbf{u} = (u_1, \dots, u_m)^T \sim N(\mathbf{0}, \mathbf{I}_m)$, where \mathbf{I}_m unit matrix is the variance-covariance matrix of random effects, and $y_{ij}|u_i \sim \text{Bin}(n_{ij}, p_{ij}), i = 1, \dots, m, j = 1, \dots, n_i$. The variance parameter and the regression parameters are estimated using the method of simulated moments (MSM) through the Newton-Raphson iterative formula. Furthermore, the empirical best predictor for p_{ij} and the weighted sum of probabilities for small area i were approximated via Monte Carlo (MC) simulation. Last, the mean squares error (MSE) and the error corrections of the parameters were obtained via MOM and MC simulation.

The hierarchical likelihood approach is not only widely used in estimating the random effects and fixed effects, but also used in error correction, which occurs in the process of estimating random effects and fixed effects. Lee et al. (2011) proposed hierarchical likelihood prediction intervals for random effects and fixed effects using the h -likelihood approach (Lee et al., 2011). The authors showed that the prediction interval from HL is very accurate compared to the results from penalized quasi-likelihood and fully Bayesian methods using a lip cancer dataset in areas of Scotland and an infant mortality dataset in British Columbia, Canada.

Furthermore, HGLMs are also considered in Bayesian approaches through MCMC methods to estimate parameters, and the methodology is illustrated using state-level and hospital-level auxiliary data to describe the cluster-specific rates of utilization for both hospitals and states. Huang and Wolfe considered h -likelihood using the EM algorithm together with MCMC and stated that it could also be considered as a modification of the MCMC. These approaches discuss the variations due to clustering and cluster size. Additionally, the estimates can be used to conclude higher levels using the hierarchical structure of the data (Daniels & Gatsonis, 1999; X. Huang & Wolfe, 2002). Ghosh et al. (1998) considered Hierarchical Bayes GLM to model discrete and continuous data using MCMC integration to obtain the joint posterior distribution avoiding high-dimensional numerical integration (Ghosh et al., 1998).

Most of the researchers have considered uncorrelated random effects, but it is reasonable to consider the correlation between neighboring areas in many practical application problems. The correlation approaches zero when the distance between neighborhood areas increases. Spatial hierarchical models are considered when the random effects $u_i, i = 1, \dots, m$ are not *iid*, i.e., they are correlated. The most common spatial small area model is the conditional auto-regression (CAR) spatial model, which assumes that the conditional distribution of small area i $Z_i u_i$ in equation (1.1), given the area effects for the other areas (neighboring areas for area i) can be obtained using the information of area i , where $Z_i = 1/\sqrt{C_i}$ in CAR spatial model, C_i is census count for small area i ,

and $Z_i = 1$ in the basic FH model. The CAR spatial model is used in estimating US Census undercount of certain areas based on spatial dependence (Cressie, 1991).

In 2006, Alessandra et al. considered a spatially correlated random effects models and proposed a methodology based on the EBLUP estimator using a simultaneously autoregressive (SAR) model (Petrucci & Salvati, 2006). The authors proposed an estimator for MSE combining EBLUP with a SAR model, called spatial EBLUP ($\theta(\hat{\sigma}_u^2, \hat{\rho})$) estimators which were evaluated via MC simulation of spatially correlated random effects data using a soil erosion dataset in a southwest watershed. The estimated spatial autocorrelation coefficients from the ML and REML method show a strong spatial relationship between the random effects. Under the SAR model, the synthetic estimator in equation (1.1) has the form $\hat{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}(\mathbf{I} - \rho\mathbf{W})^{-1}\mathbf{u} + \boldsymbol{\epsilon}$, where \mathbf{W} is the spatial weight matrix for $\boldsymbol{\theta}$, $\rho \in (-1,1)$ is the spatial autoregressive coefficient.

Most importantly, it is required to concentrate on point estimates as well as prediction errors for each small area, but not the average prediction error in SAE problems. The prediction error or the accuracy of the estimates plays a significant contribution to the accuracy of estimates since SAE considers essential practical applications, such as health disease estimation, fund allocation estimation, environmental health-related estimations, US census undercount estimation, etc. Many researchers have proposed prediction MSE (PMSE) estimators using various estimation techniques (Jiang, Lahiri, & Wan, 2002; N. Prasad & Rao, 1986; N. N. Prasad & Rao, 1990). In 1990, Prasad et al. developed PMSE estimators which occur when estimating EBLUP of $\boldsymbol{\beta}$ and variance components $\boldsymbol{\theta}$ under the linear mixed models where the random effects are independent normally distributed with variance σ_u^2 which is estimated using MOM method (N. N. Prasad & Rao, 1990).

Chapter 3. Small Area Estimation Using Calibrated Hierarchical Likelihood Approach

3.1 CH with Bias Correction for Generalized Family Distributions

Our main goal is to adopt the hierarchical likelihood (h -likelihood, also known as h -loglikelihood) approach with the bias correction to obtain the improved maximum likelihood estimates for the model parameters in SAE. The proposed method is called the Calibrated Hierarchical Likelihood with Bias Correction (CHBC). As stated above, a unique aspect of the h -likelihood approach is that it avoids multi-dimensional integration of the nuisance variables when obtaining the parameter estimates, the variables $\mathbf{u} = (u_1, \dots, u_m)^t$ are treated as parameters and jointly estimated along with $\boldsymbol{\beta}$, and σ^2 .

First, consider the h -likelihood for the HGLM model,

$$h = \ell(\theta, \phi; \mathbf{y}|\mathbf{u}) + \ell(\theta; \mathbf{u}), \quad (3.1)$$

where $\ell(\theta, \phi; \mathbf{y}|\mathbf{u})$ is the natural logarithm of the conditional likelihood of $\mathbf{y}|\mathbf{u}$, and $\ell(\theta; \mathbf{u})$ is the logarithm of the probability density function of random effect \mathbf{u} . Table 3.1 provides some examples of HGLMs.

Table 3.1: HGLM models with components of some canonical GLM families.

HGLM Model	$\mathbf{y} \mathbf{u}$	$\theta(\mu)$	$b(\theta)$	ϕ	$g(\cdot)$	\mathbf{u}
Normal-Normal	Normal	μ	$\theta^2/2$	σ^2	identity	Normal
Binomial-Normal	Binomial	$\log \mu$	$\exp \theta$	1	logit	Normal
Poisson-Normal	Poisson	$\log \mu/(1 - \mu)$	$\log(1 + \exp \theta)$	1	log	Normal
Gamma-Normal	Gamma	$-\mu^{-1}$	$-\log(-\theta)$	$1/\alpha$	log	Normal
Poisson-Gamma	Poisson	$\log \mu/(1 - \mu)$	$\log(1 + \exp \theta)$	1	log	Gamma

The CHBC method involves parameter estimation using an iterative approximation based on the h -likelihood. Hence, it requires the partial derivatives of h -likelihood with respect to model parameters. The proposed CHBC can be generalized into a broad class of canonical GLM family distributions of $\mathbf{y}|\mathbf{u}$, and \mathbf{u} , which mainly depends on four key elements in the score function \mathcal{S} and the Fisher information matrix \mathcal{J} ; the score statistic of GLM model $s(\theta; \mathbf{y})$, the weight matrix \mathbf{W} , the first partial derivative of $\ell_{\mathbf{u}} \nabla_{\ell_{\mathbf{u}}}^1$, and the second partial derivative of $\ell_{\mathbf{u}} \nabla_{\ell_{\mathbf{u}}}^2$. This generalization simplifies the process of taking partial derivative with respect to all the parameters in the model.

First, from (3.1), consider the log-likelihood of the GLM family distribution of $\mathbf{y}|\mathbf{u}$, followed by the definition of HGLMs by Lee and Nelder (1996) (Youngjo Lee & John A Nelder, 1996),

$$\ell(\theta, \phi; \mathbf{y}|\mathbf{u}) = \frac{\sum\{\mathbf{y}\theta - b(\theta)\}}{\phi} + c(\mathbf{y}, \phi), \quad (3.2)$$

where θ is the canonical parameter, $b(\cdot)$, and $c(\cdot)$ are known functions, and ϕ is the dispersion parameter. The random effect \mathbf{u} follows a distribution from the GLM family.

Take the partial derivative of (3.2) with respect to β for j^{th} observation

$$\frac{\partial \ell_j}{\partial \beta_j} = \frac{\partial \ell_j}{\partial \theta_j} \frac{\partial \theta_j}{\partial \mu_j} \frac{\partial \mu_j}{\partial \eta_j} \frac{\partial \eta_j}{\partial \beta_j}$$

$$\frac{\partial \ell_j}{\partial \theta_j} = \frac{y_j - b'(\theta_j)}{\phi_j}.$$

Since $\mu_j = b'(\theta_j)$, $\partial \mu_j / \partial \theta_j = b''(\theta_j)$. For canonical GLM, $g(\mu_j) = \eta_j = X_{ij}\beta_j + Z_{ij}u_j = \theta_j$, thus

$$\frac{\partial \mu_j}{\partial \eta_j} = \frac{\partial g^{-1}(\eta_j)}{\partial \eta_j} = \frac{\partial \mu_j}{\partial \theta_j} = b''(\theta_j), \text{ and } \frac{\partial \eta_j}{\partial \beta_j} = X_{ij}.$$

Now

$$\frac{\partial \ell_j}{\partial \beta_j} = \frac{y_j - b'(\theta_j)}{\phi_j} (b''(\theta_j))^{-1} b'''(\theta_j) X_{ij} = \frac{y_j - b'(\theta_j)}{\phi_j} X_{ij}.$$

Similarly,

$$\frac{\partial \ell_j}{\partial u_j} = \frac{y_j - b'(\theta_j)}{\phi_j} u_{ij}.$$

Thus, the score function of (3.2) can be expressed as (McCullagh, 2018),

$$s(\theta; \mathbf{y}) = \frac{\partial}{\partial \theta} \ell(\theta, \phi; \mathbf{y} | \mathbf{u}) = \frac{\mathbf{y} - b'(\theta)}{\phi}. \quad (3.3)$$

The MHLEs are obtained through the iterative Newton Raphson method using the h -likelihood. Then, the MHLEs are improved based on the bias correction method via the Regression Calibration (RC), which is discussed later in this chapter. First, consider the Newton Raphson iterative approximation (Whiteside, 1967)

$$\begin{aligned} \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k+1)} \\ \hat{\mathbf{u}}^{(k+1)} \end{pmatrix} &= \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k)} \\ \hat{\mathbf{u}}^{(k)} \end{pmatrix} + (\mathcal{J}^{-1} \mathcal{S}(\tau))|_{(\boldsymbol{\beta}, \mathbf{u}) = (\hat{\boldsymbol{\beta}}^{(k)}, \hat{\mathbf{u}}^{(k)})} \\ &= \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k)} \\ \hat{\mathbf{u}}^{(k)} \end{pmatrix} + \begin{pmatrix} -\frac{\partial^2 h}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} \\ \frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} & -\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \mathbf{u}} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \mathbf{u}} \end{pmatrix} \Bigg|_{(\boldsymbol{\beta}, \mathbf{u}) = (\hat{\boldsymbol{\beta}}^{(k)}, \hat{\mathbf{u}}^{(k)})}, \end{aligned} \quad (3.4)$$

$\boldsymbol{\beta}$, \mathbf{u} in \mathcal{J} and $\mathcal{S}(\tau)$ are replaced with $\hat{\boldsymbol{\beta}}^{(k)}$ and $\hat{\mathbf{u}}^{(k)}$, respectively in each iteration. The variance-covariance matrix $\boldsymbol{\vartheta}$ will be estimated by MHLE of $\boldsymbol{\vartheta}$ iteratively as described in section 3.5 via adjusted h -likelihood (h_A) approach.

Now, the \mathcal{S} using (3.3), and \mathcal{J} in the generalized CHBC method can be expressed as follows,

$$\mathcal{S} = \begin{pmatrix} \frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T s(\theta; \mathbf{y}) \\ \mathbf{Z}^T s(\theta; \mathbf{y}) + \boldsymbol{\nabla}_{\mathbf{u}}^1 \end{pmatrix}, \quad (3.5)$$

$$\mathcal{J} = \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \mathbf{V}_{\ell_u}^2 \end{pmatrix}, \quad (3.6)$$

where $\boldsymbol{\mu} = E(\mathbf{Y}) = b'(\boldsymbol{\theta})$, $\mathbf{W} = \text{Diag}(1/\text{Var}(\mathbf{y})(g'(\boldsymbol{\mu}))^2)$, $g(\cdot) = \text{link function}$, $\boldsymbol{\eta} = g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$, $\mathbf{V}_{\ell_u}^1 = \partial \ell_u / \partial \mathbf{u}$, $\mathbf{V}_{\ell_u}^2 = -\partial^2 \ell_u / \partial \mathbf{u}^2$. We illustrate the proposed approach using Binomial-Normal HGLM and Poisson-Normal HGLM, which are discussed in the following sections of this chapter.

3.1.1 Binomial-Normal HGLM

In this section, we explain the proposed method using binary outcomes with normally distributed random effects, known as the Binomial-Normal HGLM or the mixed logit model. For binary outcomes in small area estimation, consider the mixed logit model

$$\text{logit } P(y_{ij} = 1) = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i, \quad (3.7)$$

where $\mathbf{y} = (y_{ij}, 1 \leq i \leq m, 1 \leq j \leq n_i)$ is the binary response vector which is independent given the random effects u_1, \dots, u_m . Let $\mathbf{u} = (u_i)_{1 \leq i \leq m}$, $\mathbf{x}_{ij} = (x_{ijk})_{1 \leq k \leq p}$ is a vector of covariates, and $\boldsymbol{\beta} = (\beta_k)_{1 \leq k \leq p}$ is a vector of unknown fixed effects. The density of $\mathbf{y}|\mathbf{u}$ follows a Bernoulli distribution with the joint density of (\mathbf{y}, \mathbf{u}) as $f(\mathbf{y}, \mathbf{u}) = f(\mathbf{y}|\mathbf{u}) \times f(\mathbf{u})$,

$$f(\mathbf{y}, \mathbf{u}) = \prod_{ij} P(y_{ij} = 1|u_i)^{y_{ij}} (1 - P(y_{ij} = 1|u_i))^{1-y_{ij}} \times f(u_i), \quad (3.8)$$

where $P(y_{ij} = 1) = \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i) / (1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i))$, from (3.7). Furthermore, suppose u_1, \dots, u_m are independent and identically distributed as $N(0, \sigma^2)$ with $f(u_i) = (2\pi\sigma^2)^{-1/2} \exp(-1/(2\sigma^2)u_i^2)$. This assumption is reasonable to consider as n becomes large, the difference in MHLEs and actual values of fixed and random effects converge to a normal distribution, i.e. $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}, \widehat{\mathbf{u}} - \mathbf{u}) \rightarrow N(0, \boldsymbol{\tau}^2)$ where $\boldsymbol{\tau}$ is the limit of the inverse of the observed Fisher information matrix (\mathcal{J}^{-1}) .

Now, consider the joint log-density of \mathbf{y} and \mathbf{u} for the mixed logit model,

$$\begin{aligned}
\ell &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} \log P(y_{ij} = 1|u_i) + (1 - y_{ij}) \log \left(1 - P(y_{ij} = 1|u_i) \right) \right] + \sum_{i=1}^m \log f(u_i) \\
&= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} \log \left(\frac{(1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1}}{1 - (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1}} \right) + \log \left(1 - (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1} \right) \right] \\
&\quad - \frac{1}{2} \sum_{i=1}^m \log(2\pi) - \frac{1}{2} \sum_{i=1}^m \log(\sigma^2) + \sum_{i=1}^m \log \left[\exp \left(-\frac{u_i^2}{2\sigma^2} \right) \right] \\
&= \sum_{i=1}^m \sum_{j=1}^{n_i} [y_{ij}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i) - \log(1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i))] - \frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2.
\end{aligned}$$

Equivalently, from (3.3), we have

$$h = \mathbf{y}^T (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})) - \frac{1}{2} \log(\det(\mathbf{G})) - \frac{1}{2} \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u} + c, \quad (3.9)$$

where $c = -\frac{m}{2} \log 2\pi$ is a constant, and $\mathbf{G} = \sigma^2 \mathbf{I}$ is the variance-covariance matrix of \mathbf{u} .

3.1.2 Poisson-Normal HGLM

Consider the Poisson-Normal HGLM (also known as Poisson mixed model), where

$$y_{ij}|u_i \sim \text{Poisson}(\lambda), u_i \sim \mathcal{N}(0, \sigma^2), \quad (3.10)$$

λ is the mean or the event rate of discrete events, and σ^2 is the variance of the latent random variable u_i .

$$\log \lambda_{ij} = \mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a $(p \times 1)$ vector of unknown regression coefficients for the fixed effects, $\mathbf{x}_{ij}^T = (x_{1ij}, \dots, x_{p ij})^T$ is auxiliary information for p variables, and superscript T indicates transpose, y_{ij} is the response value for j^{th} object in i^{th} small area, σ^2 is the unknown variance parameter of random effect u_i for i^{th} small area.

Now from (3.1), consider the h -likelihood for Poisson-Normal HGLM

$$h = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i},$$

where ℓ_{1ij} is the log-likelihood of a Poisson random variable $y_{ij}|u_i$, and ℓ_{2i} is the log-likelihood of u_i , respectively,

$$\ell_{1ij} = \log \left(\frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}} \right),$$

$$\ell_{2i} = -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2.$$

$$\begin{aligned} h &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[\log \left(\frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} e^{-\lambda_{ij}} \right) \right] + \sum_i \left[-\frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2 \right] \\ &= \sum_{i=1}^m \sum_{j=1}^{n_i} \left[y_{ij} (\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i) - \log y_{ij}! - e^{(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i)} \right] + \sum_i \left[-\frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2 \right] + c, \quad (3.11) \end{aligned}$$

where $c = -\frac{m}{2} \log 2\pi$.

3.2 Maximum Hierarchical Likelihood Estimates of $(\boldsymbol{\beta}, \mathbf{u})$

Let $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$ be the fixed effect and random effect parameters. Given the variance parameter $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_q)^T = \sigma^2$, the maximum hierarchical (h)-likelihood estimators (MHLEs) of $\hat{\boldsymbol{\tau}} = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$ are obtained by solving the score function $\partial h / \partial \boldsymbol{\tau} = 0$. Since the solution does not have a closed-form, the MHLEs for $(\boldsymbol{\beta}, \mathbf{u})$ are obtained using the Newton-Raphson approximation through an iterative procedure using equation (3.4) as described above.

Here, we provide both methods of obtaining the elements of $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{J}}$ to illustrate how the CHBC can be applied to any GLM family distributions using Poisson-Normal HGLM. First, consider taking partial derivatives of h -likelihood with respect to $\boldsymbol{\tau}$. Holding the variance parameter $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_q)^T = \sigma^2 I_{m \times m}$ constant, the partial derivative of h -likelihood in equation (3.11) with respect to $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$ is

$$\begin{aligned}
h &= \sum_{i=1}^m \sum_{j=1}^{n_i} [y_{ij}(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i) - \log(1 + \exp(\mathbf{x}_{ij}^T \boldsymbol{\beta} + u_i))] - \frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2 \\
\frac{\partial h}{\partial \beta_r} &= \sum_{ij} [y_{ij} x_{ijr} - (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1} x_{ijr}] \\
&= \sum_{ij} [y_{ij} x_{ijr} - p_{ij} x_{ijr}],
\end{aligned}$$

where $p_{ij} = (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1}$, $r = 1, \dots, p$ the number of fixed covariates. For $i = 1, \dots, m$ random effects, we have

$$\begin{aligned}
\frac{\partial h}{\partial u_i} &= \sum_{j=1}^{n_i} [y_{ij} - (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-1}] - (\sigma^2)^{-1} u_i \\
&= \sum_{j=1}^{n_i} [y_{ij} - p_{ij}] - (\sigma^2)^{-1} u_i.
\end{aligned}$$

The entries of the observed information matrix \mathbf{J} of $\boldsymbol{\beta}$ and u_i are obtained as follows.

For $r = 1, \dots, p, s = r$,

$$\begin{aligned}
-\frac{\partial^2 h}{\partial \beta_r \partial \beta_s} &= \sum_{ij} [\exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i) (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-2} x_{ijr} x_{ijs}] \\
&= \sum_{ij} [p_{ij} (1 - p_{ij}) x_{ijr} x_{ijs}] \\
&= \sum_{ij} [W_{ij} x_{ijr} x_{ijs}];
\end{aligned}$$

for $r = 1, \dots, p, i = 1, \dots, m$,

$$\begin{aligned}
-\frac{\partial^2 h}{\partial \beta_r \partial u_i} &= \sum_{j=1}^{n_i} [x_{ijr} \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i) (1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-2}] \\
&= \sum_{j=1}^{n_i} [x_{ijr} p_{ij} (1 - p_{ij})] \\
&= \sum_{j=1}^{n_i} [x_{ijr} W_{ij}];
\end{aligned}$$

for $i = 1, \dots, m, s = 1, \dots, p,$

$$\begin{aligned} -\frac{\partial^2 h}{\partial u_i \partial \beta_s} &= \sum_{j=1}^{n_i} \left[(1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-2} \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i) x_{ijs} \right] \\ &= \sum_{j=1}^{n_i} [p_{ij}(1 - p_{ij})x_{ijs}] \\ &= \sum_{j=1}^{n_i} [W_{ij}x_{ijs}]; \end{aligned}$$

for $i = 1, \dots, m, l = i,$

$$\begin{aligned} -\frac{\partial^2 h}{\partial u_i \partial u_l} &= \sum_{j=1}^{n_i} \left[(1 + \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i))^{-2} \exp(-\mathbf{x}_{ij}^T \boldsymbol{\beta} - u_i) \right] + (\sigma^2)^{-1} \\ &= \sum_{j=1}^{n_i} p_{ij}(1 - p_{ij}) - (\sigma^2)^{-1} \\ &= \sum_{j=1}^{n_i} W_{ij} - (\sigma^2)^{-1}; \end{aligned}$$

where $W_{ij} = p_{ij}(1 - p_{ij})$, for $i = 1, \dots, m, l \neq i$, we have $\frac{\partial^2 h}{\partial u_i \partial u_l} = 0$.

The above expressions can be expressed in matrix form as follows,

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}), \quad (3.11)$$

$$\frac{\partial h}{\partial \mathbf{u}} = \mathbf{Z}^T (\mathbf{y} - \boldsymbol{\pi}) - \mathbf{G}^{-1} \mathbf{u}, \quad (3.12)$$

where $\boldsymbol{\pi} = (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1}$. Let \mathbf{y} be a $(N \times 1)$ vector of outcome measures, \mathbf{X} be a $(N \times p)$ matrix of fixed effects information with p explanatory variables, $\boldsymbol{\beta}$ be a $(p \times 1)$ vector of fixed effects' regression coefficients, \mathbf{Z} be the $(N \times m)$ design matrix of explanatory variables with random effects, \mathbf{u} be a $(m \times 1)$ vector of random effects, and $\mathbf{G} = \sigma^2 \mathbf{I}$ be the $(m \times m)$ be the variance-covariance matrix of random effects.

Now, consider the CHBC for canonical GLM families. h -likelihood in matrix form,

$$h = \mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})) - \frac{1}{2}\mathbf{u}^T \mathbf{G}^{-1}\mathbf{u} - \frac{1}{2}\log(\det(\mathbf{G})) + c,$$

compared with (3.1),

$$b(\boldsymbol{\theta}) = \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})), \boldsymbol{\theta} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) \text{ and } \phi = 1,$$

hence

$$b(\boldsymbol{\theta}) = \log(1 + \exp \boldsymbol{\theta}),$$

taking the partial derivative with respect to $\boldsymbol{\theta}$

$$b'(\boldsymbol{\theta}) = (1 + \exp \boldsymbol{\theta})^{-1} \exp \boldsymbol{\theta} = (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} = \boldsymbol{\pi}.$$

Thus,

$$\mathbf{X}^T(\mathbf{y} - \boldsymbol{\pi}) = \mathbf{X}^T(\mathbf{y} - b'(\boldsymbol{\theta})) = \mathbf{X}^T \mathcal{S}(\boldsymbol{\theta}; \mathbf{y}).$$

Similarly, we can show that

$$\mathbf{Z}^T(\mathbf{y} - \boldsymbol{\pi}) - \mathbf{G}^{-1}\mathbf{u} = \mathbf{Z}^T \mathcal{S}(\boldsymbol{\theta}; \mathbf{y}) - \mathbf{G}^{-1}\mathbf{u} = \mathbf{Z}^T \mathcal{S}(\boldsymbol{\theta}; \mathbf{y}) + \mathbf{V}_{\boldsymbol{\ell}_u}^1.$$

Therefore, the score function \mathcal{S} (eq. 3.5) can be used as the generalized version of the \mathcal{S} , which can be easily expressed using $b'(\boldsymbol{\theta})$ and ϕ .

3.3 Bias Correction

As stated above, random effects and fixed effects are jointly estimated using the HMLEs procedure through the Newton Raphson approximation. However, it is shown that directly plugging in the estimated \mathbf{u} , $\hat{\mathbf{u}}$, which was estimated using h -likelihood will result in biasedness when estimating the parameters of interest, especially based on nonlinear likelihood functions. Moreover, the HMLEs of $\hat{\mathbf{u}}$ will lead to bias and inconsistent estimations of regression coefficients and parameters in the random effects, such as variance parameters. Researchers have considered and proposed many methods to avoid this biasedness in parameter estimation (Carroll, Ruppert, Stefanski, & Crainiceanu, 2006; Hausman, Newey, & Powell, 1995; C. Wang, Hsu, Feng, & Prentice, 1997). We use the regression calibration approach proposed by Wang et al. (1997) to correct for biasedness in the estimators. The use of a correction factor is an important step to

alleviate the biasedness of MLEs, which MLEs are adjusted based on the sample size and its' variance. Furthermore, the regression calibration method (RCM) is computationally efficient and has been applied widely in correcting measurement errors in many applications since it provides reliable estimates than direct estimates with no bias correction.

Wang et al. (1997) used the idea from the large sample theory and proposed the regression calibration method for measurement error correction. The primary purpose of the RCM is to adjust $\hat{\mathbf{u}}$ using the variance of $\mathbf{u}|\hat{\mathbf{u}}$. From large sample theory, when N becomes large, maximum likelihood estimators of the likelihood function have the asymptotic normality property $\sqrt{N}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \approx N\left(0, (\mathbf{J}(\boldsymbol{\tau}))^{-1}\right)$, where $\boldsymbol{\tau} = (\boldsymbol{\beta}, \mathbf{u})$, $\hat{\boldsymbol{\tau}}$ is a consistent and asymptotically efficient estimator of $\boldsymbol{\tau}$, which is also known as the Cramer-Rao Lower Bound (Casella & Berger, 2002). The variance of $\mathbf{u}|\hat{\mathbf{u}}$, $\boldsymbol{\gamma}$ is the $(m \times m)$ right lower corner matrix of $(\mathbf{J}(\boldsymbol{\tau}))^{-1}$ the asymptotic variance-covariance matrix of MHLEs.

Since $\mathbf{u} \sim N(0, \sigma^2)$, $(\mathbf{u}, \hat{\mathbf{u}})$ has a joint-normal distribution,

$$\begin{pmatrix} \mathbf{u} \\ \hat{\mathbf{u}} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & \sigma^2 + \boldsymbol{\gamma}^2 \end{pmatrix}\right).$$

Then, the conditional distribution of \mathbf{u} given $\hat{\mathbf{u}}$ can be derived as

$$\mathbf{u}|\hat{\mathbf{u}} \sim N(\boldsymbol{\zeta}\hat{\mathbf{u}}, \sigma^2(1 - \boldsymbol{\zeta})), \text{ where } \boldsymbol{\zeta} = \frac{\sigma^2}{\sigma^2 + \boldsymbol{\gamma}^2}, \boldsymbol{\gamma} \text{ is the limit of } \mathbf{J}^{-1},$$

which is a vector, obtained from the diagonal elements of \mathbf{J}^{-1} .

Now, under the regression calibration method, we replace \mathbf{u} by

$$E[\mathbf{u}|\hat{\mathbf{u}}] = \boldsymbol{\zeta}\hat{\mathbf{u}} = \frac{\sigma^2}{\sigma^2 + \boldsymbol{\gamma}^2} \hat{\mathbf{u}}. \quad (3.13)$$

Note that, if the variance of $\hat{\mathbf{u}}$, $\boldsymbol{\gamma} \approx 0$, then $E[\mathbf{u}|\hat{\mathbf{u}}] \approx 1$, which means the MHLEs with RCM and the original MHLEs are similar. However, when $\text{Var}(\hat{\mathbf{u}})$ is large, the MHLE results in the biased estimates, thus the RCM adjusts the biasedness of the estimates.

3.4 The MHLEs of Dispersion Parameters (ϑ)

From equation (2.13), the adjusted h -likelihood

$$h_A = h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} - \frac{1}{2} \log\{\det(\mathcal{J})\}|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{p+m}{2} \log 2\pi,$$

where m is the number of dispersion parameters, and p is the number of fixed effects.

The generic way of obtaining the maximum adjusted profile h -likelihood estimator for ϑ is by solving the score function $\partial h_A / \partial \vartheta = 0$, but this becomes complicated when there is no closed form for ϑ . Thus, it requires some numerical approximation. First, consider the score function of the adjusted profile h -likelihood from the equation (2.14)

$$\frac{\partial h_A}{\partial \vartheta} = \frac{\partial h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}}{\partial \vartheta} - \frac{1}{2} \text{trace} \left(\mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \vartheta} \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}, \quad (3.14)$$

which is used to estimate the variance-covariance matrix $\boldsymbol{\vartheta}$ of random effect \mathbf{u} . Since $\partial \ell_{1ij} / \partial \vartheta = 0$, the score function can be written as

$$\frac{\partial h}{\partial \vartheta} = 0 + \frac{\partial}{\partial \vartheta} \left(\sum_{i=1}^m \ell_{2i} \right)$$

The adjusted profile log-likelihood of $\mathbf{u} \sim N(0, \vartheta)$ for small area i

$$\ell_{2i} = -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \vartheta - \frac{1}{2\vartheta} \sum_{i=1}^m u_i^2 = -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \vartheta - \frac{1}{2\vartheta} \mathbf{u}^T \mathbf{u},$$

where $\vartheta I_{m \times m} = \mathbf{G} = \sigma^2 I_{m \times m}$. Now, from $h = \sum_{ij} \ell_{1ij} + \sum_i \ell_{2i}$, $\partial h / \partial \vartheta$ given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \mathbf{u} = \hat{\mathbf{u}}$ can be represented as

$$\begin{aligned} \frac{\partial h|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}}{\partial \vartheta} &= 0 + \frac{\partial}{\partial \vartheta} \left(-\frac{m}{2} \log \vartheta - \frac{1}{2\vartheta} \sum_{i=1}^m \hat{u}_i^2 \right) \\ &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}}. \end{aligned}$$

The partial derivative of the observed information matrix with respect to ϑ given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}, \mathbf{u} = \hat{\mathbf{u}}$

$$\frac{\partial \mathbf{J}}{\partial \vartheta} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} = \frac{\partial}{\partial \vartheta} \begin{pmatrix} \left(-\frac{\partial^2 h}{\partial \boldsymbol{\beta}^2} \right)_{p \times p} & \left(-\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} \right)_{p \times m} \\ \left(-\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} \right)_{m \times p} & \left(-\frac{\partial^2 h}{\partial \mathbf{u}^2} \right)_{m \times m} \end{pmatrix},$$

From (3.5)

$$\begin{aligned} \frac{\partial h_A}{\partial \vartheta} &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &\quad - \frac{1}{2} \text{trace} \left(\begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}^{-1} \frac{\partial}{\partial \vartheta} \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{Z}^T \mathbf{W} \mathbf{Z} + (\vartheta \mathbf{I}_{m \times m})^{-1} \end{pmatrix} \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} - \frac{1}{2} \text{trace} \left(\begin{pmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\vartheta^{-2} \mathbf{I}_{m \times m} \end{pmatrix} \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} - \frac{1}{2} \text{trace} \left(\begin{pmatrix} \mathbf{J}_{11}^* & \mathbf{J}_{12}^* \\ \mathbf{J}_{21}^* & \mathbf{J}_{22}^* \end{pmatrix} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\vartheta^{-2} \mathbf{I}_{m \times m} \end{pmatrix} \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} - \frac{1}{2} \text{trace} \begin{pmatrix} \mathbf{0} & -\mathbf{J}_{12}^* \vartheta^{-2} \mathbf{I}_{m \times m} \\ \mathbf{0} & -\mathbf{J}_{22}^* \vartheta^{-2} \mathbf{I}_{m \times m} \end{pmatrix} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \\ &= \left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2\vartheta^2} \text{trace}(\mathbf{J}_{22}^*) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \end{aligned}$$

Set $\partial h_A / \partial \vartheta = 0$

$$\left(-\frac{m}{2\vartheta} + \frac{1}{2\vartheta^2} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2\vartheta^2} \text{trace}(\mathbf{J}_{22}^*) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} = 0.$$

Thus,

$$\hat{\vartheta} = \frac{1}{m} \left(\sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{m} \text{trace}(\mathbf{J}_{22}^*) \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}, \mathbf{u}=\hat{\mathbf{u}}} \quad (3.15)$$

where $\hat{\vartheta}$ is the MLE of the variance parameter ϑ . p is the number of fixed effects, \mathbf{J}_{22}^* is the lower right block matrix of \mathbf{J}^{-1} with dimensions being $m \times m$, which is calculated using current

estimates of $\boldsymbol{\beta}$ and \mathbf{u} . As described in appendix C, by the asymptotic properties of MHLE of \mathbf{u} , $\mathbf{J}_{22}^* = (-\partial^2 h | \partial \mathbf{u}^2 |_{\mathbf{u}=\hat{\mathbf{u}}})^{-1}$.

Algorithm

1. Set $\boldsymbol{\beta}_0$, \mathbf{u}_0 , and $\vartheta_0(\sigma_0^2)$.
2. Evaluate quantities $\mathcal{S}(\boldsymbol{\tau})$ and \mathbf{J} using (3.5) and (3.6), respectively.
3. Estimate $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^{(k)}$, and $\mathbf{u} = \hat{\mathbf{u}}^{(k)}$ using Newton Raphson (eqn. (3.4)), where $k = 1, 2, \dots$ indicates the iteration.
4. Update $\mathbf{J}^{(k)}$ using $\hat{\boldsymbol{\beta}}^{(k)}$ and $\hat{\mathbf{u}}^{(k)}$ and estimate $\mathbf{J}_{22}^{*(k)}$ using $\mathbf{J}^{-1(k)}$ matrix.
5. Estimate $\hat{\vartheta}^{(k)}$ using equation (3.15), by replacing \mathbf{u} with $E[\mathbf{u} | \hat{\mathbf{u}}^{(k)}] = \hat{\boldsymbol{\zeta}}^{(k)} \hat{\mathbf{u}}^{(k)} = \hat{\mathbf{u}}^{c(k)}$, where $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_m)$, $\zeta_i^{(k)} = \frac{\hat{\vartheta}^{(k-1)}}{\hat{\vartheta}^{(k-1)} + \tau_i^{(k)}}$, $\hat{\vartheta}^{(k-1)} = \text{Var}(\hat{\mathbf{u}}^{(k-1)})$, $\tau_i^{(k)} = \mathbf{J}_{ii}^{*(k)}$, \mathbf{J}_{ii}^* is the (i, i) diagonal element of the lower right corner matrix of 2×2 \mathbf{J}^{-1} block matrix $((\mathbf{J}^{-1})_{22})$ evaluated at t^{th} iteration. By asymptotic properties of $\hat{\mathbf{u}}$, $(\mathbf{J}^{-1})_{22} = (\mathbf{J}_{22})^{-1}$ (Youngjo Lee & John A Nelder, 1996).
6. Update h by replacing \mathbf{u} with $\hat{\mathbf{u}}^{c(k)}$, and $\exp(\mathbf{u})$ with $E[\exp(\mathbf{u}) | \hat{\mathbf{u}}^{(k)}] = \exp\left(\hat{\boldsymbol{\zeta}}^{(k)} \hat{\mathbf{u}}^{(k)} + \left(\hat{\vartheta}^{(k)}(1 - \hat{\boldsymbol{\zeta}}^{(k)})\right) / 2\right)$.
7. Repeat steps 2 to 6 until it meets the convergence criteria, which is defined as

$$\max \left\{ \left| \hat{\boldsymbol{\beta}}^{(k+1)} - \hat{\boldsymbol{\beta}}^{(k)} \right|, \left| \hat{\vartheta}^{(k+1)} - \hat{\vartheta}^{(k)} \right| \right\} < \delta$$

where δ is a predetermined tolerance limit. The MHLEs of $\boldsymbol{\beta}$, \mathbf{u} , and σ are $\hat{\boldsymbol{\beta}}$, $\hat{\mathbf{u}}$, and $\hat{\sigma}^2 (= \hat{\vartheta}^{(t)})$, respectively.

3.5 Asymptotic Properties of MHLEs

3.5.1 Asymptotic Efficiency of $\hat{\boldsymbol{\beta}}$

The advantage of MHLE is to offer a hierarchical estimation procedure to accommodate a wide range of distributions in mixed effect models. The total sample size ($N = \sum n_i$) becomes large when either the number of small areas (m) increases or the sample size of each small area increases (n_i), $i = 1, \dots, m$. However, the asymptotic properties in mixed models do not hold for any scenario. However, the asymptotic properties of MHLE $\hat{\boldsymbol{\beta}}$ has been claimed based on the ML estimation process of $\boldsymbol{\beta}$ using $E[\partial h / \partial \boldsymbol{\beta} | y, \hat{\boldsymbol{\beta}}^{(p)}] = \mathbf{0}$. It is proven that the difference of MHLE of $\boldsymbol{\beta}$ and marginal MLE of $\boldsymbol{\beta}$ is $\mathcal{O}(N^{-1})$, and hence both have asymptotically common variance. Additionally, the MHLE $\hat{\boldsymbol{\beta}}$ had been claimed to be asymptotically efficient, since marginal MLE $\hat{\boldsymbol{\beta}}$ is asymptotically efficient (Youngjo Lee & John A Nelder, 1996), i.e., as $N \rightarrow \infty$, $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}$, and asymptotically normal, $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \rightarrow \mathcal{N}(\mathbf{0}, I(\boldsymbol{\beta})^{-1}|_{\hat{\boldsymbol{\beta}}})$, and the variance of MHLE is given by the Cramer Rao lower bound; $N^{-1}I(\hat{\boldsymbol{\beta}})^{-1}$. The Laplace Approximation often provides extremely accurate results, as well as a negligible contribution of relative error from the additional terms ($\mathcal{O}(N^{-1})$) as n_i becomes large. In contrast, when n_i and m are relatively small, the additional error term $\mathcal{O}(N^{-1})$ might not be negligible (Booth & Hobert, 1998; Youngjo Lee & John A Nelder, 1996; Liu & Pierce, 1993).

Lee and Nelder (1996) discussed the marginal MLE of $\boldsymbol{\beta}$ in GLMMs could be obtained using the Laplace approximation to the marginal likelihood $\hat{\ell} \propto h - 1/2 \log(\det(\mathcal{J}^*))$, where $\mathcal{J}^* = -\partial^2 h | \partial \mathbf{u}^T \partial \mathbf{u} |_{\mathbf{u}=\hat{\mathbf{u}}}$, as described in Breslow and Clayton (1993) and Liu and Pierce (1993) (Breslow & Clayton, 1993; Liu & Pierce, 1993). Under certain conditions, they showed that the MLE of $\boldsymbol{\beta}$ could be obtained using $\partial \ell / \partial \boldsymbol{\beta} |_{\mathbf{u}=\hat{\mathbf{u}}} = \mathbf{0}$ which leads to $\partial h(\mathbf{y}; \boldsymbol{\beta}, \hat{\mathbf{u}}) / \partial \boldsymbol{\beta} |_{\mathbf{u}=\hat{\mathbf{u}}} = \mathbf{0}$

assuming $\partial \mathcal{J}^*/\partial \boldsymbol{\beta}|_{\mathbf{u}=\hat{\mathbf{u}}}$ is negligible. Then the Taylor series expansion to $\partial h(\mathbf{y}; \boldsymbol{\beta}, \hat{\mathbf{u}})/\partial \boldsymbol{\beta}|_{\mathbf{u}=\hat{\mathbf{u}}}$ can be expressed as

$$\frac{\partial h(\mathbf{y}; \boldsymbol{\beta}, \hat{\mathbf{u}})}{\partial \boldsymbol{\beta}} = \frac{\partial h(\mathbf{y}; \boldsymbol{\beta}, \mathbf{u})}{\partial \boldsymbol{\beta}} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} + (\mathbf{u} - \hat{\mathbf{u}}) C_1|_{\mathbf{u}=\hat{\mathbf{u}}} + \frac{1}{2!} (\mathbf{u} - \hat{\mathbf{u}})^T C_2 (\mathbf{u} - \hat{\mathbf{u}}) \Big|_{\mathbf{u}=\hat{\mathbf{u}}} + \dots, \quad (B.1)$$

where $C_1 = (\partial/\partial \boldsymbol{\beta})(\partial h(\mathbf{y}; \boldsymbol{\beta}, \mathbf{u})/\partial \mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}$, and $C_2 = (\partial/\partial \boldsymbol{\beta})(\partial^2 h(\mathbf{y}; \boldsymbol{\beta}, \mathbf{u})/\partial \mathbf{u}^T \partial \mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}$.

Furthermore, Lee and Nelder (1996) claimed that the asymptotic properties of $\hat{\boldsymbol{\beta}}$ based on the ML estimation process of $\boldsymbol{\beta}$ using $E[\partial h/\partial \boldsymbol{\beta}|\mathbf{y}] = 0$. From (A.3) and, (A.4), $E[\mathbf{u}|\mathbf{y}] + \mathcal{O}(N^{-1}) = \hat{\mathbf{u}}$ and $\text{Var}(\mathbf{u}|\mathbf{y}) = (-\partial^2 h|\partial \mathbf{u}^T \partial \mathbf{u}|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1} + \mathcal{O}(N^{-1})$, then, $N(E(\mathbf{u}|\mathbf{y}) - \hat{\mathbf{u}}) = N(\mathcal{O}(N^{-1})) = \mathcal{O}(1)$, and,

$$E[C_1(\mathbf{u} - \hat{\mathbf{u}})|_{\mathbf{u}=\hat{\mathbf{u}}}] = E\left[\frac{C_1}{N} \mathcal{O}(1)\right] = \frac{C_1}{N} = \mathcal{O}(1).$$

Similarly, from (A.4),

$$E[(\mathbf{u} - \hat{\mathbf{u}})^T C_2 (\mathbf{u} - \hat{\mathbf{u}})|_{\mathbf{u}=\hat{\mathbf{u}}}] = \frac{C_2}{n} = \mathcal{O}(1).$$

Now,

$$\begin{aligned} E\left[\frac{\partial h}{\partial \boldsymbol{\beta}}\right] &= \frac{\partial h}{\partial \boldsymbol{\beta}} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} + E[C_1(\mathbf{u} - \hat{\mathbf{u}})|_{\mathbf{u}=\hat{\mathbf{u}}}] + \frac{1}{2!} E[(\mathbf{u} - \hat{\mathbf{u}})^T C_2 (\mathbf{u} - \hat{\mathbf{u}})|_{\mathbf{u}=\hat{\mathbf{u}}}] + \dots \\ &= \frac{\partial h}{\partial \boldsymbol{\beta}} \Big|_{\mathbf{u}=\hat{\mathbf{u}}} + \mathcal{O}(1), \end{aligned}$$

where $C_j/n = \mathcal{O}(1)$, $j = 1, 2$.

Now, the MLEs are obtained using $E[\partial h/\partial \boldsymbol{\beta}|\mathbf{y}] = 0$ and the difference of MHLE of $\boldsymbol{\beta}$ and the marginal MLE of $\boldsymbol{\beta}$ can be shown as $\mathcal{O}(N^{-1})$.

3.5.2 Asymptotic Properties of MHLE of $\hat{\mathbf{u}}$

The MHLE of the random effect $\hat{\mathbf{u}}$ can be shown as an asymptotically best-unbiased predictor (ABUP) for \mathbf{u} with an additional term $\mathcal{O}(N^{-1})$ compared to marginal MLE. Lee and Nelder (1996)

explains that the necessary condition for MHLE of \mathbf{u} to have the asymptotic properties is $(-\partial^2 h / \partial u_i^2)^{-1} |_{\mathbf{u}=\hat{\mathbf{u}}} = \mathcal{O}(N^{-1})$ for all i . This condition holds as $n_i \rightarrow \infty$ for fixed m . Still, $\hat{\mathbf{u}}$ will not improve much as $m \rightarrow \infty$ for fixed n_i . Using the second-order Taylor series expansion, it can be shown that $E(u_i | \mathbf{y}) = \hat{u}_i + \mathcal{O}(N^{-1})$ and $\text{Var}(u_i | \mathbf{y}) = E[(u_i^2 | \mathbf{y})] - (E[u_i | \mathbf{y}])^2 = \left(-\partial^2 h / \partial u_i^2 |_{u_i=\hat{u}_i} \right)^{-1} + \mathcal{O}(N^{-1})$.

Let $\mathbf{u} = \hat{\mathbf{u}}$ is a solution to $\partial h(\mathbf{y}; \mathbf{u}, \hat{\boldsymbol{\beta}}) / \partial \mathbf{u} = \mathbf{0}$ for given $\boldsymbol{\beta}$. If $\hat{\mathbf{u}} = E[\mathbf{u} | \mathbf{y}]$, then $\hat{\mathbf{u}}$ is called the best unbiased predictor for \mathbf{u} , where $E[\mathbf{u} | \mathbf{y}]$ is the conditional expectation of \mathbf{u} given \mathbf{y} .

$$E[\mathbf{u} | \mathbf{y}] = \frac{\int \mathbf{u} f(\mathbf{u}, \mathbf{y}) d\mathbf{u}}{\int f(\mathbf{u}, \mathbf{y}) d\mathbf{u}} = \frac{\int \mathbf{u} \exp h(\mathbf{u}) d\mathbf{u}}{\int \exp h(\mathbf{u}) d\mathbf{u}}.$$

Now, consider the Taylor Series expansion of the joint log-likelihood, similarly h around $\mathbf{u} = \hat{\mathbf{u}}$, then the numerator of $E[\mathbf{u} | \mathbf{y}]$

$$\begin{aligned} \int \mathbf{u} \exp h(\mathbf{u}) d\mathbf{u} &= \int \mathbf{u} \exp \left\{ h(\hat{\mathbf{u}}) + \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) + \dots + \mathcal{O}(N^{-1}) \right\} d\mathbf{u} \\ &\approx \exp(h(\hat{\mathbf{u}})) \left\{ \int (\mathbf{u} - \hat{\mathbf{u}}) \exp \left\{ \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) \right\} d\mathbf{u} \right. \\ &\quad \left. + \int \hat{\mathbf{u}} \exp \left\{ \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) \right\} d\mathbf{u} \right\}. \end{aligned}$$

Note that,

$$\int (\mathbf{u} - \hat{\mathbf{u}}) \exp \left\{ \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) \right\} d\mathbf{u} = 0 \text{ at } \mathbf{u} = \hat{\mathbf{u}}.$$

Thus,

$$\begin{aligned} \int \mathbf{u} \exp h(\mathbf{u}) d\mathbf{u} &= \int \hat{\mathbf{u}} \exp(h(\hat{\mathbf{u}})) \exp \left\{ \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) \right\} d\mathbf{u} \\ &= \hat{\mathbf{u}} | -2\pi (h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}})^{-1} |^{1/2} \exp\{h(\hat{\mathbf{u}})\}. \end{aligned}$$

Similarly, the denominator of $E[\mathbf{u} | \mathbf{y}]$

$$\int \exp h(\mathbf{u}) d\mathbf{u} = \int \exp \left\{ h(\hat{\mathbf{u}}) + \frac{1}{2} (\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u}) |_{\mathbf{u}=\hat{\mathbf{u}}} (\mathbf{u} - \hat{\mathbf{u}}) + \dots + \mathcal{O}(N^{-1}) \right\} d\mathbf{u},$$

$$\begin{aligned} &\approx \exp(h(\hat{\mathbf{u}})) \int \exp\left\{\frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}})\right\} d\mathbf{u}, \\ &= \left| -2\pi h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}^{-1} \right|^{1/2} \exp(h(\hat{\mathbf{u}})) \end{aligned}$$

Now, by taking the ratio of $\int \mathbf{u} \exp h(\mathbf{u}) d\mathbf{u}$ and $\int \exp h(\mathbf{u}) d\mathbf{u}$

$$E[\mathbf{u}|\mathbf{y}] \approx \hat{\mathbf{u}}, \quad (\text{A. 2})$$

which is the first order Laplace Approximation to the $\mathbf{u}|\mathbf{y}$. However, in general, the Laplace approximation incurs an error term of $O(N^{-1})$, then the expression (A. 2) can be expressed as

$$E[\mathbf{u}|\mathbf{y}] + O(N^{-1}) = \hat{\mathbf{u}} \quad (\text{A. 3})$$

Note that the relative error of $O(N^{-1})$ in the asymptotic order of N terms of the Taylor Series expansion $\lim_{N \rightarrow \infty} O(N^{-1}) = 0$.

Next, consider the $E[(\mathbf{u}\mathbf{u}^T|\mathbf{y})]$,

$$\begin{aligned} &E[(\mathbf{u}\mathbf{u}^T|\mathbf{y})] \\ &= \frac{\int \mathbf{u}\mathbf{u}^T \exp\left\{h(\hat{\mathbf{u}}) + \frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}}) + \dots + O(N^{-1})\right\} d\mathbf{u}}{(-2\pi(h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1})^{1/2} \exp(h(\hat{\mathbf{u}}))} \\ &\approx \left| -\frac{1}{2\pi} h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} \right|^{1/2} \int \mathbf{u}\mathbf{u}^T \exp\left\{\frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}})\right\} d\mathbf{u} \\ &= \left| -\frac{1}{2\pi} h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} \right|^{1/2} \left(\int (\mathbf{u} - \hat{\mathbf{u}})^T (\mathbf{u} - \hat{\mathbf{u}}) \exp\left\{\frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}})\right\} d\mathbf{u} \right. \\ &\quad \left. - \int \hat{\mathbf{u}}\hat{\mathbf{u}}^T \exp\left\{\frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}})\right\} d\mathbf{u} \right. \\ &\quad \left. + 2 \int \mathbf{u}^T \hat{\mathbf{u}} \exp\left\{\frac{1}{2}(\mathbf{u} - \hat{\mathbf{u}})^T h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}}(\mathbf{u} - \hat{\mathbf{u}})\right\} d\mathbf{u} \right) \\ &= \left| -\frac{1}{2\pi} h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} \right|^{1/2} \left((h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1} \left| -\frac{1}{2\pi} h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} \right|^{-1/2} + \mathbf{u}\mathbf{u}^T \left| -\frac{1}{2\pi} h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} \right|^{-1/2} \right) \\ &E[(\mathbf{u}\mathbf{u}^T|\mathbf{y})] = (h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1} + \hat{\mathbf{u}}\hat{\mathbf{u}}^T, \quad (\text{A. 4}) \end{aligned}$$

where $h''(\mathbf{u})|_{\mathbf{u}=\hat{\mathbf{u}}} = \partial^2 h|\partial\mathbf{u}\partial\mathbf{u}^T|_{\mathbf{u}=\hat{\mathbf{u}}}$. Now the conditional variance of $\mathbf{u}|\mathbf{y}$ from (A. 3) and (A. 4),

$$\text{Var}(\mathbf{u}|\mathbf{y}) = (-\partial^2 h|\partial\mathbf{u}\partial\mathbf{u}^T|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1} + \mathcal{O}(N^{-1}). \quad (\text{A.5})$$

The conditional variance $\mathbf{u}|\mathbf{y}$ can be simplified as $\text{Var}(\mathbf{u}|\mathbf{y}) = (-\partial^2 h|\partial\mathbf{u}\partial\mathbf{u}^T|_{\mathbf{u}=\hat{\mathbf{u}}})^{-1}$ since $\lim_{N \rightarrow \infty} \mathcal{O}(N^{-1}) = 0$.

3.5.3 Wald Confidence Interval

By the properties of MLEs, $\hat{\boldsymbol{\tau}} \xrightarrow{p} \boldsymbol{\tau}_0$ as $n \rightarrow \infty$, which means MLE $\hat{\boldsymbol{\tau}}$ converges to the true parameter value $\boldsymbol{\tau}_0$ in probability,

$$\lim_{n \rightarrow \infty} P(\boldsymbol{\tau}_0 - \epsilon < \hat{\boldsymbol{\tau}} < \boldsymbol{\tau}_0 + \epsilon) = 1, \forall \epsilon > 0,$$

and the limiting (asymptotic) distribution

$$\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}_0) \approx N(0, \mathbf{I}(\boldsymbol{\tau})^{-1}|_{\boldsymbol{\tau}_0}),$$

where $\mathbf{I}(\boldsymbol{\tau})|_{\boldsymbol{\tau}_0}$ is the expected Fisher information evaluated at $\boldsymbol{\tau} = \boldsymbol{\tau}_0$ (Lehmann & Romano, 2006).

$$\text{Var}(\hat{\boldsymbol{\tau}}) = (\mathbf{I}(\boldsymbol{\tau})|_{\hat{\boldsymbol{\tau}}})^{-1}, SE(\hat{\boldsymbol{\tau}}) = (\mathbf{I}(\boldsymbol{\tau})|_{\boldsymbol{\tau}_0})^{-1/2}.$$

The standard error is obtained from the diagonal elements of the variance-covariance matrix. Now, the properties of the asymptotic distribution of MHLEs can be used to obtain the $(1 - \alpha)\%$ confidence interval for the MHLEs

$$\hat{\boldsymbol{\tau}} \pm Z_{\alpha/2}(SE(\hat{\boldsymbol{\tau}})).$$

Chapter 4. Joint Modeling of Multiple Outcomes in Small Area

Estimation

The joint modeling approach is generally used with longitudinal or repeated measures data analysis where the measurements are taken over time, and time is considered as the random effects component. Here, a similar idea is applied in small area estimation where small areas (clusters) are regarded as the random effects component. It is often possible to observe multiple outcomes from the same individual, which might be associated, and such associations might provide additional information to obtain more precise estimations (Benavent & Morales, 2016; Burgard et al., 2020; Datta & Ghosh, 1991; González-Manteiga et al., 2008; Ha et al., 2017; Lee et al., 2017; Tsiatis & Davidian, 2004). Building separate models through univariate analysis without considering this association might not provide accurate results or might lose some vital information.

To the best of our knowledge, the joint modeling approach had not been studied, accounting for the association among multiple outcomes in SAE. Therefore, in this chapter, we considered joint analysis using a joint modeling approach to account for the association among the outcomes. Those outcomes are joint through unobserved area-specific random effects, which explain the association between multiple outcomes, so it will ignore the biased results, which could occur by conducting a separate analysis of multiple outcomes. Thus, we consider a joint modeling approach through two different ways; 1) multivariate analysis with different variances of random effects, 2) joint modeling of multiple outcomes through a shared parameter, based on the Hierarchical (h)-likelihood approach with bias correction of random effects.

The multivariate models are considered in small area estimation in different scenarios, such as the sampling variability in auxiliary information; hence the correlation between sampling sub-domains is considered through various variance components of the latent random effects (Benavent & Morales, 2016; Ubaidillah et al., 2019). The multivariate Fay-Herriot model has been widely applied in different situations in SAE. Most researchers have shown that multivariate analysis often

provides more efficient estimations since it accounts for the correlations, unlike the univariate analysis (Datta & Ghosh, 1991; González-Manteiga et al., 2008; Gueorguieva, 2001; John NK Rao & Molina, 2015). Datta et al. (1991) applied the multivariate model to obtain estimates for the population median income for four-person families. The authors considered the correlation with the population median income of three-person families and five-person families in the same state, which results in improved estimation for four-person families compared to the univariate analysis, ignoring the variability among these two groups of the population.

The multivariate modeling approach is also used in estimating the under-count US Census through poststratification using adjustment factors obtained by considering the correlation of available Census counts (Isaki, Tsay, & Fuller, 2000). Moreover, the spatial models are considered to account for the correlation between neighboring areas (Cressie, 1991; Lee, Alam, Noh, Rönnegård, & Skarin, 2016). Cressie (1991) obtained estimations for the US Census undercount, considering the spatial correlation of neighboring areas. As stated above, we study two situations of analyzing multiple outcomes through a joint modeling approach to account for the association between the outcomes that could occur from the same individual, and also through a multivariate modeling approach considering different variances among the random effects to account for the association of multiple outcomes. While the joint model includes a shared parameter that explains the association, the multivariate model includes the variance-covariance matrix to explain the association among the multiple outcomes. Section 4.1 illustrates the multivariate joint model and h -likelihood through multivariate random effects, and section 4.2 covers the joint model through shared random effects and its h -likelihood.

4.1 Multivariate Joint Model and Hierarchical (h)-Likelihood

The multiple outcomes from the same individual, for the j^{th} individual, $\mathbf{y} = (y_{1j}, \dots, y_{kj})^T, j = 1, \dots, n_i, i = 1, \dots, m$. The inherent association between such multiple outcomes can be studied

through an unobservable random effect $\mathbf{u}_i = (\mathbf{u}_{1i}, \dots, \mathbf{u}_{ki})^T = (u_{11}, \dots, u_{1m}, \dots, u_{k1}, \dots, u_{km})^T$, is a km -dimensional vector of random effects, where \mathbf{u}_{1i} , and \mathbf{u}_{ki} account for the association between y_{1j} , and y_{kj} . The multi-dimensional random effects requires a multivariate distribution for k outcomes with each small area having k random effects, and each random effect is a vector of $n_i, i = 1, \dots, m$ observations. In the case of multivariate normal random effects $\mathbf{u}_{ri} \sim \mathcal{N}(0, \sigma_r^2)$, $\mathbf{u}_i = (\mathbf{u}_{1i}, \dots, \mathbf{u}_{ri})^T \sim MN(0, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ is the $k \times k$ the variance-covariance matrix of \mathbf{u}_i for i^{th} small area. The variance-covariance matrix $\mathbf{\Sigma}$ explains the association between r random effects, a strong association between $\mathbf{u}_{1i}, \dots, \mathbf{u}_{ri}$ implies that the existence of a strong association between y_1, \dots, y_r . Thus, the association between multiple outcomes will require modeling them jointly to inherit the correlation and obtain accurate model estimates for each outcome.

First, consider the multivariate h -likelihood

$$h = \mathcal{L}_{(\theta, \phi; y_1 | \mathbf{u}_1)}(y_1 | \mathbf{u}_1) + \dots + \mathcal{L}_{(\theta, \phi; y_k | \mathbf{u}_k)}(y_k | \mathbf{u}_k) + \mathcal{L}_{(\mathbf{U}; \phi)}(\mathbf{U}), \quad (4.1)$$

where $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_k)^T$,

$$\mathcal{L}_{(\theta_r, \phi_r; y_r)}(\theta_r, \phi_r; y_r | \mathbf{u}) = \frac{\{y_r \theta_r - b(\theta_r)\}}{a(\phi_r)} + c(y_r, \phi_r), r = 1, \dots, k, \quad (4.2)$$

$\theta_r = \theta(\boldsymbol{\mu}_r)$ is the canonical parameter, $E(y_r | \mathbf{U}) = \boldsymbol{\mu}_r$, $\text{Var}(y_r | \mathbf{U}) = a(\phi_r) b''(\theta_r)$, $\boldsymbol{\eta}_r = g(\boldsymbol{\mu}_r) = \mathbf{X}_r \boldsymbol{\beta}_r + \mathbf{ZU}$, $g(\cdot)$ is the link function for the GLM, $a(\cdot)$ is a function of dispersion parameters. The systematic (or nonrandom) component relates a parameter $\boldsymbol{\eta}$ to the covariate information in the model which connects with the conditional mean $E[\mathbf{y} | \mathbf{X}] = \boldsymbol{\mu}$ with a link function $g(\cdot)$.

As described above, we obtain the MHLEs through iterative approximation based on the score equations and Newton Raphson method. We illustrate the proposed technique for canonical GLM family distributions of $\mathbf{y} | \mathbf{u}$. First, consider the score equation of (4.2),

$$s(\boldsymbol{\theta}_r; \mathbf{y}) = \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}_r, \boldsymbol{\phi}_r; \mathbf{y}_r | \mathbf{u}) = \frac{\mathbf{y} - b'(\boldsymbol{\theta}_r)}{a(\boldsymbol{\phi}_r)}, \quad (4.3)$$

where $b'(\boldsymbol{\theta}_r) = \partial b(\boldsymbol{\theta}_r) / \partial \boldsymbol{\theta}_r$. Now, the proposed generalized score function \mathcal{S} , for $r = 1, \dots, k$ outcomes, can be expressed as

$$\mathcal{S} = \begin{pmatrix} \mathbf{X}^T s(\boldsymbol{\theta}; \mathbf{y}) \\ \mathbf{Z}^T s(\boldsymbol{\theta}; \mathbf{y}) + \mathbf{V}_{\mathcal{L}U}^1 \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T \frac{(\mathbf{y} - b'(\boldsymbol{\theta}))}{a(\boldsymbol{\phi})} \\ \mathbf{Z}^T \frac{(\mathbf{y} - b'(\boldsymbol{\theta}))}{a(\boldsymbol{\phi})} + \mathbf{V}_{\mathcal{L}U}^1 \end{pmatrix}, \quad (4.4)$$

where \mathbf{X} is a block matrix with off-diagonal matrices being null matrices and each diagonal matrix $\mathbf{X}_r (N_r \times p_r)$ represents the design matrix for each outcome variable. The matrix \mathbf{Z} is a block-diagonal matrix with off-diagonal matrices being null matrices and each diagonal matrix $\mathbf{Z}_r (N_r \times m)$ represents the design matrix for each random effect.

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{X}_k \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{Z}_k \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \end{pmatrix}, \quad b'(\boldsymbol{\theta}) = \begin{pmatrix} b'(\boldsymbol{\theta}_1) \\ \vdots \\ b'(\boldsymbol{\theta}_k) \end{pmatrix}, \quad \mathbf{U} =$$

$$\begin{pmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_k \end{pmatrix}, \quad \mathbf{V}_{\mathcal{L}U}^1 = \frac{\partial \mathcal{L}U}{\partial \mathbf{U}} \text{ and the variance-covariance matrix } \boldsymbol{\Sigma} \text{ of } \mathbf{U} \text{ can have different covariance}$$

patterns, such as,

1. Independent (variance components), no correlation between multiple outcomes,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 & \dots & 0 \\ 0 & \sigma_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \sigma_{kk} \end{pmatrix},$$

2. Compound symmetry, the same correlation between each outcome,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{ee} + \sigma_{11} & \dots & \sigma_{11} \\ \vdots & \ddots & \vdots \\ \sigma_{11} & \dots & \sigma_{ee} + \sigma_{11} \end{pmatrix},$$

3. Autoregressive, spatially decreasing correlation,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho & \dots & \rho^{k-1} \\ \rho & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{k-1} & \rho^{k-2} & \dots & 1 \end{pmatrix},$$

4. Unstructured, different correlation between each outcome,

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \dots & \dots & \sigma_{kk} \end{pmatrix}.$$

The generalized Fisher information matrix \boldsymbol{J} for $r = 1, \dots, k$, can be derived as

$$\boldsymbol{J} = \begin{pmatrix} [\boldsymbol{J}_{11}]_{(\mathcal{N} \times \mathcal{N})} & [\boldsymbol{J}_{12}]_{(\mathcal{N} \times \mathcal{M})} \\ [\boldsymbol{J}_{21}]_{(\mathcal{M} \times \mathcal{P})} & [\boldsymbol{J}_{22}]_{(\mathcal{M} \times \mathcal{M})} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} & \boldsymbol{X}^T \boldsymbol{W} \boldsymbol{Z} \\ \boldsymbol{Z}^T \boldsymbol{W} \boldsymbol{X} & \boldsymbol{Z}^T \boldsymbol{W} \boldsymbol{Z} + \boldsymbol{V}_{\mathcal{L}_U}^2 \end{pmatrix}, \quad (4.5)$$

where $\boldsymbol{V}_{\mathcal{L}_U}^2 = -\partial^2 \mathcal{L}_U / \partial \boldsymbol{U}^T \partial \boldsymbol{U}$, the dimension of \boldsymbol{J} is $((\mathcal{N} + \mathcal{M}) \times (\mathcal{N} + \mathcal{M}))$, $\mathcal{N} = \sum_r^k N_r$, \mathcal{P} in the total number of covariates, and $\mathcal{M} = km$. Here, we make the working assumption that each outcome has an equal amount of covariates, i.e., $\mathcal{P} = n_1 = n_2 = \dots = n_k$. Note that, if $n_1 \neq n_2 \neq \dots \neq n_k$, then $\mathcal{P} = \max(n_1, \dots, n_k)$. Let $\boldsymbol{V}_{\mathcal{L}_U}^2 = \boldsymbol{Q}$. The weight matrix \boldsymbol{W}

$$\boldsymbol{W} = \begin{pmatrix} \text{Diag} \left(\frac{1}{\text{Var}(\boldsymbol{y}_1)(g'(\boldsymbol{\mu}_1))^2} \right) & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \text{Diag} \left(\frac{1}{\text{Var}(\boldsymbol{y}_k)(g'(\boldsymbol{\mu}_k))^2} \right) \end{pmatrix}, \quad (4.6)$$

where $g'(\cdot)$ is the first derivative of link function. $\text{Var}(\boldsymbol{y}_r) = a(\boldsymbol{\phi}_r)b''(\boldsymbol{\theta}_r)$.

4.1.1 Estimation of Dispersion Parameters in Multivariate Joint Model

The MHLEs of dispersion parameters are obtained using the adjusted h -likelihood(h_A), by using the Newton Raphson approximation. First, consider h_A

$$h_A = h|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} + \frac{1}{2} \log\{\det(2\pi\boldsymbol{J}^{-1})\}|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}, \quad (4.7)$$

where $\hat{\tau} = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}})$. Suppose $\boldsymbol{\psi}$ is a vector of dispersion parameters, which includes $k(k+1)/2$ variance parameters for multivariate (for k responses) random effects, and variance parameters from response variables. For example, for Binomial-Normal model with bivariate normal random effects has a total of four dispersion parameters, three variance parameters ($\sigma_{11}, \sigma_{12}, \sigma_{22}$) from the bivariate normal random effect and the variance of the normal response variable (σ^2). Now, for the general case, taking the first partial derivative of h_A with respect to $\boldsymbol{\psi}$,

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\psi}} h_A &= \frac{\partial}{\partial \boldsymbol{\psi}} h|_{\tau=\hat{\tau}} + \frac{1}{2} \frac{\partial}{\partial \boldsymbol{\psi}} \log\{\det(2\pi \mathcal{J}^{-1})\}|_{\tau=\hat{\tau}} \\ \frac{\partial h_A}{\partial \boldsymbol{\psi}} &= \frac{\partial}{\partial \boldsymbol{\psi}} (\mathcal{L}_{(U;\phi)}(\mathbf{U})) \Big|_{\tau=\hat{\tau}} - \frac{1}{2} \text{trace} \left(\mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} \right) \Big|_{\tau=\hat{\tau}},\end{aligned}\quad (4.8)$$

and the second partial derivative of h_A with respect to $\boldsymbol{\psi}$,

$$\frac{\partial^2 h_A}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} = \frac{\partial^2}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} (\mathcal{L}_{(U;\phi)}(\mathbf{U})) \Big|_{\tau=\hat{\tau}} - \frac{1}{2} \text{trace} \left(-\mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} \mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} + \mathcal{J}^{-1} \frac{\partial^2 \mathcal{J}}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} \right) \Big|_{\tau=\hat{\tau}} \quad (4.9)$$

where

$$\begin{aligned}\mathcal{J} &= \begin{pmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \mathcal{Q} \end{pmatrix}, \quad \mathcal{Q} = \mathbf{V}_{L_U}^2 \\ \frac{\partial \mathcal{J}}{\partial \boldsymbol{\psi}} &= \begin{pmatrix} \mathbf{X}^T \mathbf{W}' \mathbf{X} & \mathbf{X}^T \mathbf{W}' \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W}' \mathbf{X} & \mathbf{Z}^T \mathbf{W}' \mathbf{Z} + \mathcal{Q}' \end{pmatrix}, \quad \frac{\partial^2 \mathcal{J}}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} = \begin{pmatrix} \mathbf{X}^T \mathbf{W}'' \mathbf{X} & \mathbf{X}^T \mathbf{W}'' \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W}'' \mathbf{X} & \mathbf{Z}^T \mathbf{W}'' \mathbf{Z} + \mathcal{Q}'' \end{pmatrix},\end{aligned}$$

The MHLE of $\boldsymbol{\psi}$ is obtained using Newton Raphson approximation,

$$\boldsymbol{\psi}^{(i+1)} = \boldsymbol{\psi}^{(i)} + \left(-\frac{\partial^2 h_A}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} \right)^{-1} \frac{\partial h_A}{\partial \boldsymbol{\psi}} \Big|_{\boldsymbol{\Sigma}=\hat{\boldsymbol{\Sigma}}^{(i)}}. \quad (4.10)$$

The expressions for multivariate normal random effects can be expressed as,

$$\mathcal{Q} = \mathbf{V}_{L_U}^2 = \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m}, \quad \mathcal{Q}' = \frac{\partial}{\partial \boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m} = -(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1}) \otimes I_{m \times m},$$

$$\mathbf{Q}'' = \frac{\partial}{\partial \boldsymbol{\Sigma}} \mathbf{Q}' = 2(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1}) \otimes I_{m \times m},$$

$$\mathbf{Q}' = \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(-\frac{\partial^2 \mathcal{L}_{\mathbf{U}}}{\partial \mathbf{U}^T \partial \mathbf{U}} \right) = \frac{\partial \mathbf{Q}}{\partial \boldsymbol{\Sigma}}, \quad \mathbf{Q}'' = \frac{\partial^2}{\partial \boldsymbol{\Sigma}^T \partial \boldsymbol{\Sigma}} \left(-\frac{\partial^2 \mathcal{L}_{\mathbf{U}}}{\partial \mathbf{U}^T \partial \mathbf{U}} \right) = \frac{\partial^2 \mathbf{Q}}{\partial \boldsymbol{\Sigma}^T \partial \boldsymbol{\Sigma}},$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix of \mathbf{U} , \mathbf{W}' , and \mathbf{W}'' are the first and second partial derivative of \mathbf{W} with respect to $\boldsymbol{\Sigma}$. Note that, $\mathbf{W} \neq f(\boldsymbol{\Sigma})$, hence, $\mathbf{W}' = \mathbf{0}$, and $\mathbf{W}'' = \mathbf{0}$. Thus,

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{\Sigma}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}' \end{pmatrix}, \text{ and } \frac{\partial^2 \mathcal{J}}{\partial \boldsymbol{\Sigma}^T \partial \boldsymbol{\Sigma}} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}'' \end{pmatrix}.$$

Note that the proposed approach can be generalized to any GLM family distributed random effects.

Now, for multivariate normal random effects, (4.6) can be expressed as,

$$\begin{aligned} \frac{\partial h_A}{\partial \boldsymbol{\Sigma}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(\mathcal{L}_{(\mathbf{U}, \phi)}(\mathbf{U}) \right) \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace}(\mathbf{J}_{22}^{-1} \mathbf{Q}') \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(-\frac{m}{2} \log(\det(\boldsymbol{\Sigma})) - \frac{1}{2} \mathbf{U}^T \boldsymbol{\Sigma}^{-1} \mathbf{U} \right) \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace}(\mathbf{J}_{22}^{-1} \mathbf{Q}') \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(-\frac{m}{2} \log(\det(\boldsymbol{\Sigma})) - \frac{1}{2} \mathbf{U}^T \mathbf{Q} \mathbf{U} \right) \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace}(\mathbf{J}_{22}^{-1} \mathbf{Q}') \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= -\frac{m}{2} \text{trace}(\boldsymbol{\Sigma}^{-1}) - \frac{1}{2} \mathbf{U}^T \mathbf{Q}' \mathbf{U} \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace}(\mathbf{J}_{22}^{-1} \mathbf{Q}') \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \end{aligned} \quad (4.11)$$

and (4.9) can be expressed as,

$$\begin{aligned} \frac{\partial^2 h_A}{\partial \boldsymbol{\Sigma}^T \partial \boldsymbol{\Sigma}} &= \frac{\partial}{\partial \boldsymbol{\Sigma}} \left(-\frac{m}{2} \text{trace}(\boldsymbol{\Sigma}^{-1}) - \frac{1}{2} \mathbf{U}^T \mathbf{Q}' \mathbf{U} \right) \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace}(-\mathbf{J}_{22}^{-1} \mathbf{Q}' \mathbf{J}_{22}^{-1} \mathbf{Q}' + \mathbf{J}_{22}^{-1} \mathbf{Q}'') \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}} \\ &= \frac{m}{2} \text{trace}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1}) - \frac{1}{2} (\mathbf{U}^T \mathbf{Q}'' \mathbf{U}) - \frac{1}{2} \text{trace}(\mathbf{J}^{-1} \mathbf{Q}'' - \mathbf{Q}' \mathbf{J}^{-1} \mathbf{Q}' \mathbf{J}^{-1}) \Big|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}}, \end{aligned} \quad (4.12)$$

where $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m}$, $\mathbf{Q}' = -\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m}$, $\mathbf{Q}'' = 2\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m}$, $\mathbf{J}^{-1} = (\mathbf{Z}^T \mathbf{W} \mathbf{Z} + \mathbf{Q})^{-1}$.

4.1.2 Bias Correction in Multivariate Joint Model

We consider the bias correction of random effects to mitigate the biasedness that could occur due to the use of current estimates of unobserved random effects ($\hat{\mathbf{u}}$) to obtain estimations for other parameters in the model. The most common bias correction approach is the Regression Calibration (RC) method, which is simple and provides improved estimations compared to other bias corrections approaches. The RCM replaces the current estimate of the unobserved random effect by its adjusted values, i.e., the conditional expectation of $\mathbf{u}|\hat{\mathbf{u}}$ ($E[\mathbf{u}|\hat{\mathbf{u}}]$). The main advantages of RCM are its simplicity and ability to apply for any regression models with non-Gaussian data even though it is widely used with Gaussian data (Carroll et al., 2006; Fraser & Stram, 2012; Freedman, Midthune, Carroll, & Kipnis, 2008; S. Y. Huang, 2005; Spiegelman, Logan, & Grove, 2011; Spiegelman, McDermott, & Rosner, 1997).

Given $\mathbf{u}_1, \dots, \mathbf{u}_m \sim MN(\mathbf{0}, \Sigma)$, the joint distribution of $(\mathbf{u}_i, \hat{\mathbf{u}}_i)$ for i^{th} small area is multivariate normal,

$$\begin{pmatrix} \mathbf{u}_i \\ \hat{\mathbf{u}}_i \end{pmatrix} \sim MN \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma \\ \Sigma & \Sigma + \mathbf{V}_i \end{pmatrix} \right),$$

and the conditional distribution of $\mathbf{u}_i|\hat{\mathbf{u}}_i$ is also multivariate normal, $\mathbf{u}_i|\hat{\mathbf{u}}_i \sim MN(\Sigma(\Sigma + \mathbf{V}_i)^{-1}\hat{\mathbf{u}}_i, (\Sigma - \Sigma(\Sigma + \mathbf{V}_i)^{-1}\Sigma))$, where Σ is the variance-covariance matrix of \mathbf{u}_i , and $\mathbf{V}_i((ki + p) \times (ki + p))$ is the variance-covariance matrix of $\hat{\mathbf{u}}_i|\mathbf{u}_i \sim MN(\mathbf{u}_i, \mathbf{V}_i)$ which is obtained by \mathbf{J}^{-1} evaluated at current estimates of $\beta_r, \mathbf{U}_r, r = 1, \dots, k$,

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{J}_{a \times a}^{-1} & \dots & \mathbf{J}_{a \times b}^{-1} \\ \vdots & \ddots & \vdots \\ \mathbf{J}_{b \times a}^{-1} & \dots & \mathbf{J}_{b \times b}^{-1} \end{pmatrix}, \quad (4.13)$$

where $a = (ki + p - (k - 1))$, and $b = (ki + p)$. Now, the corrected random effect for $\hat{\mathbf{u}}_i$ based on RCM is the conditional expectation of $\mathbf{u}_i|\hat{\mathbf{u}}_i, E[\mathbf{u}_i|\hat{\mathbf{u}}_i] = \tilde{\boldsymbol{\mu}}_i$, and, $E[(\exp \mathbf{u}_i)|\hat{\mathbf{u}}_i] = \exp \left\{ \tilde{\boldsymbol{\mu}}_i + \frac{1}{2} \text{diag}(\tilde{\Sigma}_i) \right\}$, where $\tilde{\boldsymbol{\mu}}_i = \Sigma(\Sigma + \mathbf{V}_i)^{-1}\hat{\mathbf{u}}_i$, and $\tilde{\Sigma}_i = (\Sigma - \Sigma(\Sigma + \mathbf{V}_i)^{-1}\Sigma)$.

4.1.3 Bivariate Joint Model and H -Likelihood Function

For simplicity, we illustrate the multivariate joint model considering a bivariate ($k = 2$) Poisson-Normal model with $\mathbf{y}_1 | \mathbf{u} \sim \text{Pois}(\lambda)$, $\mathbf{y}_2 | \mathbf{u} \sim N(\mu, \mathbf{G})$, $\mathbf{y}_1 = (y_{11}, \dots, y_{1N_1})^T$, $\mathbf{y}_2 = (y_{21}, \dots, y_{2N_2})^T$, and $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)^T \sim BN(0, \mathbf{\Sigma})$, $\mathbf{\Sigma}$ is the 2×2 variance-covariance matrix of the bivariate latent random variable \mathbf{u} . The joint h -likelihood for the bivariate model can be expressed as,

$$h = \mathcal{L}_{(\theta, \phi; \mathbf{y}_1 | \mathbf{u}_1)}(\mathbf{y}_1 | \mathbf{u}_1) + \mathcal{L}_{(\theta, \phi; \mathbf{y}_2 | \mathbf{u}_2)}(\mathbf{y}_2 | \mathbf{u}_2) + \mathcal{L}_{(\theta, \phi; \mathbf{u})}(\mathbf{u}), \quad (4.14)$$

where,

$\mathcal{L}_{(\theta, \phi; \mathbf{y}_1 | \mathbf{u}_1)}(\mathbf{y}_1 | \mathbf{u}_1) = \mathbf{y}_1^T (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_1 \mathbf{U}_1) - \mathbf{1}^T \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_1 \mathbf{U}_1) - (\log \mathbf{y}_1)^T \mathbf{1}_{N_1 \times 1}$, $\mathbf{y}_1 = (y_{11}!, \dots, y_{1N_1}!)^T$, from (2.11), $\boldsymbol{\theta}_1 = (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_1 \mathbf{U}_1)$, $b(\boldsymbol{\theta}_1) = \mathbf{1}^T \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_1 \mathbf{U}_1)$, $a(\boldsymbol{\phi}_1) = 1$, $c(\mathbf{y}_1, \boldsymbol{\phi}_1) = (\log \mathbf{y}_1)^T \mathbf{1}_{N_1 \times 1}$, and

$$\begin{aligned} & \mathcal{L}_{(\theta, \phi; \mathbf{y}_2 | \mathbf{u}_2)}(\mathbf{y}_2 | \mathbf{u}_2) \\ &= -\frac{N_2}{2} \log(\det(\mathbf{G})) - \frac{1}{2} (\mathbf{y}_2 - (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2))^T \mathbf{G}^{-1} (\mathbf{y}_2 - (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2)) \\ &+ c_2, \end{aligned}$$

can be expressed in the form of (4.12) as

$$\begin{aligned} &= \mathbf{G}^{-1} \left\{ \mathbf{y}_2^T (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2) - \frac{1}{2} (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2)^T (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2) \right\} \\ &\quad - \frac{1}{2} \{ \mathbf{G}^{-1} \mathbf{y}_2^T \mathbf{y}_2 + \log(2\pi \mathbf{G}^{-1}) \}, \end{aligned}$$

where $\boldsymbol{\theta}_2 = (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2)$, $b(\boldsymbol{\theta}_2) = \frac{1}{2} (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2)^T (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2)$, $a(\boldsymbol{\phi}_2) = \mathbf{G}^{-1}$, and $c(\mathbf{y}_2, \boldsymbol{\phi}_2) = -\frac{1}{2} \{ \mathbf{G}^{-1} \mathbf{y}_2^T \mathbf{y}_2 + \log(2\pi \mathbf{G}^{-1}) \}$. The dimensions of the matrices are $\mathbf{y}_1 (N_1 \times 1)$, $\mathbf{y}_2 (N_2 \times 1)$, $\mathbf{X}_1 (N_1 \times p_1)$, $\mathbf{X}_2 (N_2 \times p_2)$, $\boldsymbol{\beta}_1 (p_1 \times 1)$, $\boldsymbol{\beta}_2 (p_2 \times 1)$, $\mathbf{Z}_1 (N_1 \times m)$, $\mathbf{Z}_2 (N_2 \times m)$, $\mathbf{U}_1 (m \times 1)$, $\mathbf{U}_2 (m \times 1)$, and $\mathbf{U} (2m \times 1)$.

The log-likelihood of random effect $\mathbf{U} = (\mathbf{U}_1, \mathbf{U}_2)^T$,

$$\mathcal{L}_{(\boldsymbol{\theta}, \boldsymbol{\phi}; \mathbf{U})}(\mathbf{U}) = -\frac{1}{2} \mathbf{U}^T \boldsymbol{\Sigma}^{-1} \mathbf{U} - \frac{m}{2} \log(\det(\boldsymbol{\Sigma})) + c_u, \quad (4.15)$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{22}^* \end{pmatrix}, c_2 = -\frac{N_2}{2} \log(2\pi), c_u = -\frac{m}{2} \log(2\pi).$$

Now, the quantities of the score function \mathcal{S} , $b'(\boldsymbol{\theta})$, $a(\boldsymbol{\phi})$, and $\nabla_{\mathcal{L}_U}^1$ in (2.13) can be expressed as,

$$b'(\boldsymbol{\theta}_1) = \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z}_1 \mathbf{U}_1), a(\boldsymbol{\phi}_1) = 1,$$

$$b'(\boldsymbol{\theta}_2) = (\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z}_2 \mathbf{U}_2), a(\boldsymbol{\phi}_2) = \mathbf{G}^{-1},$$

$$\nabla_{\mathcal{L}_{u_1}}^1 = -(\sigma_{11}^* \mathbf{U}_1 + \sigma_{12}^* \mathbf{U}_2), \text{ and } \nabla_{\mathcal{L}_{u_2}}^1 = -(\sigma_{21}^* \mathbf{U}_1 + \sigma_{22}^* \mathbf{U}_2), \text{ then,}$$

$$\begin{aligned} s(\boldsymbol{\theta}; \mathbf{y}) &= \begin{pmatrix} \mathbf{X}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^T \\ \mathbf{Z}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2^T \end{pmatrix} \begin{pmatrix} s(\boldsymbol{\theta}_1; \mathbf{y}_1) \\ s(\boldsymbol{\theta}_2; \mathbf{y}_2) \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \nabla_{\mathcal{L}_{u_1}}^1 \\ \nabla_{\mathcal{L}_{u_2}}^1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{X}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2^T \\ \mathbf{Z}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2^T \end{pmatrix} \begin{pmatrix} \frac{1}{a(\boldsymbol{\phi}_1)} (\mathbf{y}_1 - b'(\boldsymbol{\theta}_1)) \\ \frac{1}{a(\boldsymbol{\phi}_2)} (\mathbf{y}_2 - b'(\boldsymbol{\theta}_2)) \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \nabla_{\mathcal{L}_{u_1}}^1 \\ \nabla_{\mathcal{L}_{u_2}}^1 \end{pmatrix}, \end{aligned} \quad (4.16)$$

$$\text{where } \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{22}^* \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}^{-1}, \nabla'_{\mathcal{L}_U} = \begin{pmatrix} \nabla'_{\mathcal{L}_{u_1}} & \nabla'_{\mathcal{L}_{u_2}} \end{pmatrix}^T = -\mathbf{U}(\boldsymbol{\Sigma}^{-1} \otimes I_{m \times m}).$$

The components of (4.6) are as follows,

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_2 \end{pmatrix}, \nabla''_{\mathcal{L}_U} = -\boldsymbol{\Sigma}^{-1} \otimes I_{m \times m},$$

$$\mathbf{W}_r = \text{Diag} \left(\left((g'(\boldsymbol{\mu}_r))^T \right)^{-1} (\text{Var}(\mathbf{y}_r))^{-1} (g'(\boldsymbol{\mu}_r))^{-1} \right), r = 1, 2,$$

where $g'(\boldsymbol{\mu}_r)$ is the first partial derivative of link function $g(\boldsymbol{\mu}_r) = \log \boldsymbol{\lambda}_r$ (for Poisson, $\boldsymbol{\mu}_r = \boldsymbol{\lambda}_r$) with respect to $\boldsymbol{\mu}_r$. For $r = 1$,

$$g'(\boldsymbol{\mu}_r) = \frac{\partial}{\partial \lambda_r} (\log \lambda_r) \left(\frac{\partial \lambda_r}{\partial \boldsymbol{\mu}_r} \right) = \frac{1}{\lambda_r},$$

$$\mathbf{W}_r = \text{Diag} \left(\frac{1}{\lambda_r \left(\frac{1}{\lambda_r} \right)^2} \right) = \text{Diag}(\lambda_r).$$

Next, (4.18), and (4.19) are used to obtain MHLEs for $\boldsymbol{\psi} = (\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{22})$. As described in section 4.2.2, the estimates are fine-tuned through the bias correction approach using RCM by replacing $\mathbf{u}_i | \hat{\mathbf{u}}_i$ with $E[\mathbf{u}_i | \hat{\mathbf{u}}_i] = \tilde{\boldsymbol{\mu}}_i$, and $E[(\exp \mathbf{u}_i) | \hat{\mathbf{u}}_i] = \exp \left\{ \tilde{\boldsymbol{\mu}}_i + \frac{1}{2} \text{diag}(\tilde{\boldsymbol{\Sigma}}_i) \right\}$, where

$$\tilde{\boldsymbol{\mu}}_i = \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \boldsymbol{\nu}_i)^{-1} \hat{\mathbf{u}}_i, \tilde{\boldsymbol{\Sigma}}_i = (\boldsymbol{\Sigma} - \boldsymbol{\Sigma}(\boldsymbol{\Sigma} + \boldsymbol{\nu}_i)^{-1} \boldsymbol{\Sigma}), \mathbf{u}_i = (\mathbf{u}_{1i}, \mathbf{u}_{2i})^T,$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}, \text{ and } \boldsymbol{\nu}_i = \begin{pmatrix} \mathbf{J}_{(2i+p-1) \times (2i+p-1)}^{-1} & \mathbf{J}_{(2i+p-1) \times (2i+p)}^{-1} \\ \mathbf{J}_{(2i+p) \times (2i+p-1)}^{-1} & \mathbf{J}_{(2i+p) \times (2i+p)}^{-1} \end{pmatrix}.$$

4.1.4 MHLE Algorithm – Multivariate Joint Model

The MHLEs of fixed effects are estimated through iterative approximation as described below,

1. Initialize $\boldsymbol{\beta}_1^{(0)}, \boldsymbol{\beta}_2^{(0)}, \mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}$, and $\boldsymbol{\Sigma}^{(0)}$.
 1. Obtain $\boldsymbol{\beta}_1^{(0)}$, and $\boldsymbol{\beta}_2^{(0)}$ using Poisson regression without random effects.
 2. Obtain $\mathbf{u}_1^{(0)}, \mathbf{u}_2^{(0)}$ sampled from multivariate normal for a given $\boldsymbol{\Sigma}^{(0)}$.
2. Evaluate \mathcal{S} , and \mathcal{J} using (4.4), and (4.5).
3. Estimate $\hat{\boldsymbol{\beta}}^{(k)} = \left(\hat{\boldsymbol{\beta}}_1^{(k)}, \hat{\boldsymbol{\beta}}_2^{(k)} \right)^T$ and $\hat{\mathbf{u}}^{(k)} = \left(\hat{\mathbf{u}}_1^{(k)}, \hat{\mathbf{u}}_2^{(k)} \right)^T$ using

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k)} \\ \hat{\mathbf{u}}^{(k)} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{\beta}}^{(k-1)} \\ \hat{\mathbf{u}}^{(k-1)} \end{pmatrix} + (\mathcal{J}^{-1} \mathcal{S})|_{(\hat{\boldsymbol{\beta}}^{(k-1)}, \hat{\mathbf{u}}^{(k-1)})}, k = 1, 2, \dots$$

4. Update $\mathcal{J}^{(k)}$ using current estimates of $\boldsymbol{\beta}$, and \mathbf{u} , $\hat{\boldsymbol{\beta}}^{(k)} = \left(\hat{\boldsymbol{\beta}}_1^{(k)}, \hat{\boldsymbol{\beta}}_2^{(k)} \right)^T$ and $\hat{\mathbf{u}}^{(k)} = \left(\hat{\mathbf{u}}_1^{(k)}, \hat{\mathbf{u}}_2^{(k)} \right)^T$.

5. Estimate dispersion parameters $\boldsymbol{\psi}(= \boldsymbol{\Sigma}^{(k)})$ by replacing $\hat{\mathbf{u}}^{(k)}$ with $\hat{\mathbf{u}}^{c(k)} = \zeta_i^{(k)} \hat{\mathbf{u}}_i^{(k)}$ and by replacing $\exp(\hat{\mathbf{u}}^{(k)})$ with $\hat{\mathbf{u}}^{CExp(k)}$, where $\boldsymbol{\Sigma}^{(k-1)}$ is variance-covariance matrix of $\boldsymbol{\Sigma}$ at the $(k-1)^{th}$ iteration, $\boldsymbol{\nu}_i^{(k)} = (\mathbf{J}_{22}^{-1})^{(k)} = (\mathbf{Z}^T \mathbf{W} \mathbf{Z} + \mathbf{Q})^{-1} |_{\hat{\boldsymbol{\beta}}^{(k)}, \hat{\mathbf{u}}^{(k)}, \boldsymbol{\Sigma}^{(k-1)}}$ obtained from $\mathbf{J}^{(k)}$,

$$\boldsymbol{\Sigma}^{(k)} = \boldsymbol{\Sigma}^{(k-1)} + \left(-\frac{\partial^2 h_A}{\partial \boldsymbol{\Sigma}^T \partial \boldsymbol{\Sigma}} \right)^{-1} \frac{\partial h_A}{\partial \boldsymbol{\psi}} \Bigg|_{\boldsymbol{\Sigma} = \tilde{\boldsymbol{\Sigma}}^{(k-1)}},$$

where, $\zeta_i^{(k)} = \boldsymbol{\Sigma}^{(k-1)} \left(\boldsymbol{\Sigma}^{(k-1)} + \boldsymbol{\nu}_i^{(k)} \right)^{-1}$, so, $\zeta_i^{(k)} \hat{\mathbf{u}}_i^{(k)} = \boldsymbol{\Sigma}^{(k-1)} \left(\boldsymbol{\Sigma}^{(k-1)} + \boldsymbol{\nu}_i^{(k)} \right)^{-1} \hat{\mathbf{u}}_i^{(k)}$,

$$\hat{\mathbf{u}}^{CExp(k)} = \exp \left\{ \zeta_i^{(k)} \hat{\mathbf{u}}_i^{(k)} + \frac{1}{2} \text{diag} \left(\tilde{\boldsymbol{\Sigma}}_i^{(k)} \right) \right\},$$

$$\tilde{\boldsymbol{\Sigma}}_i^{(k)} = \left(\boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^{(k-1)} \left(\boldsymbol{\Sigma}^{(k-1)} + \boldsymbol{\nu}_i^{(k)} \right)^{-1} \boldsymbol{\Sigma}^{(k)} \right) = \boldsymbol{\Sigma}^{(k)} \left(1 - \zeta_i^{(k)} \right),$$

$$\frac{\partial h_A}{\partial \boldsymbol{\psi}} = -\frac{m}{2} \text{trace} \left((\boldsymbol{\Sigma}^{-1})^{(k-1)} \right) - \frac{1}{2} \mathbf{U}^T \mathbf{Q}' \mathbf{U} \Bigg|_{\hat{\boldsymbol{\beta}}^{(k)}, \hat{\mathbf{u}}^{(k)}} - \frac{1}{2} \text{trace} \left(\mathbf{J}_{22}^{-1} \mathbf{Q}' \right) \Bigg|_{\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}},$$

$$\frac{\partial^2 h_A}{\partial \boldsymbol{\psi}^T \partial \boldsymbol{\psi}} = \frac{m}{2} \text{trace} \left(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \right) - \frac{1}{2} \left(\mathbf{U}^T \mathbf{Q}'' \mathbf{U} \right) - \frac{1}{2} \text{trace} \left(\mathbf{J}_{22}^{-1} \mathbf{Q}'' - \mathbf{J}_{22}^{-1} \mathbf{Q}' \mathbf{J}_{22}^{-1} \mathbf{Q}' \right),$$

$$\mathbf{Q}' = -\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m} |_{\boldsymbol{\Sigma}^{(k-1)}}, \text{ and } \mathbf{Q}'' = 2\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \otimes I_{m \times m} |_{\boldsymbol{\Sigma}^{(k-1)}}.$$

Note: for two response variables with count data having bivariate normal \mathbf{u} , $\boldsymbol{\psi} = \boldsymbol{\Sigma}$.

6. Update h -likelihood by replacing $\hat{\mathbf{u}}^{(k)}$ with $\hat{\mathbf{u}}^{c(k)}$, and $\exp(\hat{\mathbf{u}}^{(k)})$ with $\hat{\mathbf{u}}^{CExp(k)}$.
7. Repeat steps 2 – 6 until it meets the convergence criteria

$$\max \left\{ \left(\hat{\boldsymbol{\beta}}^{(k)} - \hat{\boldsymbol{\beta}}^{(k-1)} \right), \left(\boldsymbol{\Sigma}^{(k)} - \boldsymbol{\Sigma}^{(k-1)} \right) \right\} < 10^{-5}.$$

4.2 Joint Modeling Through Shared Random Effects

Suppose $\mathbf{Y}_1, \dots, \mathbf{Y}_k$ is a vector of k outcomes of interest, which are measured on the number of individuals in small areas. Conducting joint analysis on multiple outcomes explains the

association among them and provide better results than separate analysis on each outcome.

Consider the joint model in SAE for $\mathbf{Y}_1, \dots, \mathbf{Y}_k$ outcomes

$$\begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_k \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta}_1 + \rho_1\mathbf{Z}\mathbf{u} \\ \mathbf{X}\boldsymbol{\beta}_2 + \rho_2\mathbf{Z}\mathbf{u} \\ \vdots \\ \mathbf{X}\boldsymbol{\beta}_k + \rho_k\mathbf{Z}\mathbf{u} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_k \end{pmatrix}, \quad (4.17)$$

where $\rho_1, \rho_2, \dots, \rho_k$ are the shared parameters between outcomes, \mathbf{X} is covariate information, $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_k$ are fixed effects coefficient vectors, $\boldsymbol{\beta}^{(\cdot)} = (\beta_1, \dots, \beta_p)$, p is the number of fixed effects for each outcome, \mathbf{Z} is the design matrix for random effects whole diagonal elements being one and off-diagonal elements being zero, $\mathbf{u} = (u_1, \dots, u_m)$ is a vector of random effects, and m is the number of small areas.

Now, we define the joint density for multiple outcomes, $(\mathbf{Y}_1, \dots, \mathbf{Y}_k)$

$$\begin{aligned} f(\mathbf{Y}_1, \dots, \mathbf{Y}_k) &= f(\mathbf{Y}_1, \dots, \mathbf{Y}_k | \mathbf{u}) \times f(\mathbf{u}) \\ &= f(\mathbf{Y}_1 | \mathbf{u}) \times f(\mathbf{Y}_2 | \mathbf{u}) \times \dots \times f(\mathbf{Y}_k | \mathbf{u}) \times f(\mathbf{u}) \end{aligned}$$

Here, we assume that the outcomes are conditionally independent, that is $\mathbf{Y}_1 | \mathbf{u}, \dots, \mathbf{Y}_k | \mathbf{u}$ are independent. Now, h -likelihood can be written as

$$h = \ell(\mathbf{Y}_1 | \mathbf{u}) + \dots + \ell(\mathbf{Y}_k | \mathbf{u}) + l(\mathbf{u}). \quad (4.18)$$

However, for the canonical GLM family, the matrices \mathcal{S} and \mathcal{J} can be expressed as

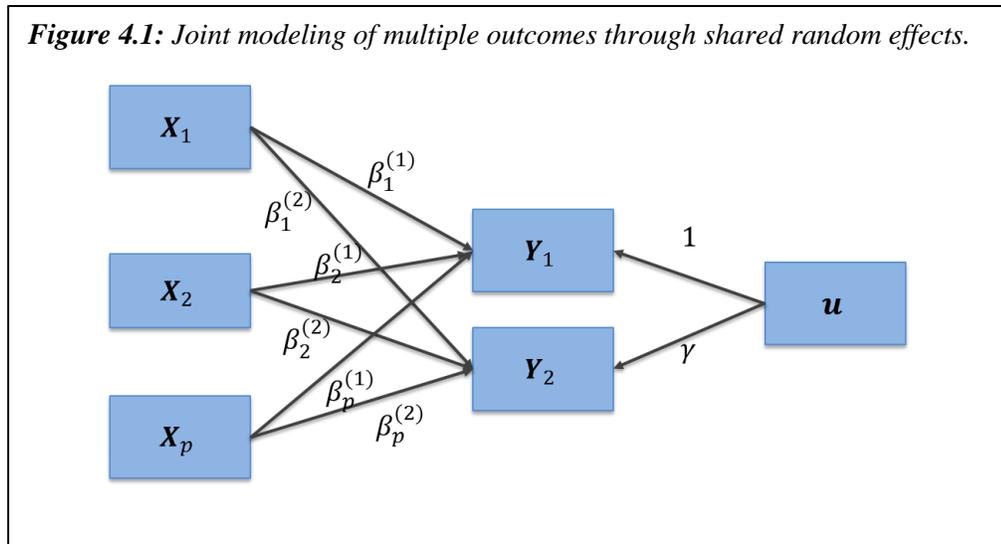
$$\mathcal{S} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Z} \end{pmatrix}^T \circ \begin{pmatrix} \mathbf{y} - \mathbf{1}_{kN \times 1} \circ \mathbf{b}'(\boldsymbol{\theta}) \\ \mathbf{y} - \boldsymbol{\gamma}_{kN \times 1} \circ \mathbf{b}'(\boldsymbol{\theta}) \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nabla}_{\ell_u}^1 \end{pmatrix}, \quad (4.19)$$

and

$$\mathcal{J} = \begin{pmatrix} \mathbf{X}^T \circ \boldsymbol{\mathcal{W}} \circ \mathbf{X} & \boldsymbol{\gamma} \circ \mathbf{X}^T \circ \boldsymbol{\mathcal{W}} \circ \mathbf{Z} \\ \boldsymbol{\gamma} \circ \mathbf{Z}^T \circ \boldsymbol{\mathcal{W}} \circ \mathbf{X} & \boldsymbol{\gamma} \circ \boldsymbol{\gamma} \circ (\mathbf{Z}^T \boldsymbol{\mathcal{W}} \mathbf{Z}) + \boldsymbol{\nabla}_{\ell_u}^2 \end{pmatrix}, \quad (4.20)$$

where \circ indicates the componentwise multiplication, also known as the Hadamard product. Here, the matrices and vectors are $\mathbf{X} = (\mathbf{X}_1 \ \dots \ \mathbf{X}_k)^T$, $\mathbf{Z} = (\mathbf{Z}_1 \ \dots \ \mathbf{Z}_k)^T$, $\mathbf{W} = (\mathbf{W}_1 \ \dots \ \mathbf{W}_k)^T$, $\boldsymbol{\gamma} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_k \end{pmatrix}$, $\boldsymbol{\mathcal{W}} = \begin{pmatrix} \mathbf{W}_1 & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \mathbf{W}_k \end{pmatrix}$, and $\mathbf{W}_r = \text{Diag} \left(\frac{1}{(g'(\boldsymbol{\mu}_r))^T \text{Var}(\mathbf{y}_r) (g'(\boldsymbol{\mu}_r))} \right)$, $r = 1, \dots, k$, $\boldsymbol{\gamma}$ is a vector of k shared parameters, \mathbf{X} and \mathbf{Z} are vectors of k matrices.

We illustrate the joint modeling based on shared random effects using two outcomes, \mathbf{Y}_1 , and \mathbf{Y}_2 , which is displayed in Figure 4.1 with the shared parameter $\boldsymbol{\gamma}$, that explains the association between \mathbf{Y}_1 , and \mathbf{Y}_2 . $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{X}_1, \mathbf{X}_2$, and \mathbf{X}_p are vectors of $N \times 1$ if both outcome variables have the same number of observations. The regression coefficients $\beta_p^{(1)}$, and $\beta_p^{(2)}$ are fixed effects for p th covariate $\mathbf{X}_p, p = 1, \dots, p$ of \mathbf{Y}_1 , and \mathbf{Y}_2 . While $\gamma > 0$ ($\gamma < 0$) indicates the existence of a positive (negative) association between \mathbf{Y}_1 , and \mathbf{Y}_2 , and $\gamma = 0$ indicates no association between \mathbf{Y}_1 , and \mathbf{Y}_2 .



In this section, we consider a joint model with two binary response variables, $\mathbf{y} = (y_{1ij}, y_{2ij})'$, $i = 1, \dots, n_j, j = 1, \dots, m$ where, $y_{1ij}|u_{1i} \sim \text{Bernoulli}(p_1)$, $y_{2ij}|u_{2i} \sim \text{Bernoulli}(p_2)$, and $u_{1i}, u_{2i} \sim N(0, \theta)$. The auxiliary information

$$X = \begin{pmatrix} \mathbf{x}_{1ij}' & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2ij}' \end{pmatrix},$$

where $\mathbf{x}'_{1ij} = (x_{1ij1}, \dots, x_{1ijp_1})'$, and $\mathbf{x}'_{2ij} = (x_{2ij1}, \dots, x_{2ijp_2})'$ are vectors of p_1 and p_2 covariates of y_{1ij} and y_{2ij} , respectively. The logit model for y_{1ij} and y_{2ij}

$$P(y_{1ij} = 1|u_{1i}) = \text{logit}^{-1}(\mathbf{x}_{1ij}^t \boldsymbol{\beta} + u_{1i}) = \text{logit}^{-1}(\mathbf{x}_{1ij}^t \boldsymbol{\beta} + u_i),$$

$$P(y_{2ij} = 1|u_{2i}) = \text{logit}^{-1}(\mathbf{x}_{2ij}^t \boldsymbol{\delta} + u_{2i}) = \text{logit}^{-1}(\mathbf{x}_{2ij}^t \boldsymbol{\delta} + \gamma u_i),$$

where $y_r = (y_{rij}, 1 \leq i \leq m_r, 1 \leq j \leq n_i), r = 1, 2$ is the binary response vectors for 1st and 2nd outcomes of interest, which are independent given the random effects u_1, \dots, u_{m_r} , and γ is the shared parameter. Let $\mathbf{u} = (u_i)_{1 \leq i \leq m_1}$, and $\boldsymbol{\beta}_1 = (\beta_{1k})_{1 \leq k \leq p_1}$ and $\boldsymbol{\beta}_2 = (\beta_{2k})_{1 \leq k \leq p_2}$ are vectors of unknown fixed effects of y_1 and y_2 . From (4.18), for two outcome variables

$$\begin{aligned} h &= \ell_{y_1} + \ell_{y_2} + \ell_u \\ &= \mathbf{y}_1^T (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Zu}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Zu})) + \mathbf{y}_2^T (\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Zu}) \\ &\quad - \mathbf{1}^T \log(1 + \exp(\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Zu})) - \frac{1}{2} \mathbf{u}^T \boldsymbol{\theta}^{-1} \mathbf{u} - \frac{m}{2} \log(\det(\boldsymbol{\theta})) + c, \end{aligned} \quad (4.21)$$

where $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ are vectors of fixed effects, $\mathbf{u}, \gamma \mathbf{u}$ are random effects of response variables \mathbf{Y}_1 and \mathbf{Y}_2 respectively, with shared parameter γ . $\mathbf{X}_1, \mathbf{X}_2$ are corresponding design matrices, and $\boldsymbol{\theta}$ is the variance-covariance matrix of random effect \mathbf{u} . When $\gamma = 0$, \mathbf{Y}_1 and \mathbf{Y}_2 are not associated, and when $\gamma > 0$, they are positively associated. The parameters of fixed effects and random effects are estimated using the Newton Raphson approximation taking the partial derivative of joint h -likelihood with respect to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and \mathbf{u} . The dispersion parameters, $\boldsymbol{\theta}$, and γ are estimated using the adjusted profile h -likelihood (Ha et al., 2017; Lee et al., 2017).

4.2.1 Parameter Estimation of Fixed Effects and Random Effects

The partial derivative of h -likelihood equation (4.21) with respect to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$, and \mathbf{u} , respectively,

$$\begin{aligned}
\frac{\partial h}{\partial \boldsymbol{\beta}_1} &= [\mathbf{X}_1^T \mathbf{y}_1 - \mathbf{X}_1^T (1 + \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u}))^{-1} \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u})]_{p \times 1} \\
&= \mathbf{X}_1^T \mathbf{y}_1 - \mathbf{X}_1^T \boldsymbol{\pi}_1 = \mathbf{X}_1^T (\mathbf{y}_1 - \boldsymbol{\pi}_1), \\
\frac{\partial h}{\partial \boldsymbol{\beta}_2} &= [\mathbf{X}_2^T \mathbf{y}_2 - \mathbf{X}_2^T (1 + \exp(\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Z} \mathbf{u}))^{-1} \exp(\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Z} \mathbf{u})]_{p \times 1} \\
&= \mathbf{X}_2^T \mathbf{y}_2 - \mathbf{X}_2^T \boldsymbol{\pi}_2 = \mathbf{X}_2^T (\mathbf{y}_2 - \boldsymbol{\pi}_2), \\
\frac{\partial h}{\partial \mathbf{u}} &= \mathbf{Z}^T \mathbf{y}_1 - \mathbf{Z}^T \boldsymbol{\pi}_1 + \gamma \mathbf{Z}^T \mathbf{y}_2 - \gamma \mathbf{Z}^T \boldsymbol{\pi}_2 - \boldsymbol{\theta}^{-1} \mathbf{u} \\
&= \mathbf{Z}^T (\mathbf{y}_1 + \gamma \mathbf{y}_2 - (\boldsymbol{\pi}_1 + \gamma \boldsymbol{\pi}_2)) - \boldsymbol{\theta}^{-1} \mathbf{u},
\end{aligned}$$

where $\boldsymbol{\pi}_1 = 1/(1 + \exp(-\mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z} \mathbf{u}))$, and $\boldsymbol{\pi}_2 = 1/(1 + \exp(-\mathbf{X}_2 \boldsymbol{\beta}_2 - \gamma \mathbf{Z} \mathbf{u}))$. Now, the score function of joint h -likelihood can be written as

$$\boldsymbol{\mathcal{S}}(\boldsymbol{\tau}) = \begin{pmatrix} \frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \boldsymbol{\delta}} \\ \frac{\partial h}{\partial \mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_\beta^T (\mathbf{y}_1 - \boldsymbol{\pi}_1) \\ \mathbf{X}_\delta^T (\mathbf{y}_2 - \boldsymbol{\pi}_2) \\ \mathbf{Z}^T (\mathbf{y}_1 + \gamma \mathbf{y}_2 - (\boldsymbol{\pi}_1 + \gamma \boldsymbol{\pi}_2)) - \boldsymbol{\theta}^{-1} \mathbf{u} \end{pmatrix}, \quad (4.22)$$

where $\boldsymbol{\tau} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u})$ are vectors for fixed effects coefficients of output variables y_1, y_2 , and random effects, respectively.

Now, consider the Fisher information matrix of joint h -likelihood

$$\boldsymbol{\mathcal{J}} = \begin{bmatrix} -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_1^T \partial \boldsymbol{\beta}_1} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_1^T \partial \boldsymbol{\beta}_2} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_1^T \partial \mathbf{u}} \\ -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_1} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_2^T \partial \boldsymbol{\beta}_2} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta}_2^T \partial \mathbf{u}} \\ -\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \boldsymbol{\beta}_1} & -\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \boldsymbol{\beta}_2} & -\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \mathbf{u}} \end{bmatrix}. \quad (4.23)$$

The elements of $\boldsymbol{\mathcal{J}}$ are obtained as in chapter 3.3 taking the partial derivative of score function $\boldsymbol{\mathcal{S}}(\boldsymbol{\tau})$ with respect to $\boldsymbol{\tau}$. Taking partial derivatives of $\boldsymbol{\pi}_1$, and $\boldsymbol{\pi}_2$ with respect to $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and \mathbf{u} ,

$$\frac{\partial \boldsymbol{\pi}_1}{\partial \mathbf{u}} = -(1 + \exp(-\mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z} \mathbf{u}))^{-1} \exp(-\mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z} \mathbf{u}) \mathbf{Z} = -\boldsymbol{\pi}_1 (1 - \boldsymbol{\pi}_1) \mathbf{Z},$$

$$\frac{\partial \pi_1}{\partial \beta_1} = -(1 + \exp(-X_1 \beta_1 - Z u))^{-1} \exp(-X_1 \beta_1 - Z u) X_1 = -\pi_1(1 - \pi_1) X_1,$$

$$\frac{\partial \pi_2}{\partial u} = -\gamma(1 + \exp(-X_2 \beta_2 - \gamma Z u))^{-1} \exp(-X_2 \beta_2 - \gamma Z u) = -\gamma \pi_2(1 - \pi_2) Z,$$

$$\frac{\partial \pi_2}{\partial \beta_2} = -(1 + \exp(-X_2 \beta_2 - \gamma Z u))^{-1} \exp(-X_2 \beta_2 - \gamma Z u) X_2 = -\pi_2(1 - \pi_2) X_2.$$

The diagonal elements of the matrix \mathcal{J} ,

$$\begin{aligned} -\frac{\partial^2 h}{\partial \beta_1^T \partial \beta_1} &= X_1^T (1 + \exp(-X_1 \beta_1 - Z u))^{-1} \exp(-X_1 \beta_1 - Z u) X_1 \\ &= X_1^T \pi_1 (1 - \pi_1) X_1 \\ &= X_1^T W_1 X_1, \end{aligned}$$

$$\begin{aligned} -\frac{\partial^2 h}{\partial \beta_2^T \partial \beta_2} &= X_2^T (1 + \exp(-X_2 \beta_2 - \gamma Z u))^{-1} \exp(-X_2 \beta_2 - \gamma Z u) X_2 \\ &= X_2^T \pi_2 (1 - \pi_2) X_2 \\ &= X_2^T W_2 X_2, \end{aligned}$$

$$\begin{aligned} -\frac{\partial^2 h}{\partial u^T \partial u} &= Z^T \pi_1 (1 - \pi_1) Z + \gamma^2 Z^T \pi_2 (1 - \pi_2) Z + \theta^{-1} \\ &= Z^T W_1 Z + Z^T (\gamma^2 W_2) Z + \theta^{-1}. \end{aligned}$$

The off-diagonal elements of the matrix \mathcal{J} ,

$$\begin{aligned} -\frac{\partial^2 h}{\partial \beta_1 \partial \beta_2} &= -\frac{\partial^2 h}{\partial \beta_2 \partial \beta_1} = 0, \\ -\frac{\partial^2 h}{\partial u \partial \beta_1} &= -\frac{\partial h}{\partial u} (X_1^T (y_1 - \pi_1)) = X_1^T \pi_1 (1 - \pi_1) Z = X_1^T W_1 Z, \\ -\frac{\partial^2 h}{\partial u \partial \beta_2} &= -\frac{\partial}{\partial u} (X_2^T (y_2 - \pi_2)) = X_2^T \gamma \pi_2 (1 - \pi_2) Z = X_2^T (\gamma W_2) Z, \\ -\frac{\partial^2 h}{\partial \beta_1 \partial u} &= -\frac{\partial}{\partial \beta_1} (Z^T (y_1 + \gamma y_2 - (\pi_1 + \gamma \pi_2)) - \theta^{-1} u) \\ &= Z^T \pi_1 (1 - \pi_1) X_1 = Z^T W_1 X_1, \end{aligned}$$

$$\begin{aligned}
-\frac{\partial^2 h}{\partial \boldsymbol{\beta}_2 \partial \mathbf{u}} &= -\frac{\partial}{\partial \boldsymbol{\beta}_2} (\mathbf{Z}^T (\mathbf{y}_1 + \gamma \mathbf{y}_2 - (\boldsymbol{\pi}_1 + \gamma \boldsymbol{\pi}_2)) - \boldsymbol{\theta}^{-1} \mathbf{u}) \\
&= \mathbf{Z}^T \gamma \boldsymbol{\pi}_2 (\mathbf{1} - \boldsymbol{\pi}_2) \mathbf{X}_2 = \mathbf{Z}^T (\gamma \mathbf{W}_2) \mathbf{X}_2.
\end{aligned}$$

Using the above quantities \mathcal{S} (eqn. 4.20) and \mathcal{J} (eqn. 4.21) can be updated, where

$$\begin{aligned}
\mathbf{X} &= \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} 1 \\ \gamma \end{pmatrix}^T, \quad \boldsymbol{W} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{0} \\ \mathbf{0} & \gamma^2 \mathbf{W}_2 \end{pmatrix}, \\
\boldsymbol{\beta} &= \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}, \quad \boldsymbol{w}_r = \boldsymbol{\pi}_r (\mathbf{1} - \boldsymbol{\pi}_r), \quad \boldsymbol{v}_{\ell_u}^1 = -\boldsymbol{\theta}^{-1} \mathbf{u}, \quad \boldsymbol{v}_{\ell_u}^2 = -\boldsymbol{\theta}^{-1}.
\end{aligned}$$

Now, the MHLEs are obtained using Newton Raphson approximation,

$$\hat{\boldsymbol{\tau}}^{(k+1)} = \hat{\boldsymbol{\tau}}^{(k)} + (\mathcal{J}^{-1} \mathcal{S}(\boldsymbol{\tau}))|_{\boldsymbol{\tau}=\hat{\boldsymbol{\tau}}^{(k)}}, \quad (4.24)$$

where $\boldsymbol{\tau} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \mathbf{u})$, and $\hat{\boldsymbol{\tau}} = (\widehat{\boldsymbol{\beta}}_1, \widehat{\boldsymbol{\beta}}_2, \widehat{\mathbf{u}})$.

However, since this is a canonical GLM family, we can easily obtain \mathcal{S} and \mathcal{J} using the equations (4.19) and (4.20). Consider the h -likelihood

$$\begin{aligned}
h &= \mathbf{y}_1^T (\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u})) + \mathbf{y}_2^T (\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Z} \mathbf{u}) - \mathbf{1}^T \log(1 + \\
&\quad \exp(\mathbf{X}_2 \boldsymbol{\beta}_2 + \gamma \mathbf{Z} \mathbf{u})) - \frac{1}{2} \mathbf{u}^T \boldsymbol{\vartheta}^{-1} \mathbf{u} - \frac{m}{2} \log(\det(\boldsymbol{\vartheta})) + c,
\end{aligned}$$

where $\boldsymbol{\theta}_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u}$, $b(\boldsymbol{\theta}_1) = \log(1 + \exp(\mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{Z} \mathbf{u}))$, $\boldsymbol{\theta}_2 = \mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z} \mathbf{u}$, $b(\boldsymbol{\theta}_2) = \log(1 + \exp(\mathbf{X}_2 \boldsymbol{\beta}_2 + \mathbf{Z} \mathbf{u}))$, and $g(\boldsymbol{\mu}_r) = \log \frac{\pi_r}{1 - \pi_r} = \log \frac{\mu_r}{1 - \mu_r}$, $r = 1, 2$. Now, \mathbf{X} , \mathbf{Z} , \mathbf{y} , $\boldsymbol{\gamma}$, $\boldsymbol{v}_{\ell_u}^1$, and $\boldsymbol{v}_{\ell_u}^2$ in (4.19) and (4.20) are

$$\boldsymbol{\gamma} = \begin{pmatrix} 1 \\ \gamma \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \mathbf{Z}_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}, \quad \boldsymbol{v}_{\ell_u}^1 = -\boldsymbol{\vartheta}^{-1} \mathbf{u}, \quad \boldsymbol{v}_{\ell_u}^2 = -\boldsymbol{\vartheta}^{-1},$$

and taking partial derivatives of $b(\boldsymbol{\theta}_1)$ and $b(\boldsymbol{\theta}_2)$ with respect to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$

$$\begin{aligned}
b'(\boldsymbol{\theta}_1) &= \frac{\partial}{\partial \boldsymbol{\theta}_1} \mathbf{1}^T \log(1 + \exp(\boldsymbol{\theta}_1)) = (1 + \exp(-\boldsymbol{\theta}_1))^{-1} \\
&= (1 + \exp(-\mathbf{X}_1 \boldsymbol{\beta}_1 - \mathbf{Z}_1 \mathbf{u})),
\end{aligned}$$

$$\begin{aligned}
b'(\boldsymbol{\theta}_2) &= \frac{\partial}{\partial \boldsymbol{\theta}_2} \mathbf{1}^T \log(1 + \exp(\boldsymbol{\theta}_2)) = \gamma(1 + \exp(-\boldsymbol{\theta}_2))^{-1} \\
&= \gamma(1 + \exp(-\mathbf{X}_2\boldsymbol{\beta}_2 - \gamma\mathbf{Z}_2\mathbf{u})).
\end{aligned}$$

Next, the link function

$$g(\boldsymbol{\mu}_r) = \log \frac{\boldsymbol{\pi}_r}{\mathbf{1} - \boldsymbol{\pi}_r} = \log \frac{\boldsymbol{\mu}_r}{\mathbf{1} - \boldsymbol{\mu}_r} \Rightarrow g'(\boldsymbol{\mu}_r) = (\boldsymbol{\mu}_r(\mathbf{1} - \boldsymbol{\mu}_r))^{-1},$$

and $\text{Var}(\mathbf{y}_r) = \boldsymbol{\mu}_r(\mathbf{1} - \boldsymbol{\mu}_r)$, $r = 1, 2$. Then the weight matrices \boldsymbol{W} and \boldsymbol{W}

$$\begin{aligned}
\boldsymbol{W}_r &= \text{Diag} \left(\frac{1}{(g'(\boldsymbol{\mu}_r))^T \text{Var}(\mathbf{y}_r) (g'(\boldsymbol{\mu}_r))} \right) = \text{Diag}(\boldsymbol{\mu}_r(\mathbf{1} - \boldsymbol{\mu}_r)), \\
\boldsymbol{w} &= \begin{pmatrix} \boldsymbol{W}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{W}_2 \end{pmatrix}, \quad \boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_1 \\ \boldsymbol{W}_2 \end{pmatrix}.
\end{aligned}$$

Now, the above expressions can be used to obtain (4.19) and (4.20) to obtain MHLEs of fixed effects and random effects.

4.2.2 Parameter Estimation of Dispersion Parameters

In this model, the dispersion parameters would be $\boldsymbol{\vartheta} = (\sigma^2, \gamma)$. The MHLEs of σ^2 and γ are obtained using the adjusted h -likelihood by solving the score equations $\partial h_A / \partial \sigma^2 = \partial h_A / \partial \gamma = 0$. First, consider the equation (3.5) to estimate σ^2 at current estimates $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$

$$\frac{\partial h_A}{\partial \sigma^2} = \frac{\partial h}{\partial \sigma^2} \Big|_{\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace} \left(\boldsymbol{J}^{-1} \frac{\partial \boldsymbol{J}}{\partial \sigma^2} \right) \Big|_{\hat{\boldsymbol{\tau}}}. \quad (4.25)$$

Note that $\partial \ell_{1ij} / \partial \sigma^2 = \partial \ell_{2ij} / \partial \sigma^2 = 0$, the score function can be written as

$$\frac{\partial h}{\partial \sigma^2} = 0 + \frac{\partial}{\partial \sigma^2} \left(\sum_{i=1}^m \ell_{3i} \right).$$

The log-likelihood of $u_i \sim N(0, \sigma^2)$ for small area i

$$\ell_{3i} = -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m u_i^2 = -\frac{m}{2} \log 2\pi - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{u},$$

where $\sigma^2 I_{m \times m}$. Now, from $h = \sum_{ij} \ell_{1ij} + \sum_{ij} \ell_{2ij} + \sum_i \ell_{2i}$, $\partial h / \partial \sigma^2$ given $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$, $\mathbf{u} = \hat{\mathbf{u}}$ can be represented as

$$\begin{aligned} \left. \frac{\partial h}{\partial \sigma^2} \right|_{\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}} &= 0 + \frac{\partial}{\partial \sigma^2} \left(-\frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^m \hat{u}_i^2 \right) \\ &= \left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\boldsymbol{\tau}}}. \end{aligned}$$

The partial derivative of the observed information matrix with respect to σ^2 given $\boldsymbol{\tau} = \hat{\boldsymbol{\tau}}$

$$\begin{aligned} \left. \frac{\partial \mathcal{J}}{\partial \sigma^2} \right|_{\hat{\boldsymbol{\tau}}} &= \frac{\partial}{\partial \sigma^2} \begin{bmatrix} -\frac{\partial^2 h}{\partial \boldsymbol{\beta}^2} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \boldsymbol{\delta}} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} \\ -\frac{\partial^2 h}{\partial \boldsymbol{\delta} \partial \boldsymbol{\beta}} & -\frac{\partial^2 h}{\partial \boldsymbol{\delta}^2} & -\frac{\partial^2 h}{\partial \boldsymbol{\delta} \partial \mathbf{u}} \\ -\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} & -\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\delta}} & -\frac{\partial^2 h}{\partial \mathbf{u}^2} \end{bmatrix}, \\ &= \frac{\partial}{\partial \sigma^2} \begin{bmatrix} \mathbf{X}_\beta^T \mathbf{W}_1 \mathbf{X}_\beta & \mathbf{0} & \mathbf{X}_\beta^T \mathbf{W}_1 \mathbf{Z} \\ \mathbf{0} & \mathbf{X}_\delta^T \mathbf{W}_2 \mathbf{X}_\delta & \mathbf{X}_\delta^T (\gamma \mathbf{W}_2) \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W}_1 \mathbf{X}_\beta & \mathbf{Z}^T (\gamma \mathbf{W}_2) \mathbf{X}_\delta & \mathbf{Z}^T \mathbf{W}_1 \mathbf{Z} + \mathbf{Z}^T (\gamma^2 \mathbf{W}_2) \mathbf{Z} + (\sigma^2)^{-1} \end{bmatrix}, \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -(\sigma^2)^{-2} I_{m \times m} \end{bmatrix}. \end{aligned}$$

From (4.25)

$$\begin{aligned} \left. \frac{\partial h_A}{\partial \sigma^2} \right|_{\hat{\boldsymbol{\tau}}} &= \left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace} \left(\begin{bmatrix} \mathcal{J}_{11}^* & \mathcal{J}_{12}^* & \mathcal{J}_{13}^* \\ \mathcal{J}_{21}^* & \mathcal{J}_{22}^* & \mathcal{J}_{23}^* \\ \mathcal{J}_{31}^* & \mathcal{J}_{32}^* & \mathcal{J}_{33}^* \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -(\sigma^2)^{-2} I_{m \times m} \end{bmatrix} \right) \Big|_{\hat{\boldsymbol{\tau}}} \\ &= \left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\boldsymbol{\tau}}} - \frac{1}{2} \text{trace} \left(\begin{bmatrix} \mathbf{0} & \mathbf{0} & -\mathcal{J}_{13}^* (\sigma^2)^{-2} I_{m \times m} \\ \mathbf{0} & \mathbf{0} & -\mathcal{J}_{23}^* (\sigma^2)^{-2} I_{m \times m} \\ \mathbf{0} & \mathbf{0} & -\mathcal{J}_{33}^* (\sigma^2)^{-2} I_{m \times m} \end{bmatrix} \right) \Big|_{\hat{\boldsymbol{\tau}}} \\ &= \left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\boldsymbol{\tau}}} + \frac{1}{2\sigma^4} \text{trace}(\mathcal{J}_{33}^*) \Big|_{\hat{\boldsymbol{\tau}}} \end{aligned}$$

Set $\partial h_A / \partial \sigma^2 = 0$,

$$\left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\tau}} + \frac{1}{2\sigma^4} \text{trace}(\mathcal{J}_{33}^*) \Big|_{\hat{\tau}} = 0.$$

Thus,

$$\widehat{\sigma^2} = \frac{1}{m} \left(\sum_{i=1}^m \hat{u}_i^2 \right) \Big|_{\hat{\tau}} + \frac{1}{m} \text{trace}(\mathcal{J}_{33}^*) \Big|_{\hat{\tau}}. \quad (4.26)$$

MHLE of σ^2 is obtained using (4.26). Now, the MHLE of γ is obtained using the partial derivative of (3.5) with respect to γ

$$\frac{\partial h_A}{\partial \gamma} = \frac{\partial h}{\partial \gamma} \Big|_{\hat{\tau}} - \frac{1}{2} \text{trace} \left(\mathcal{J}^{-1} \frac{\partial \mathcal{J}}{\partial \gamma} \right) \Big|_{\hat{\tau}}. \quad (4.27)$$

Consider the first term of (4.27)

$$\begin{aligned} \frac{\partial h}{\partial \gamma} \Big|_{\hat{\tau}} &= (\mathbf{Z}\mathbf{u})^T \mathbf{y}_2 - (\mathbf{Z}\mathbf{u})^T (\mathbf{1} + \exp(\mathbf{X}_\delta \boldsymbol{\delta} + \gamma \mathbf{Z}\mathbf{u}))^{-1} \exp(\mathbf{X}_\delta \boldsymbol{\delta} + \gamma \mathbf{Z}\mathbf{u}) \\ &= ((\mathbf{Z}\mathbf{u})^T \mathbf{y}_2 - (\mathbf{Z}\mathbf{u})^T \boldsymbol{\pi}_2) \Big|_{\hat{\tau}} = (\mathbf{Z}\mathbf{u})^T (\mathbf{y}_2 - \boldsymbol{\pi}_2) \Big|_{\hat{\tau}} \end{aligned} \quad (4.28)$$

The second term of (4.27)

$$\begin{aligned} \frac{\partial \mathcal{J}}{\partial \gamma} &= \frac{\partial}{\partial \gamma} \begin{bmatrix} \mathbf{X}_\beta^T \mathbf{W}_1 \mathbf{X}_\beta & \mathbf{0} & \mathbf{X}_\beta^T \mathbf{W}_1 \mathbf{Z} \\ \mathbf{0} & \mathbf{X}_\delta^T \mathbf{W}_2 \mathbf{X}_\delta & \gamma \mathbf{X}_\delta^T \mathbf{W}_2 \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W}_1 \mathbf{X}_\beta & \gamma \mathbf{Z}^T \mathbf{W}_2 \mathbf{X}_\delta & \mathbf{Z}^T (\mathbf{W}_1 + \gamma^2 \mathbf{W}_2) \mathbf{Z} + \boldsymbol{\theta}^{-1} \end{bmatrix} \Big|_{\hat{\tau}} \\ &= \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_\delta^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{X}_\delta & \gamma \mathbf{X}_\delta^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{Z} + \mathbf{X}_\delta^T \mathbf{W}_2 \mathbf{Z} \\ \mathbf{0} & \gamma \mathbf{Z}^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{X}_\delta + \mathbf{Z}^T \mathbf{W}_2 \mathbf{X}_\delta & \mathbf{Z}^T \left(\gamma^2 \frac{\partial \mathbf{W}_2}{\partial \gamma} + 2\gamma \mathbf{W}_2 \right) \mathbf{Z} \end{bmatrix} \Big|_{\hat{\tau}}, \end{aligned} \quad (4.29)$$

where,

$$\frac{\partial \boldsymbol{\pi}_2}{\partial \gamma} = -(1 + \exp(-\mathbf{X}_\delta \boldsymbol{\delta} - \gamma \mathbf{Z}\mathbf{u}))^{-1} \exp(-\mathbf{X}_\delta \boldsymbol{\delta} - \gamma \mathbf{Z}\mathbf{u}) \mathbf{Z}\mathbf{u} = -\boldsymbol{\pi}_2 (\mathbf{1} - \boldsymbol{\pi}_2) \mathbf{Z}\mathbf{u}$$

then,

$$\frac{\partial \mathbf{W}_2}{\partial \gamma} = \frac{\partial}{\partial \gamma} (\boldsymbol{\pi}_2 (\mathbf{1} - \boldsymbol{\pi}_2)) = -\boldsymbol{\pi}_2 \frac{\partial \boldsymbol{\pi}_2}{\partial \gamma} + (\mathbf{1} - \boldsymbol{\pi}_2) \frac{\partial \boldsymbol{\pi}_2}{\partial \gamma}$$

$$= \boldsymbol{\pi}_2(\mathbf{1} - \boldsymbol{\pi}_2)(2\boldsymbol{\pi}_2 - \mathbf{1})\mathbf{Z}\mathbf{u}.$$

Now, we take the partial derivative of the (4.27) with respect to γ to obtain the Hessian matrix \mathbf{J}_A

$$\mathbf{J}_A = \frac{\partial^2 h_A}{\partial \gamma^2} = \frac{\partial^2 h}{\partial \gamma^2} \Big|_{\hat{\tau}} - \frac{1}{2} \frac{\partial}{\partial \gamma} \left(\text{trace} \left(\mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \gamma} \right) \right) \Big|_{\hat{\tau}}. \quad (4.30)$$

The expressions (4.29) and (4.30) will be used in the Newton-Raphson procedure to obtain MLE of γ

$$\begin{aligned} \frac{\partial^2 h}{\partial \gamma^2} \Big|_{\hat{\tau}} &= \frac{\partial}{\partial \gamma} (\mathbf{Z}\mathbf{u})^T (\mathbf{y}_2 - \boldsymbol{\pi}_2) \Big|_{\hat{\tau}} = -(\mathbf{Z}\mathbf{u})^T \frac{\partial}{\partial \gamma} \boldsymbol{\pi}_2 \Big|_{\hat{\tau}} \\ \frac{\partial^2 h}{\partial \gamma^2} \Big|_{\hat{\tau}} &= (\mathbf{Z}\mathbf{u})^T \boldsymbol{\pi}_2 (\mathbf{1} - \boldsymbol{\pi}_2) \mathbf{Z}\mathbf{u} \Big|_{\hat{\tau}}. \end{aligned} \quad (4.31)$$

The second term of (4.30)

$$\begin{aligned} &\frac{\partial}{\partial \gamma} \left(\text{trace} \left(\mathbf{J}^{-1} \frac{\partial \mathbf{J}}{\partial \gamma} \right) \right) \Big|_{\hat{\tau}} = \\ &\frac{\partial}{\partial \gamma} \left(\text{trace} \left(\begin{bmatrix} \mathbf{J}_{11}^* & \mathbf{J}_{12}^* & \mathbf{J}_{13}^* \\ \mathbf{J}_{21}^* & \mathbf{J}_{22}^* & \mathbf{J}_{23}^* \\ \mathbf{J}_{31}^* & \mathbf{J}_{32}^* & \mathbf{J}_{33}^* \end{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_\delta^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{X}_\delta & \gamma \mathbf{X}_\delta^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{Z} + \mathbf{X}_\delta^T \mathbf{W}_2 \mathbf{Z} \\ \mathbf{0} & \gamma \mathbf{Z}^T \frac{\partial \mathbf{W}_2}{\partial \gamma} \mathbf{X}_\delta + \mathbf{Z}^T \mathbf{W}_2 \mathbf{X}_\delta & \mathbf{Z}^T \left(\gamma^2 \frac{\partial \mathbf{W}_2}{\partial \gamma} + 2\gamma \mathbf{W}_2 \right) \mathbf{Z} \end{bmatrix} \right) \right) \Big|_{\hat{\tau}}. \end{aligned} \quad (4.32)$$

The MHLE of σ^2 is obtained through the iterative procedure using (4.26) and the shared parameter (γ) are obtained using (4.27) and (4.30) via Newton Raphson approximation.

Chapter 5. Simulation

5.1 Simulation – Binary HGLM

5.1.1 Data Generation

We choose the mixed logit model as an illustration because binary data are particularly problematic to estimate through GLMM when the information at the domain level from SAE is insufficient with small sample size. The proposed h -likelihood approach is evaluated through Monte Carlo simulation performing 1000 simulations to estimate fixed effects and random effects. The first simulation study was conducted to evaluate the proposed method based on a single outcome variable. We considered GLMM as our benchmark method to compare the results of the proposed method varying small areas ($q = 5, 10, 20, 30$), each small area with the sample sizes of $n_q = 10, 30, 50, 100, 500$, respectively. A total of 20 data sets, the smallest data set with 50 observations, and the largest data set with 15000 observations, were analyzed based on two discrete variables with one binary outcome variable. First, consider the binary HGLM model with $\mathbf{y}|\mathbf{u} \sim \text{Bino}(\mathbf{p})$, $\mathbf{u} \sim N(0, \sigma^2)$ with log-likelihood of $\mathbf{y}|\mathbf{u}$

$$\ell_{\mathbf{y}|\mathbf{u}} = \mathbf{y}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})),$$

$$\boldsymbol{\theta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u},$$

$$b(\boldsymbol{\theta}) = \log(1 + \exp \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}),$$

$$\phi = 1, \mathbf{u} \sim N(0, \sigma^2) \text{ with } \text{Var}(\mathbf{u}) = \sigma^2.$$

The initial random effects for each small area are simulated from a normal distribution with mean 0, and initial variance is 0.1 ($\sigma_0^2 = 0.1$). The initial values of fixed effects are assumed to be $\beta_{11} = 1.3$, $\beta_{21} = 1.5$, for two discrete variables X_1 , and X_2 , and the intercept term, $\beta_0 = -1.5$. First, the mixed logit model is considered based on fixed effects and random effects data to calculate P_0 for the j^{th} individual in i^{th} small area using

$$P_0(y_{ij}|u_i = 1) = \frac{\exp(\beta_0 + \sum_{p=1}^2 X_{ijp}\beta_p + Z_{ij}u_i)}{1 + \exp(\beta_0 + \sum_{p=1}^2 X_{ijp}\beta_p + Z_{ij}u_i)},$$

which can be expressed in matrix form

$$P_0(\mathbf{y} = \mathbf{1}) = (\mathbf{1} + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1},$$

$$\text{where } \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_{11} \\ \beta_{12} \\ \beta_{21} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} -1.5 \\ 1.3 \\ 0.0 \\ 1.5 \\ 0.0 \end{pmatrix}, \mathbf{Z} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}_{N \times m}, \mathbf{u} = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \mathbf{u} \text{ is}$$

drawn from $N(0, 0.1)$, and the exponentiation is applied elementwise to the vectors. Under this scenario, the binary response variable is simulated using the calculated probability. Now we apply the proposed h -likelihood iterative method to compute maximum likelihood estimates of $\boldsymbol{\beta}$, \mathbf{u} , and σ^2 for the first combination of 20 combinations of (q, n_q) mentioned above. 1000 simulated data sets for each combination of q and n_q were generated, which means 16000 data sets were used to evaluate the proposed method.

5.1.2 MC Simulation Results

As described in section 5.1.1, the proposed h -likelihood approach and GLMM method were applied to 20 combinations of data sets for each combination were used to 1000 different data sets for the same q and n_q . The final estimates for each scenario of (q, n_q) were obtained, averaging over 1000 simulations for MHLEs and MLEs from the GLMM.

The final MHLEs of $\boldsymbol{\beta}$, and θ are obtained by taking the average over the number of simulations (1000 simulations). The MHLE of \mathbf{u} is obtained averaging over 1000 simulations and averaging over sample sizes of each scenario of (q, n_q) .

$$\hat{\boldsymbol{\beta}}_{mean} = \frac{1}{s} \sum_{i=1}^s \hat{\boldsymbol{\beta}}, \quad (5.1)$$

$$\hat{\mathbf{u}}_{mean} = \frac{1}{n_q} \sum_{j=1}^{n_q} \left(\frac{1}{s} \sum_{i=1}^s \hat{\mathbf{u}}_{ij} \right), \quad (5.2)$$

where s is the total number of simulations ($s = 1000$) performed at each combination of q and n_q .

The performance of MHLEs $\hat{\boldsymbol{\psi}} = (\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}, \hat{\boldsymbol{\theta}})$ are evaluated using the mean squared error (MSE) and the relative bias. For (q, n_q) ,

$$Bias\left(\hat{\boldsymbol{\psi}}_{(q, n_q)}\right) = \frac{E\left(\hat{\boldsymbol{\psi}}_{(q, n_q)}\right) - \boldsymbol{\psi}_{(q, n_q)}}{\boldsymbol{\psi}_{(q, n_q)}}. \quad (5.3)$$

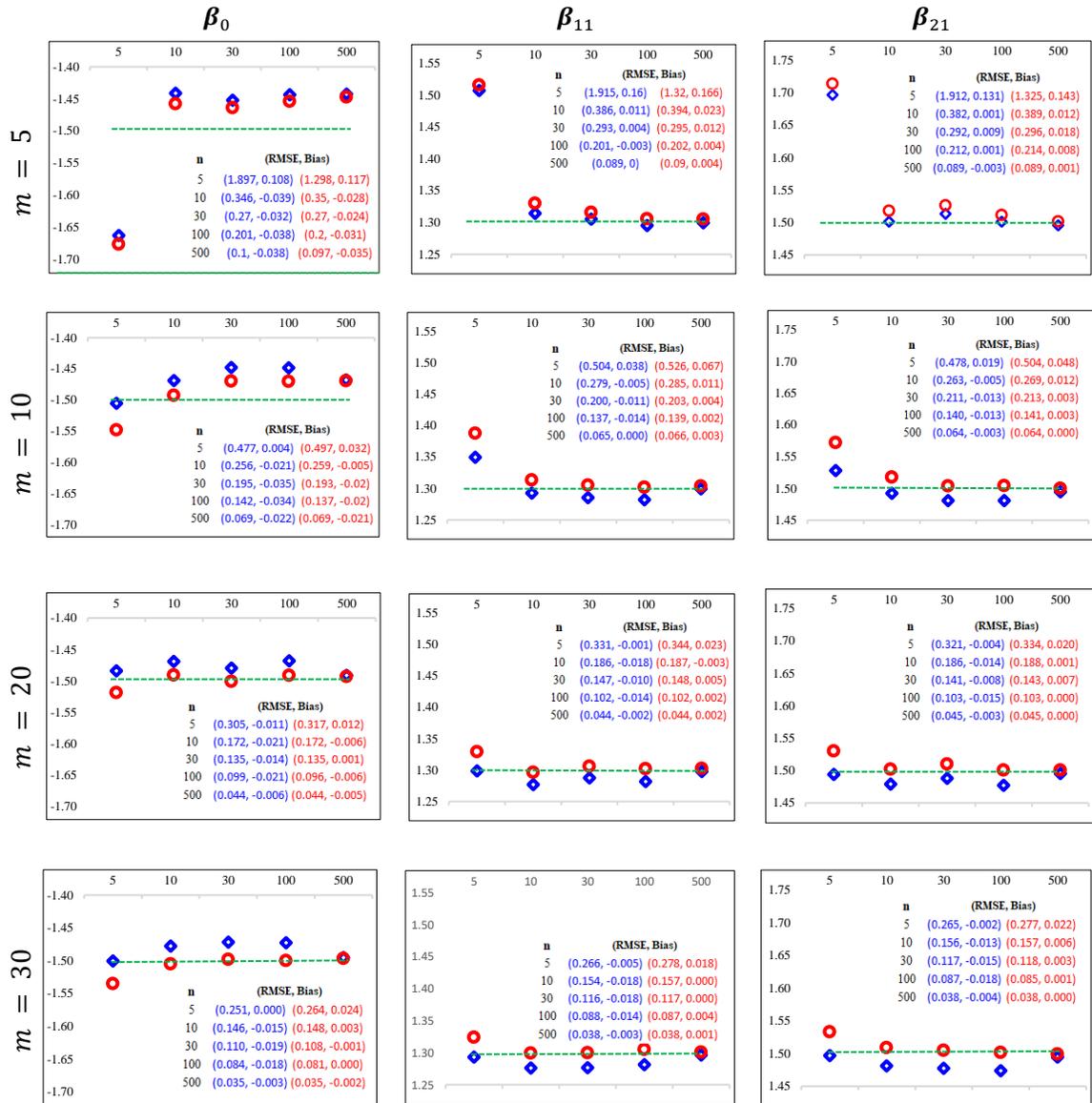
The MSE of $\hat{\boldsymbol{\beta}}$, and $\hat{\boldsymbol{\theta}}$ are obtained averaging over 1000 simulations for each scenario and the MSE for each small area ($MSE(\hat{\mathbf{u}})$) is obtained averaging over 1000 simulations and the number of small areas (q). The MHLEs and RMSEs (\sqrt{MSE}) for each estimate are given in Tables B.1 and B.2.

$$MSE\left(\hat{\boldsymbol{\beta}}_{q, n_q, k}\right) = \frac{1}{1000} \sum_{s=1}^{1000} (\hat{\beta}_{sk} - \beta_k)^2, \quad k = 1, \dots, (p + 1), \quad (5.4)$$

$$MSE\left(\hat{\mathbf{u}}_{q, n_q}\right) = \frac{1}{1000} \sum_{s=1}^{1000} \frac{1}{q} \left(\sum_{i=1}^q (\hat{u}_{si} - u_i)^2 \right). \quad (5.5)$$

Figure 5.1 shows the fixed effects estimates, root mean squared error (RMSE), and the relative bias for both CHBC and GLMM methods for each combination of $(q = m, n_q = n)$. For easiness, we considered an equal number of sample sizes for each small area. The simulation results show that the fixed effect estimates from the proposed CHBC method provide equal or slightly better results compared to GLMM, except in very few cases GLMM outperforms. While the RMSEs of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_{11}$, and $\boldsymbol{\beta}_{21}$ from GLMM for the (5, 5) case is smaller than CHBC; the relative bias is a little lower in CHBC than GLMM. The RMSEs, and the relative bias are approximately similar or better in the proposed CHBC approach in most cases.

Figure 5.1: MHLEs of fixed effects from CHBC and GLMM for mixed logit model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).



The blue diamonds, red circles, and green dotted lines display the MHLEs of fixed effects, MLEs of fixed effects, and the actual values of fixed effects. The blue highlighted values from the CHBC method, and the red highlighted values are from the GLMM for the RMSE and the relative bias for sample size $n = 5, 10, 30, 100, \text{ and } 500$ for the number of small areas $m = 5, 10, 20, \text{ and } 30$, respectively. The true values of β_0, β_{11} , and β_{21} are $-1.5, 1.3, \text{ and } 1.5$ for each combination of m and n .

The HMLEs of θ_s ($\hat{\theta}$) are displayed in Table B.1, which are equally accurate in the CHBC compared to GLMM estimates in every scenario. Both approaches slightly underestimated the variance parameter, but it is reliably accurate. However, the MLE of the θ from GLMM is somewhat precise than that of the CHBC, especially for small sample sizes. Overall, both the proposed CHBC and GLMM models perform better when the sample size increases. Furthermore, it is promising that the RMSE decreases when the sample size increases.

5.2 Simulation – Poisson HGLM

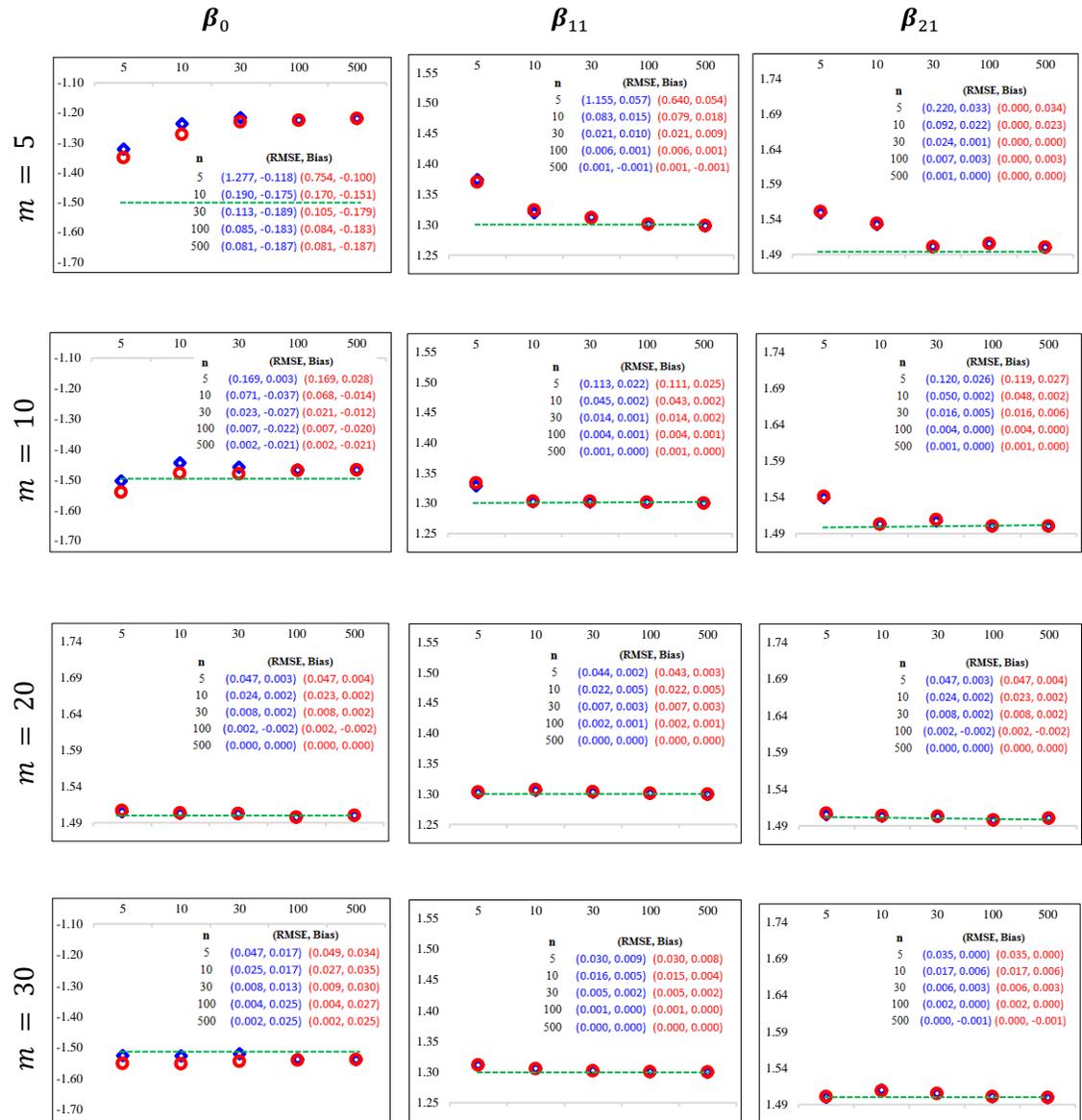
5.2.1 Data Generation

As described in section 5.1.1, the proposed CHBC method is evaluated using a second simulation study based on Poisson HGLM conducting 1000 simulations. The number of small areas $q = 5, 10, 20, 30$ and the sample size for each small area $n_q = 10, 30, 50, 100, 500$ were used to compare the results with the Poisson GLMM. Each data set with two discrete variables with a Poisson outcome variable $\mathbf{y}|\mathbf{u} \sim \text{Pois}(\boldsymbol{\lambda}), \mathbf{u} \sim N(0, \sigma^2)$ is considered. The initial fixed effect coefficients and σ^2 are similar to section 5.1.1, $\boldsymbol{\beta} = (\beta_0, \beta_{11}, \beta_{12}, \beta_{21}, \beta_{22})^T = (-1.5, 1.3, 0.0, 1.5, 0.0)^T$, and $\sigma^2 = 0.1$. The count data is generated by using $\boldsymbol{\lambda} = \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$. Also, we considered an equal number of sample sizes for each small area.

5.2.2 Simulation Results

The MHLEs of fixed effects and random effects are obtained using (5.1) and (5.2). The relative bias and RMSEs are calculated as described in section 5.1.2 using (5.3), (5.4), and (5.5). Figure 5.2 displays the MHLEs for fixed effects $\boldsymbol{\beta}_0, \boldsymbol{\beta}_{11}$, and $\boldsymbol{\beta}_{21}$ from CHBC and MLEs from GLMM for each combination of m, n . It shows that the MHLEs from both CHBC (blue diamond shapes) and GLMM (red circles) are very accurate in each scenario regardless of the sample size.

Figure 5.2: MHLEs of fixed effects from CHBC and GLMM for Poisson mixed model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).

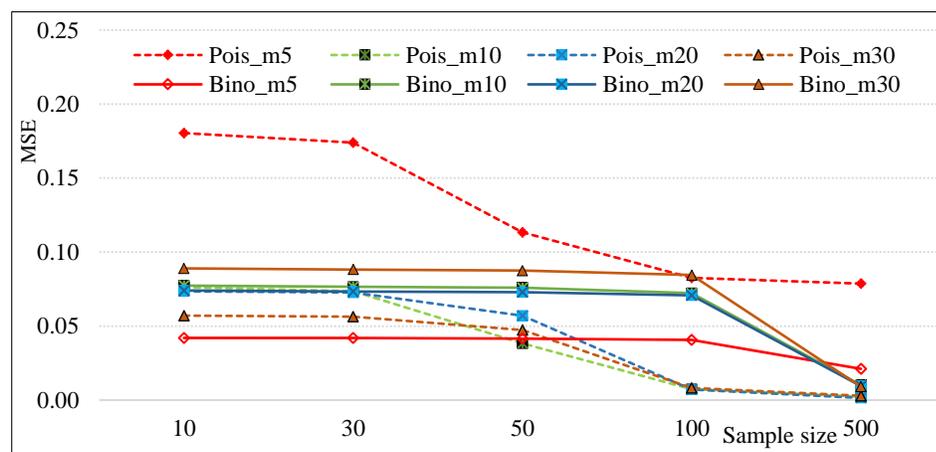


The blue diamonds, red circles, and green dotted lines display the MHLEs of fixed effects, MLEs of fixed effects, and the actual values of fixed effects. The blue highlighted values from the CHBC method, and the red highlighted values are from the GLMM for the RMSE and the relative bias for sample size $n = 5, 10, 30, 100, \text{ and } 500$ for the number of small areas $m = 5, 10, 20, \text{ and } 30$, respectively. The true values of $\beta_0, \beta_{11}, \text{ and } \beta_{21}$ are $-1.5, 1.3, \text{ and } 1.5$ for each combination of m and n .

The blue highlighted values within parenthesis are for RMSEs for both methods, the first value is for CHBC, and the second is for GLMM. Similarly, red highlighted values represent the relative bias, with the first one for CHBC, and the second is for GLMM. Both the RMSEs and the relative bias values are lower and very similar in both methods, indicating that the proposed CHBC method's performance is consistently better, like in GLMM. The results are significantly improved when the sample size and number of small areas increase. The simulation results are comparable with key findings in the literature, which is proved that estimations based on h -likelihood provide equal or better results compared to other modeling approaches (Lee & Nelder, 2005; Noh & Lee, 2007; Yun & Lee, 2004).

Figure 5.3 gives the MSE of the estimated average random effects for the mixed logit model and Poisson mixed model by varying the number of small areas (m) and each area's sample size (n) based on the proposed CHBC method. The dotted lines for the Poisson model and solid lines for the mixed logit model show considerably low MSE values for each scenario, while the improved MSE values when the number of small areas and the sample size increases.

Figure 5.3: Mean Squared Error (MSE) of the estimated random effect using CHBC for mixed logit and Poisson model.



The sample sizes 10, 30, 50, 100, 500 with 5, 10, 20, and 30 small areas were considered.

The simulation results are consistent with key findings in the literature, which is proven that estimations based on h -likelihood provide equal or better results compared to other modeling approaches (Lee & Nelder, 2005; Noh & Lee, 2007; Yun & Lee, 2004). Table B.2 shows the MHLE and MLE of the variance parameter obtained from both CHBC and GLMM for the Poisson mixed model. The RMSEs for each combination of m, n of the CHBC is lower compared to that of the GLMM, similarly the relative bias. Furthermore, the distribution of estimated random effects displayed in Figure B.1 shows that the distribution of $\hat{\mathbf{u}}$ from CHBC is very close to the actual distribution of \mathbf{u} . Based on the simulation results for both the mixed logit model and the Poisson mixed model, it indicates that the proposed CHBC method provides reliably better results.

5.3 Simulation – Joint Model Through Multivariate Random Effects

5.3.1 Data Generation

The empirical performance of the joint model through multivariate random effects was illustrated by conducting 1000 simulations using count data for two outcome variables. The number of small areas $q = 10, 20, 30, 50$ and the sample size for each small area $n_q = 10, 30, 100, 300$ were used with two binary outcome variables for each variable of interest. The initial fixed effect parameters $\beta_0^1, \beta_{11}^1, \beta_{21}^1$, for y_1 , and $\beta_0^2, \beta_{11}^2, \beta_{21}^2$ for y_2 were taken as $(-2.5, 1.3, 1.5)$, and $(-1.5, 1.3, 1.5)$. The initial values of the mean and the variance-covariance matrix for bivariate random effect \mathbf{u} was considered as

$$\boldsymbol{\mu}_0 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \boldsymbol{\Sigma}_0 = \begin{pmatrix} 1.3 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}.$$

The random effect is generated from the bivariate normal distribution given $\mathbf{u} \sim \text{Bivariate}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. For simplicity, we considered equal sample sizes ($N_1 = N_2$) for each outcome variable for each scenario. The total sample size was $N = 2N_1 = 2N_2$. The estimated

fixed effects, random effects, variance-covariance matrix, the mean squared error, and the bias are obtained as described in section 5.1.2 based on equations from (5.1) to (5.5).

5.3.2 Simulation Results

The maximum hierarchical likelihood estimates, RMSE, bias, and the 95% confidence interval for fixed effects of both variables of interest \mathbf{y}_1 , and \mathbf{y}_2 are given in Figures 5.4 and 5.5.

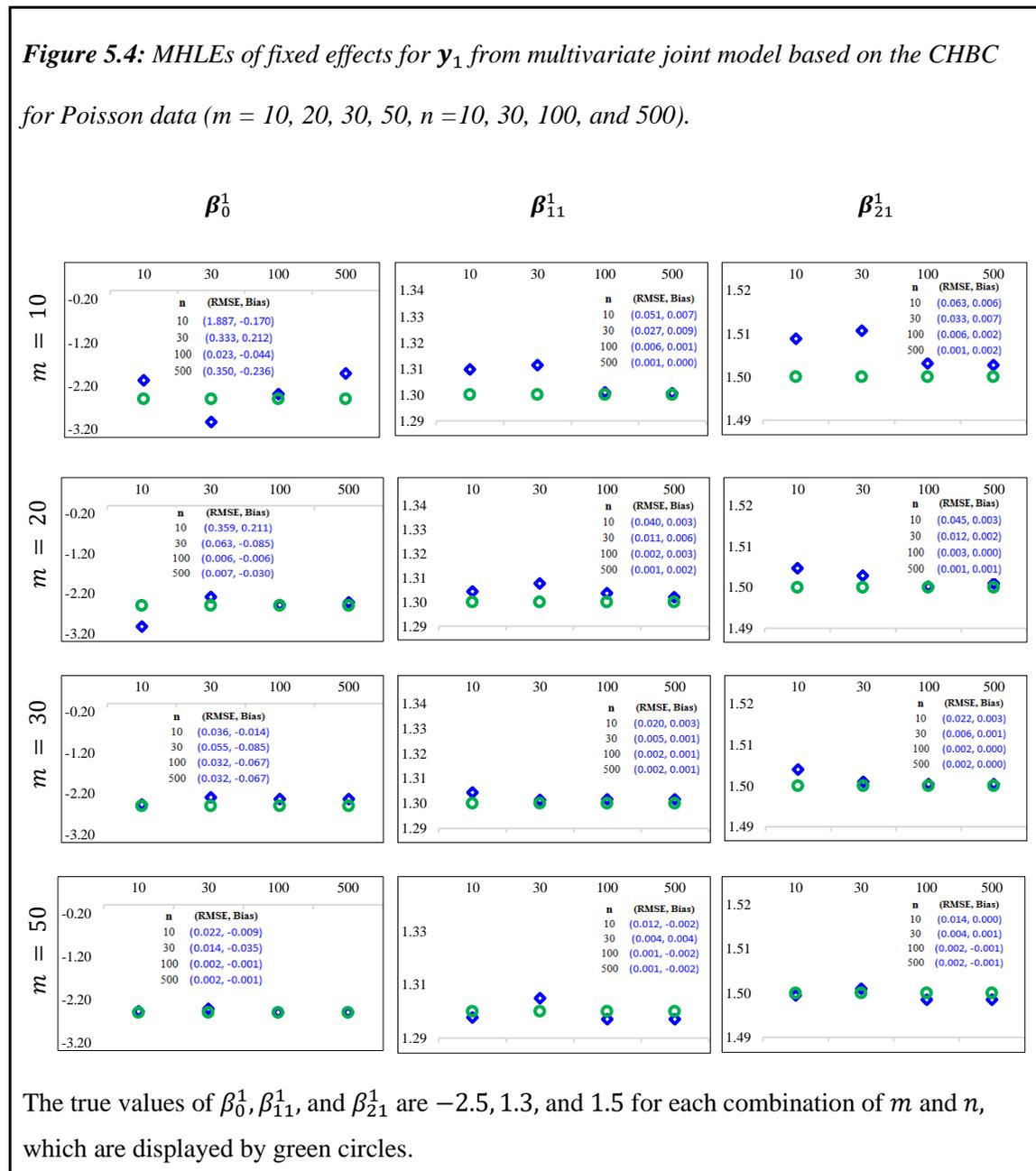
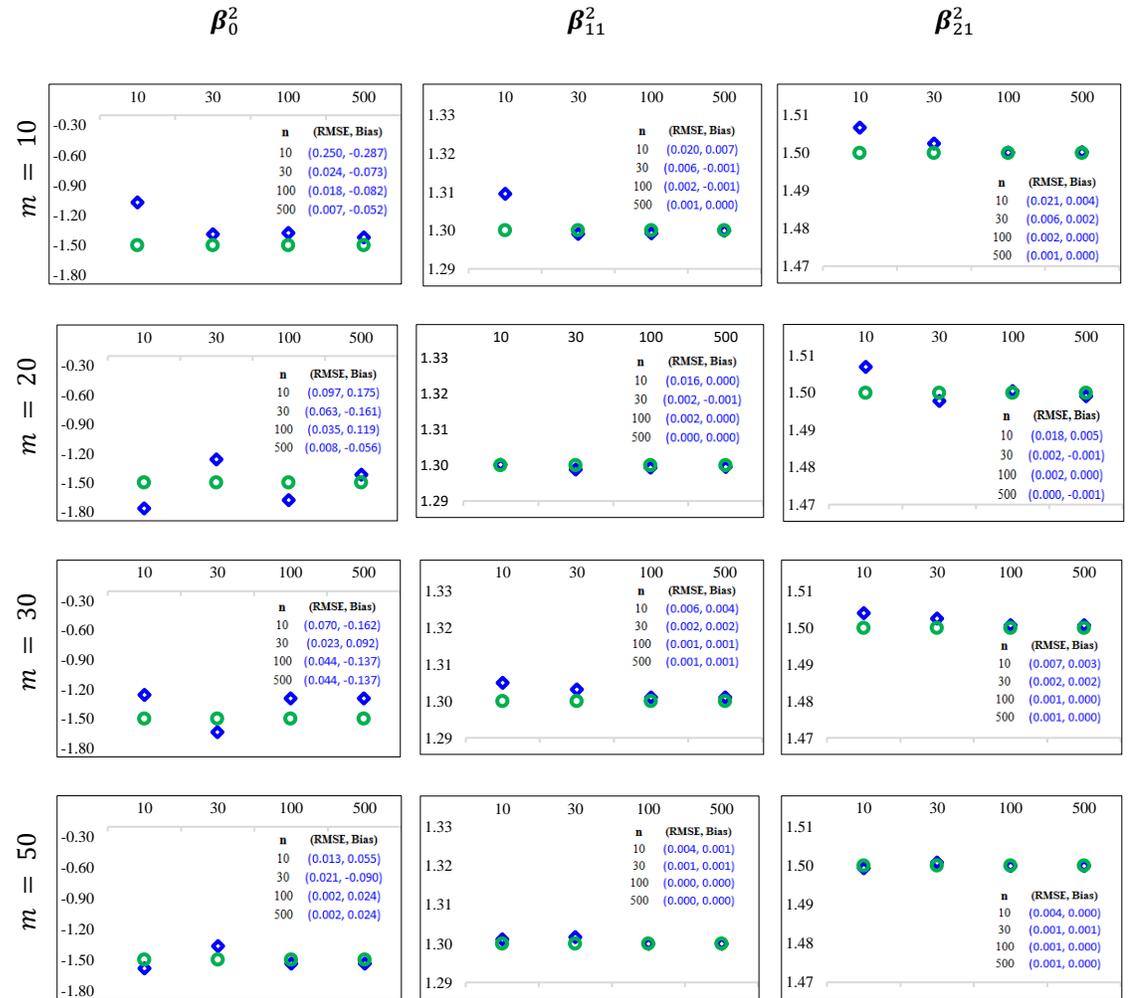


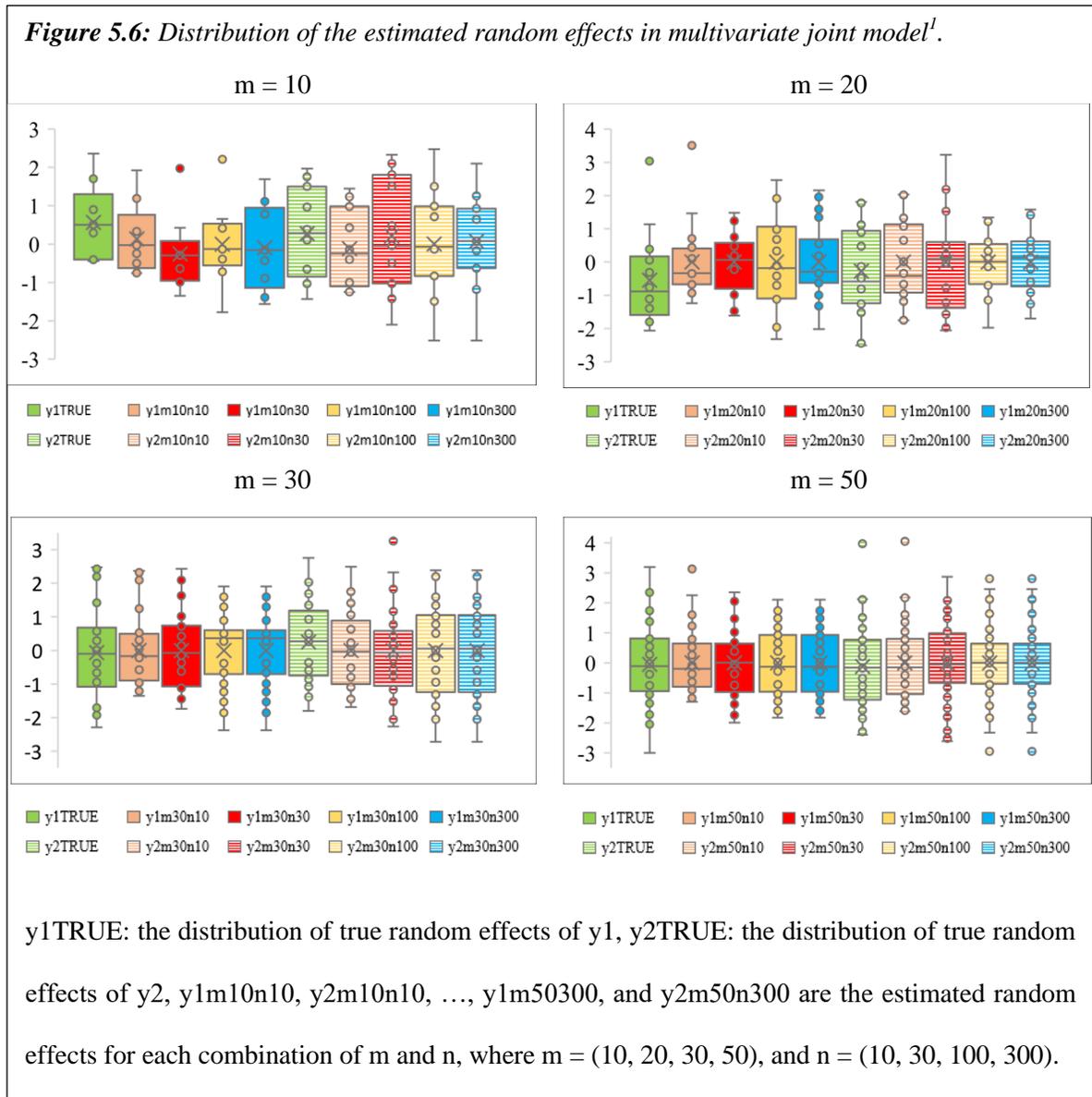
Figure 5.5: MHLEs of fixed effects for y_2 from multivariate joint model based on the CHBC for Poisson data ($m = 10, 20, 30, 50, n = 10, 30, 100, \text{ and } 500$).



The true values of $\beta_0^2, \beta_{11}^2, \text{ and } \beta_{21}^2$ are $-1.5, 1.3, \text{ and } 1.5$ for each combination of m and n , which are displayed by green circles.

The HMLEs are improved when the sample size increases, while it is a little off from the actual values for the intercept terms β_0^1 , and β_0^2 for small sample sizes. However, the HMLEs are consistently accurate throughout the univariate and multivariate analysis, indicating that the proposed CHBC performs well in both scenarios. The RMSE for β_0 for the first case, at (10,10) is 1.887, which is slightly higher compared to other scenarios. Apart from that, the RMSEs have

significantly lower values for most situations with decreasing values when the sample size increases. Additionally, the relative bias based on the estimated MHLEs and actual values of fixed effects are significantly lower in every combination of m , and n for \mathbf{y}_1 , and \mathbf{y}_2 . Based on Figure 5.6, the MHLEs for β_0^2 , β_{11}^2 , and β_{21}^2 are reliably better estimates except for the MHLEs of β_0^2 , and β_{11}^2 at (10,10) situation. Overall it shows that the MHLEs for each fixed effect are accurate from the joint model through multivariate random effects using the CHBC method.



The estimated variance parameters, absolute error, and the relative bias are displayed in Table C.1. It shows that HMLEs for variance parameters are better, and it is improved with the sample size, similarly the relative bias. Figure 5.6 displays the distribution of the estimated random effects varying with the sample size and number of small areas for both y_1 , and y_2 . It shows that the distribution of estimated random effects stays closer to the actual distribution of random effects for every combination of m and n . Overall, the simulation studies imply that the proposed CHBC method performs well in both the univariate and joint modeling scenarios in SAE.

Chapter 6. Real Data Analysis

6.1 CHBC Approach to Tobacco Smoking Data

In this section, we illustrate the proposed CHBC approach using a real data set of tobacco smoking combining 2015 and 2017 data from the Behavioral Risk Factor Surveillance System (BRFSS) (*Behavioral Risk Factor Surveillance System Survey Data (BRFSS)*, 2015). The BRFSS is a cross-sectional telephone survey that the state health department conducts monthly over landline and cellular telephones to collect prevalence data among US adults. We considered four outcome variables of interest: ever-use of E-cigarettes (EE), current-use of E-cigarettes (CE), ever-smoke (ES), and current-smoke (CS) with sample sizes of 29404, 28162, 25711, and 29396, respectively from 94 US counties. We applied the proposed CHBC method on the BRFSS data set to incorporate individual-level and area-level tobacco use behaviors and electronic cigarettes (E-Cigarettes) usage prevalence among youth at the county-level.

The auxiliary information is considered for age (≤ 12 , 13, 14, 15, 16, 17, and ≥ 18 years), race (4 groups: white, African American, Hispanic, and Others), sex (2 groups: male and female), year (2 groups: 2015, 2017), and the poverty values, which were extracted from the US census between 2015 and 2017. The race “others” group includes American Indian/Alaska Native, Asian, Native Hawaiian/other Pacific Islander, and multiple races (non-Hispanic). Table D.1 presents the summary statistics for each response variable. Youth age is stratified into different groups in some studies such as ≤ 12 , 12-17, ≥ 18 , ≤ 14 , 15-17, ≤ 18 , and > 18 years, etc. (Duke et al., 2014; *E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General*, 2016; Glasser, Abudayyeh, Cantrell, & Niaura, 2019; *Reducing Vaping Among Youth and Young Adults*, 2020). For this analysis, we grouped age into three groups (≤ 14 , 15 – 17, ≥ 18). We applied the CHBC model and GLMM to all four variables of interest ($\mathbf{y}_{EE}, \mathbf{y}_{CE}, \mathbf{y}_{ES}, \mathbf{y}_{CS}$), separately. The MHLEs of fixed effects, $\hat{\boldsymbol{\beta}}_{EE}, \hat{\boldsymbol{\beta}}_{CE}, \hat{\boldsymbol{\beta}}_{ES}$, and $\hat{\boldsymbol{\beta}}_{CS}$ are displayed in Table 6.1.

Table 6.1: Model estimates for current-use and ever-use of E-Cigarettes based on the CHBC and GLMM.

Current-use									
		CHBC				GLMM			
		$\hat{\beta}$	SE	Z Value	P(> Z)	$\hat{\beta}$	SE	Z Value	P(> Z)
Intercept		-1.625	0.132	-12.290	<0.001	-1.880	0.182	-10.332	<0.001
	<= 14 yrs	0.000				0.000			
Age	15-17 yrs	0.305	0.053	5.751	<0.001	0.418	0.083	5.012	<0.001
	>= 18 yrs	0.688	0.065	10.633	<0.001	0.844	0.096	8.787	<0.001
Race	White	0.212	0.056	3.788	<0.001	0.244	0.084	2.916	0.004
	African American	-0.243	0.073	-3.345	0.001	-0.182	0.105	-1.726	0.084
	Hispanic	0.142	0.059	2.409	0.016	0.188	0.092	2.049	0.040
	Others	0.000				0.000			
Gender	Male	0.000				0.000			
	Female	0.291	0.031	9.270	<0.001	0.245	0.045	5.387	<0.001
Year	2015	0.000				0.000			
	2017	-0.946	0.058	-16.255	<0.001	-0.785	0.077	-10.220	<0.001
Poverty Rate (%)		-0.564	0.683	-0.825	0.409	-0.079	0.902	-0.087	0.931
Ever-use									
		CHBC				GLMM			
		$\hat{\beta}$	SE	Z Value	P(> Z)	$\hat{\beta}$	SE	Z Value	P(> Z)
Intercept		-0.905	0.101	-8.988	<0.001	-0.990	0.144	-6.880	<0.001
	<= 14 yrs	0.000				0.000			
Age	15-17 yrs	0.447	0.039	11.413	<0.001	0.407	0.060	6.736	<0.001
	>= 18 yrs	0.767	0.050	15.307	<0.001	0.771	0.062	12.421	<0.001
Race	White	0.095	0.042	2.245	0.025	0.065	0.062	1.038	0.299
	African American	-0.073	0.051	-1.418	0.156	-0.018	0.077	-0.232	0.817
	Hispanic	0.311	0.044	7.040	<0.001	0.349	0.068	5.098	<0.001
	Others	0.000				0.000			
Gender	Male	0.000				0.000			
	Female	0.137	0.024	5.644	<0.001	0.127	0.035	3.612	<0.001
Year	2015	0.000				0.000			
	2017	-0.172	0.043	-4.010	<0.001	-0.106	0.061	-1.751	0.080
Poverty Rate (%)		0.575	0.522	1.101	0.271	0.653	0.739	0.883	0.377

Table 6.2: Model estimates for current-smoke and ever-smoke based on the CHBC and GLMM.

Current-smoke									
		CHBC				GLMM			
		$\hat{\beta}$	SE	Z Value	P(> Z)	$\hat{\beta}$	SE	Z Value	P(> Z)
Intercept		-3.337	0.171	-19.500	<0.001	-3.412	0.240	-14.206	<0.001
	<= 14 yrs	0.000				0.000			
Age	15-17 yrs	0.429	0.075	5.732	<0.001	0.383	0.120	3.189	0.001
	>= 18 yrs	1.099	0.086	12.776	<0.001	0.908	0.119	7.605	<0.001
	White	0.290	0.074	3.907	<0.001	0.187	0.107	1.752	0.080
Race	African American	-0.615	0.102	-6.028	<0.001	-0.591	0.146	-4.062	<0.001
	Hispanic	0.087	0.080	1.087	0.277	-0.015	0.121	-0.123	0.902
	Others	0.000				0.000			
	Male	0.000				0.000			
Gender	Female	0.235	0.041	5.721	<0.001	0.174	0.058	2.995	0.003
	2015	0.000				0.000			
Year	2017	-0.261	0.073	-3.561	<0.001	-0.198	0.099	-1.997	0.046
Poverty Rate (%)		2.655	0.843	3.151	0.002	3.297	1.151	2.864	0.004
Ever-smoke									
		CHBC				GLMM			
		$\hat{\beta}$	SE	Z Value	P(> Z)	$\hat{\beta}$	SE	Z Value	P(> Z)
Intercept		-1.939	0.138	-14.005	<0.001	-1.868	0.188	-9.945	<0.001
	<= 14 yrs	0.000				0.000			
Age	15-17 yrs	0.477	0.052	9.190	<0.001	0.337	0.071	4.732	<0.001
	>= 18 yrs	1.002	0.060	16.576	<0.001	0.857	0.072	11.920	<0.001
	White	0.177	0.050	3.573	<0.001	0.008	0.069	0.120	0.905
Race	African American	-0.302	0.062	-4.835	<0.001	-0.278	0.086	-3.242	0.001
	Hispanic	0.217	0.051	4.246	<0.001	0.165	0.075	2.197	0.028
	Others	0.000				0.000			
	Male	0.000				0.000			
Gender	Female	0.136	0.028	4.813	<0.001	0.100	0.038	2.606	0.009
	2015	0.000				0.000			
Year	2017	-0.317	0.053	-6.022	<0.001	-0.235	0.075	-3.129	0.002
Poverty Rate (%)		3.688	0.737	5.003	<0.001	3.811	1.004	3.794	<0.001

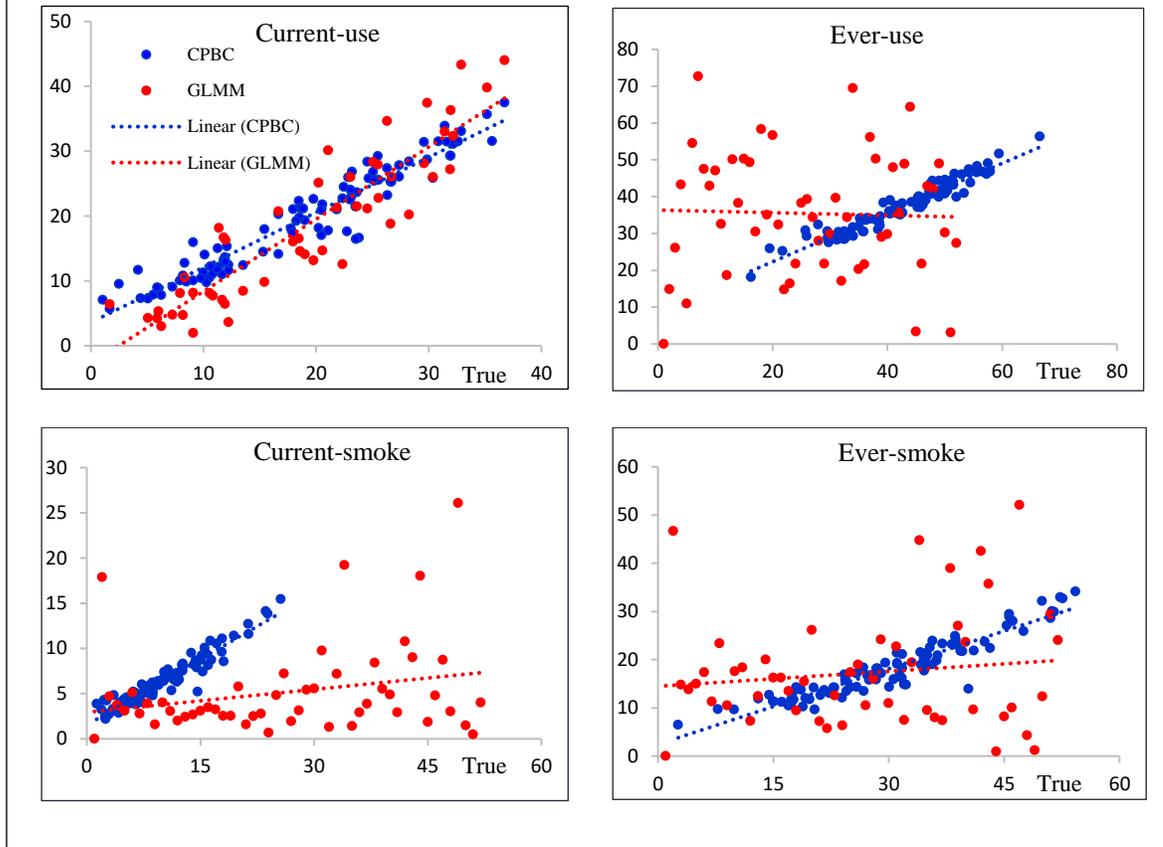
Table 6.1 displays the fixed effects estimates for age, race, gender, year, and the poverty rate for current-use and ever-use of E-Cigarettes. The coefficient effects are slightly higher for all the variables obtained from GLMM compared to CHBC except for the coefficient estimates of females for both current-use and ever-use. The most significant impact on the odds ratio of current-use and ever-use of E-Cigarettes resulted from the age group 18 or above after adjusting for other

variables in the model. The lowest effect on the odds ratio is in the year 2017, indicating that the odds ratio of current-use of E-Cigarettes has decreased by a factor of $e^{-0.946}(e^{-0.785})$ in 2017 compared to 2015, and that value of ever-use has decreased by a factor of $e^{-0.172}(e^{-0.106})$ after adjusting for age, gender, race, and poverty rate based on the CHBC (GLMM), respectively.

Table 6.2 shows that the poverty rate has a significant impact on the prevalence of both current-smoke and ever-smoke based on the estimates from CHBC and GLMM. Among the age groups ≤ 14 , 15-17, and ≥ 18 , the prevalence is higher among the individuals 18 or above compared to ≤ 14 years old. Overall, the smoking prevalence estimated from both CHBC and GLMM is very similar and consistent with the recent research studies (Hongying Dai et al., 2018; Hongying Dai & Hao, 2016).

Figure 6.1 displays the observed and the estimated prevalence obtained from CHBC and GLMM for current-use, ever-use of E-Cigarettes, current-smoke, and ever-smoke, respectively. While the observed vs. estimated prevalence is very accurate in both methods for the current-use of E-Cigarettes, it is significantly better in other outcomes from the CHBC than GLMM. The estimates for ever-use of E-Cigarettes from GLMM is not precise compared to prevalence estimates for current-smoke and ever-smoke. The results indicate that the proposed CHBC performs well in any scenario.

Figure 6.1: Observed and estimated prevalence ever-use and current-use of E-Cigarettes, current-smoke and ever-smoke in the United States based on 2015-2017 YRBSS data.



Next, we estimated the random effects for missing counties using the nearest neighboring approach, assuming that the random effects of adjacent areas are correlated, with correlation decaying to zero as distance increases. Using this fact, the estimated random effect for missing county $c_i(\tilde{\mu}_{c_i})$ can be obtained as

$$\tilde{\mu}_{c_i} = \hat{\mu}_{c_j},$$

$$s. t. \min \text{dist}(c_i, c_k), k = 1, \dots, m - 1, \quad (6.1)$$

where c_j is the closest to county c_i . Then, using the MHLEs and the estimated random effects for missing counties, we calculated the unit-level prevalence for each combination ($3 \times 2 \times 2 \times 4$) of groups using

$$\tilde{P}_{ijkc}(y_{ijkc} = 1|u_c) = \frac{\exp(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + x'_c \hat{\eta} + \hat{u}_c)}{1 + \exp(\hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + x'_c \hat{\eta} + \hat{u}_c)}, \quad (6.2)$$

where $\hat{\alpha}_i (i = 1,2,3)$, $\hat{\beta}_j (1,2)$, $\hat{\gamma}_k (k = 1,2,3,4)$, and $\hat{\eta}$ are the coefficient estimates for age, gender, race, and poverty rate, respectively. The unit-level estimations were used to obtain the county-level estimates using the U.S. Census population as in equation (6.3).

$$\tilde{P}(y_c = 1|u) = \frac{\sum_i \sum_j \sum_k \tilde{P}_{ijkc} \times \text{Pop}_{ijkc}}{\text{Pop}_c}, \quad (6.3)$$

where $\text{Pop}_c = \sum_i \sum_j \sum_k \text{Pop}_{ijkc}$ is the total population for county c . We compared the model predicted prevalence and observed prevalence using Pearson's and Spearman's correlation coefficients. The predicted proportions for "ever use of E-cigarettes", "current use of E-cigarettes", "ever-smoke", and "current-smoke" are obtained using the proposed method through MHLEs. Figure 6.3 shows the estimated county prevalence for each variable of interest.

Table D.2 gives the estimated state-level prevalence for current-use and ever-use of E-Cigarettes based on CHBC and GLMM. Based on both CHBC and GLMM estimates, North Carolina (29.22%, 34.55%), Kentucky (27.38%, 31.83%), New Mexico (27.00%, 24.10%), West Virginia (26.09%, 30.64%), Arkansas (25.89%, 29.39%), Delaware (24.04%, 23.60%), Vermont (24.04%, 22.94%), South Carolina (23.82%, 26.46%), and Kansas (23.35%, 24.71%) states have the highest prevalence of the current-use of E-Cigarettes in the US. While Wisconsin, with a prevalence of 24.99%, falls in the top ten US states based on the CHBC model, and Nevada, with 26.16%, was in the top ten based on the GLMM model. Both models identified the same eight US states out of the lowest ten rankings of the current-use of E-Cigarettes prevalence. At the same

time, Virginia and the District of Columbia were the rest two smallest prevalence from the CHBC and South Dakota and Maryland from the GLMM.

Overall, the prevalence of ever-use of E-Cigarettes is higher compared to three other prevalences. The CHBC and GLMM identify the top ten US states based on the prevalence ranking for ever-use as New Mexico (56.16%, 75.28%), Oklahoma (44.88%, 55.94%), West Virginia (44.37%, 57.36%), North Carolina (44.30%, 51.07%), Kentucky (42.58%, 49.11%), Nevada (42.06%, 49.20%), Colorado (40.89%, 44.26%), Arkansas (40.74%, 44.59%), and Delaware (40.44%, 44.03%). Based on the CHBC method, Arizona state has a higher prevalence of 41.27% of the ever-use of E-Cigarettes, while Kansas is within the top ten US states based on GLMM with a 44.47% value. The eight states identified as the lowest rankings from both methods are the same, except Georgia and Iowa were the lowest based on the CHBC, and the District of Columbia and Oregon were the lowest from the GLMM. Overall, the states with the highest prevalence for current-use also have the highest prevalence for ever-use of E-Cigarettes.

Table D.3 shows the top ten US states with the highest and the lowest prevalence for current-smoke and ever-smoke based on the estimates from CHBC and GLMM. Based on the CHBC method, West Virginia (New Mexico) has the highest prevalence for current-smoke, with 11.89% (ever-smoke, with 30.71%). The GLMM results in Iowa (New Mexico) with the highest prevalence for current-smoke, with 15.50% (ever-smoke, with 44.37%). The lowest rankings were among the same US states based on the estimates from both methods. Overall, it shows that the GLMM slightly overestimate at the higher end and underestimate at the lower end of current-smoke and ever-smoke, similar results for current-use and ever-use of E-Cigarettes based on Table D.2.

The estimated county-level prevalence for all four outcomes from CHBC and GLMM methods are displayed in the columns (a) and (b) in Figure D.1. The results imply that the current use of E-Cigarettes among youth is higher than current smoking in most US counties. Both Spearman's and Pearson's correlations of observed and estimated prevalence for each outcome ranged from 0.93 to 0.96, indicating the accuracy of the model estimations from the CHBC method.

The lower correlation coefficients for GLMM shows that the estimated prevalence from CHBC is slightly better compared to GLMM. Furthermore, the county-level distribution of the estimated prevalence is somewhat different in both methods for the current-use of E-Cigarettes. However, the estimations from both models are reliably accurate.

6.2 CHBC Approach to COVID-19 Data

We illustrate the proposed CHBC approach using the publicly available novel coronavirus data at the county-level in the US, which was downloaded from the data repository created by the John Hopkins University Center for Systems Science and Engineering (JHU CSSE) (Dong, Du, & Gardner, 2020). The novel coronavirus is also known as severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), or COVID-19. The novel coronavirus outbreak was originated in Wuhan, the capital city of Hubei province, and spread rapidly through China and globally, covering more than 150 countries worldwide. Since then, in public health-related research, a fair amount of research studies were conducted to study and to mitigate the spread of the virus by considering various types of risk factors (Dahab et al., 2020; Dong et al., 2020; P. Walker et al., 2020).

This analysis is based on US COVID-19 data as of July 31, 2020, and the confirmed COVID-19 cases were modeled using the Poisson HGLM model using the proposed CHBC. The COVID-19 data set includes data for confirmed cases and deaths of 3141 US counties. New York County in New York has the highest number of cases (225,148) and mortalities (23,531), respectively. There were 12 (0.4%) out of 3141 counties with zero cases and 25 counties with one reported case as of July 31, 2020.

The county-level COVID-19 data were available for most counties except for Utah, which was available by region, Southwest, Southeast, TriCounty, and Central, as displayed in Table 6.3. We calculated county-level COVID-19 cases and deaths by adjusting the region total by county population, as shown below.

$$\text{Count}_i = \frac{N_{\text{region}}}{\text{Pop}_{\text{region}}} * \text{Pop}_{\text{county}},$$

where Count_i is the number of COVID-19 cases (deaths), N_{region} is the regional case (death) count, $\text{Pop}_{\text{region}}$ is the regional population, and $\text{Pop}_{\text{county}}$ is the county population.

Table 6.3: COVID-19 data by regions in Utah State.

Region	Cases	Deaths	Population	FIPS
South west	2901	24	252042	49001, 49017, 49021, 49025, 49053
South east	83	0	40229	49007, 49015, 49019
Tri County	153	0	56622	49009, 49013, 49047
Central	369	2	252042	49023, 49027, 49039, 49041, 49057

It is challenging to analyze the county-level due to the lack of individual and county-level information yet. We considered some COVID-19 related variables together with the US census data at the county-level. The health professionals and medical researchers advise that the virus spreads rapidly by direct contacts or through respiratory droplets from coughing and sneezing of an infected person ("How COVID-19 Spreads,"). Hence, the Centers for Disease Control and Prevention (CDC) and the World Health Organization (WHO) strongly recommended that social distancing is the most effective approach to alleviate the spread of the virus ("Global research on coronavirus disease (COVID-19)," ; "How COVID-19 Spreads,"). The recent investigations show that the areas with an increasing trend of COVID-19 were some geographical regions where the people had close gatherings, more number of travels, and various types of social events, etc. (Banerjee & Nayak, 2020; P. Walker et al., 2020).

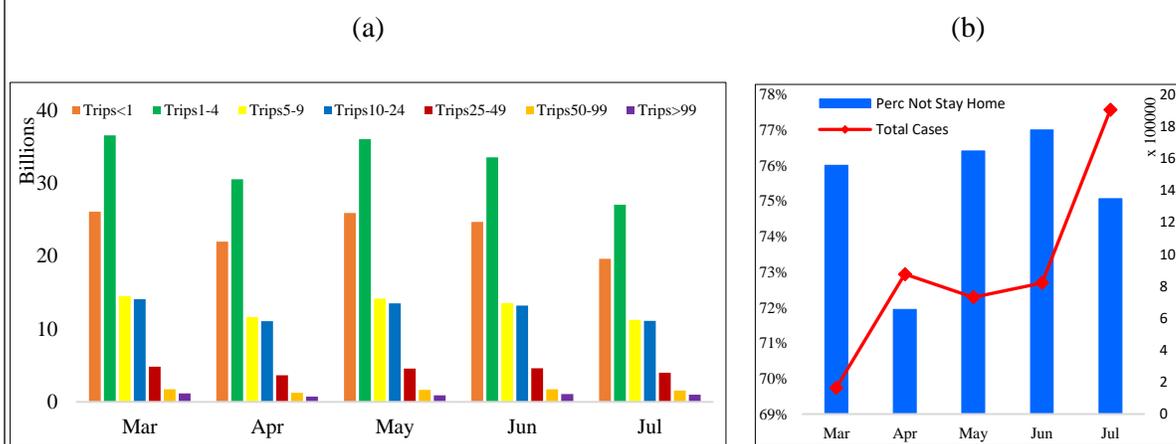
Hence, we extracted daily travel data from the Bureau of Transportation Statistics (BTS) in the United States Department of Transportation for March 01, 2020, to July 31, 2020, which explains the traveling behavior of people based on mobility data of the nation ("Daily Travel during the COVID-19 Public Health Emergency," 2020). We considered auxiliary information of

“population staying at home”, “not stay at home”, and “number of trips” (less than 1 mile, 1-4 miles, 5-9 miles, 10-24 miles, 25-49 miles, 50-99 miles, 100 miles or above). No data is available in BTS for the counties with fewer than 50 devices on any given day. The analysis includes the average number of trips for each category and the percent of the average population staying at home (Avg PSH), where

$$\text{Avg PSH} = \frac{\sum_{d=1}^{N_m} \text{PSH}_d}{\left(\sum_{d=1}^{N_m} \text{PSH}_d + \sum_{d=1}^{N_m} \text{PNSH}_d \right)},$$

N_m is the total number of days of the month (30 or 31 from March to July), PSH_d is the population staying at home on the day d , and PNSH_d is the population not staying at home on the day d .

Figure 6.2: (a) Total number of trips by month for trips less than 1 mile, 1-4 miles, 5-9 miles, 10-24 miles, 25-49 miles, 50-99 miles, and 100 miles or above and, (b) Population not stay home (%) vs. COVID-19 cases.



In column (b), the left side of y axis shows the percentage of population not stay at home in each month, and the right side of y axis represents the cumulative COVID-19 cases as on July 31, 2020.

Figure 6.2 (a) shows the average number of trips made each month for each trip mileage category. The total number of 1-4 mile trips and less than 1-mile trips are higher than the number

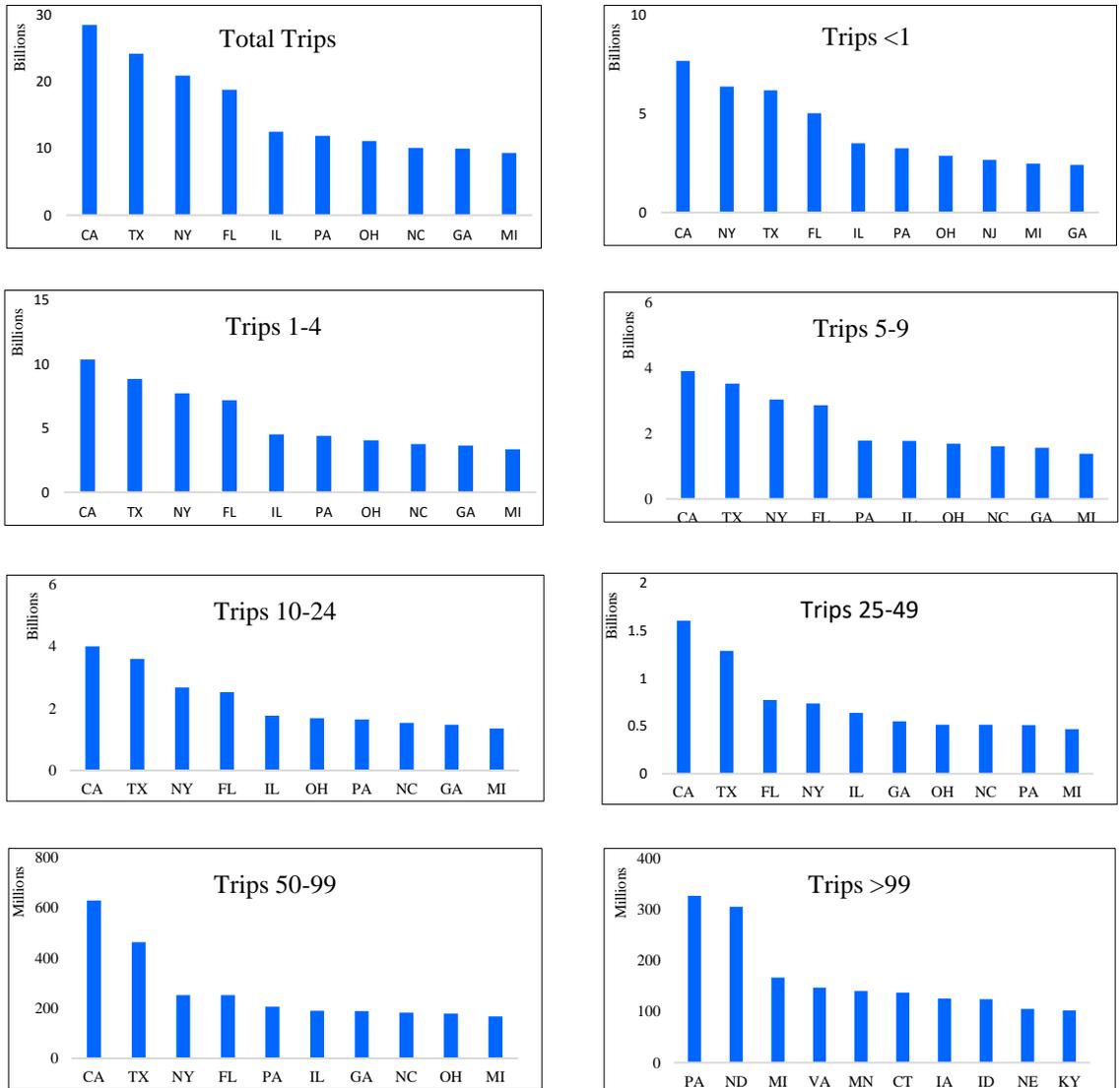
of trips of 5 miles or higher each month. Overall, there is no significant drop in the number of trips, indicating that this could potentially increase the chances of getting exposed to the virus. Furthermore, based on the tracked mobility data, Figure 6.2 (b) shows that the percentage of people not staying home was high in March, then it had decreased in April, again a significant increase in May and June. The number of total cases each month compared to the 2-3 weeks lagged number of trips provides a clear indication for the rise in cases due to less social distancing. The majority not staying home in May and June had affected to see a substantial increase in cases in July.

We also examined the distribution of the top ten US states that have the highest number of trips for each mileage group, which is displayed in Figure 6.3. As of July 31, 2020, California being the number one state with the highest number of trips for each mileage group except for the group of 100 miles or above, could be a potential risk factor for having the maximum number of COVID-19 cases (501,034) in the US. Similarly, Florida (483,280), New York (420,954), Texas (366,792), Georgia (188,828), North Carolina (185,373), New Jersey (181,012), Illinois (180,454), Pennsylvania (117,414), and Michigan (87,491) States are also among the top states with higher COVID-19 counts, which are also in the group of top states with a higher number of trips. These states have a large population as well as more diverse people. Hence, it is obvious to see many different mileage trips, and eventually, more COVID-19 cases.

Also, we considered the number of days for cases as of July 31, 2020, since the first COVID-19 cases were reported. Furthermore, the number of active cases was used as a piece of covariate information in the model. The individual level or county-level COVID-19 characteristic data are still not available for the most critical risk factors. Hence, we considered county-level demographic information from the US census, which includes age (≤ 17 years, 18-29 years, 30-44 years, 45-64 years), the proportion of individuals 65 years or above living alone. The age group 65 years or above was removed from the analysis due to a high correlation with 65 years or above

living alone. Additionally, gender (male, female), race (white, African American, Hispanic, and others), the poverty level, and the unemployment rate were considered auxiliary information.

Figure 6.3: Top ten states with highest number of trips made in each category during March 01, 2020 – July 31, 2020.



Each trip is measured in miles.

Table 6.4 displays the fixed effect estimates and the variance parameter of normally distributed random effect \mathbf{u} , from the CHBC method, and Table E.1 shows the MLEs based on the GLMM. Both tables indicate that the MHLEs and the MLEs are very similar for each covariate. The results indicate that there is a significant impact on COVID-19 cases by the percentage of people not staying at home and the number of different mileage trips. The association with COVID-19 cases is significantly higher for the people not staying at home ($\beta = 3.318, p < 0.001$) compared to staying at home after adjusting for other covariate information in the model ("Provisional COVID-19 Death Counts by Sex, Age, and State," 2020).

Besides the impact of not staying at home proportion, it also shows that the number of various mileage of trips is also a significant risk factor for seeing a higher number of cases. It is expected to see more cases from the population that made 100 miles or above mileage trips compared to less than 1-mile trips. From Figure 6.3, the top ten states with a higher number of 100 miles or above trips, Pennsylvania, North Dakota, Michigan, Virginia, Minnesota, Connecticut, Iowa, Nebraska, and Kentucky are still an increasing trend of cases. The majority of states have a large number of 100 miles or above trips in March, a slight decrease in April, and again a significant increase in June and July, which explains for most of the states to see the peak of cases in June and July. Similarly, the trips between 25-49 and 50-99 miles also have a significant positive association with an increase in cases compared to less than 1-mile trips with multiplicative factors of 9.03 and 35.05, respectively.

Additionally, a higher number of trips within 1 mile has a positive impact on COVID-19 increase than trips between 1-25 mileage, which might be mostly in very urban places with short distance to work, groceries, etc. However, they are exposed to more people, hence more likely to get infected with the virus due to population density. Besides, it also shows that a strong positive association with the poverty rate ($\beta = 1.580, p < 0.001$), while a negative association with the unemployment rate ($\beta = -3.614, p < 0.001$) with COVID-19 cases, which sounds counterintuitive. However, it is probable to have a negative association since the unemployed

people may be less likely to get exposed to the virus, also unlikely to start new employment during the pandemic as per the increase of unemployment rate from 3.8% in February to more than 15% in May. The majority of unemployed people are less educated people (in May, $\approx 18\%$ of less than high school graduates, and $\approx 7\%$ of Bachelor's degrees or higher graduates). However, it was challenging for them to be employed again due to temporary or permanent closures of small businesses, such as restaurants, shopping, retail, fitness centers, beauty, and spas, etc. Furthermore, young people have a higher unemployment rate, but fewer COVID-19 cases than middle or older ages ($\approx 25\%$ of 16-24 years, $\approx 11\%$ of 25-54 years) ("Effects of COVID-19 Pandemic on Employment and Unemployment Statistics," 2020; Kochhar, 2020). Still, COVID-19 was a high-risk factor for a substantial increase in the unemployment rate; however, the unemployment rate may not have a significant impact on the increase in COVID-19.

Also, we observed a negative correlation between the different age groups ($\beta_{25-44} = -1.017, p < 0.037, \beta_{45-64} = -3.418, p < 0.001$) and COVID-19 cases. Still, the coefficient effects are directionally correct, having a larger value for the population 65 years or older living alone. However, we could not explain the negative association of age groups with COVID-19 cases. The age group 65 years or above was removed from the analysis since it has a strong positive correlation with 65 years or above living alone (0.73) and a high negative correlation with 25-44 years (0.72). Comparatively, other variables are not highly correlated. The results also show that African Americans ($\beta = 1.633, p < 0.001$) and Hispanics ($\beta = 0.833, p < 0.001$) have a robust positive association with COVID-19 cases than Whites. However, other races (American Indian, Asian) ($\beta = -0.153, p = 0.036$) are negatively associated with cases. Furthermore, males ($\beta = 2.286, p < 0.001$) are more strongly related to COVID-19 cases than females.

Furthermore, the number of active cases ($\beta = 0.314, p < 0.001$) and the number of days since the first reported case ($\beta = 0.123, p = 0.005$) are positively associated with case counts.

There will be

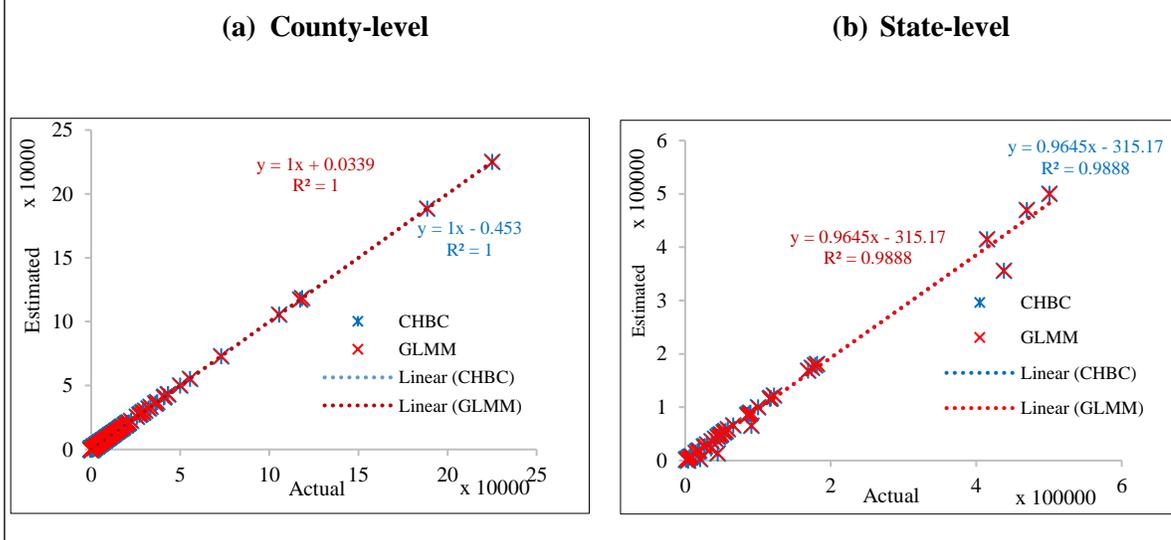
Table 6.4: MHLEs for fixed effects using the CHBC method for COVID-19 cases.

Characteristics	CHBC				95% CI	
	Estimate	SE	Z Value	P(> Z)	Lower	Upper
Intercept	-10.051	0.392	-25.617	<0.001	-10.820	-9.282
Number of days (log)	0.123	0.044	2.823	0.005	0.038	0.208
Number of active cases (log)	0.314	0.008	40.131	<0.001	0.299	0.329
Population not stay home (%)	3.818	0.271	14.095	<0.001	3.287	4.349
Number of trips (%)						
< 1 mile	0.000					
1 - 4 miles	-1.466	0.320	-4.577	<0.001	-2.094	-0.838
5 - 9 miles	-1.590	0.360	-4.420	<0.001	-2.295	-0.885
10 - 24 miles	-0.833	0.374	-2.231	0.026	-1.566	-0.101
25 - 49 miles	2.474	0.503	4.921	<0.001	1.489	3.459
50 - 99 miles	3.878	0.942	4.114	<0.001	2.030	5.725
>= 100 miles	4.996	1.735	2.880	0.004	1.596	8.397
65 Years or above living alone (%)	-0.013	0.434	-0.031	0.975	-0.864	0.837
Age (%)						
<= 24 years	0.000					
25_44 years	-1.017	0.488	-2.084	0.037	-1.973	-0.061
45_64 years	-3.418	0.481	-7.102	<0.001	-4.361	-2.475
Race (%)						
African American	1.633	0.090	18.113	<0.001	1.456	1.809
Hispanic	0.833	0.095	8.742	<0.001	0.646	1.019
Other races	-0.153	0.073	-2.100	0.036	-0.295	-0.010
White	0.000					
Gender (%)						
Male	2.286	0.521	4.384	<0.001	1.264	3.308
Female	0.000					
Poverty rate (%)	1.580	0.230	6.876	<0.001	1.130	2.031
Unemployment rate (%)	-3.614	0.773	-4.674	<0.001	-5.129	-2.098
		Std				
Variance Component	Estimate	Dev				
County-level	0.242	0.491				

The estimated counts are represented using the U.S county-level map, which is given in Figure E.1. We compared the estimations by grouping the estimated cases into nine and fifteen categories based on the cutoffs of the actual distribution. The nine groups of the confirmed COVID-19 counts are 0-18, 19-46, 47-95, 96-181, 182-318, 319-574, 575-1204, 1205-3925, and 3926-225148), which were applied to the estimated counts based on the proposed CHBC method and GLMM. The first column of the Figure E.1 displays the confirmed and estimated density of COVID-19 counts by county obtained for nine categories based on the CHBC, GLMM methods, and comparing with the density of confirmed COVID-19 cases. The range of the estimated counts and the actual counts stays within a similar range. Furthermore, It shows that the estimated count distribution across the U.S counties is almost identical to the actuals.

The Pearson's and Spearman's correlation coefficients $((\rho^P, \rho^S) = (1, 0.999))$ from both methods for the confirmed vs. expected indicate that the estimated values are significantly accurate, could slightly be overestimated. Overall, it shows that the performance of the CHBC method is as better as the GLMM. Similarly, the second column of the Figure E.1 with 15 categories, 0-11, 12-23, 14-41, 42-63, 64-95, 96-139, 140-200, 201-290, 291-389, 390-574, 575-871, 872-1439, 1440-

Figure 6.4: Actual vs. estimated COVID-19 cases based on CHBC and GLMM at county-level and state-level.



3014, 3015-8442, 8443-225148, shows that the estimations based on both CHBC and GLMM are accurate. In summary, it shows that both methods had performed significantly better determining the expected number of counts.

Additionally, we compared the confirmed and estimated counts for each county for both CHBC and GLMM. Figure 6.4(a) shows the actual versus estimated counts for both methods at the county-level, and Figure 6.4(b) represents the actual vs. estimated at the state-level. Both plots imply that the estimations from the CHBC and GLMM are very much alike, implying that both methods provide significantly precise results in estimating COVID-19 cases. The estimations are very accurate at both the lower end as well as at the higher end. The large R^2 values indicate the accuracy of the estimations, except its being almost one show some sort of overestimation from both approaches. It might be possible for various reasons, such as many variables in the model, colinearity, trend over time, etc.

Figure 6.5: Box plots for the estimated random effects based on the CHBC and GLMM.

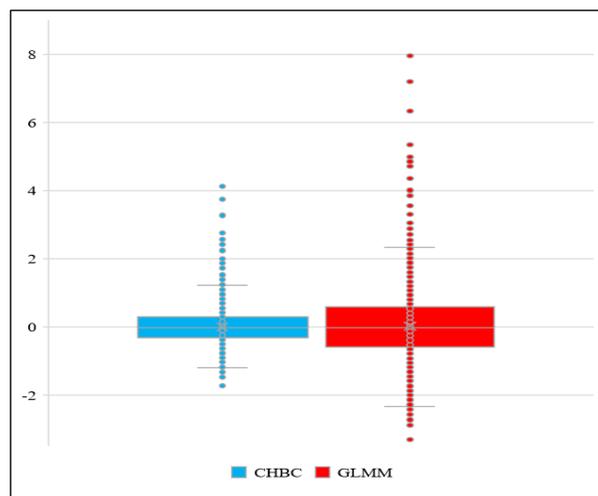


Figure 6.5 displays the distribution of estimated random effects based on the proposed CHBC method and GLMM. The estimated distributions are slightly different, with a widespread interquartile range for the estimated random effects from GLMM, indicating a wider variation in the random effects from GLMM than the CHBC method.

Based on the simulation studies and the real data analysis result imply that the proposed CHBC method performs well and

the results improve as sample m and n_m increase, except the underestimated variance parameters in SAE.

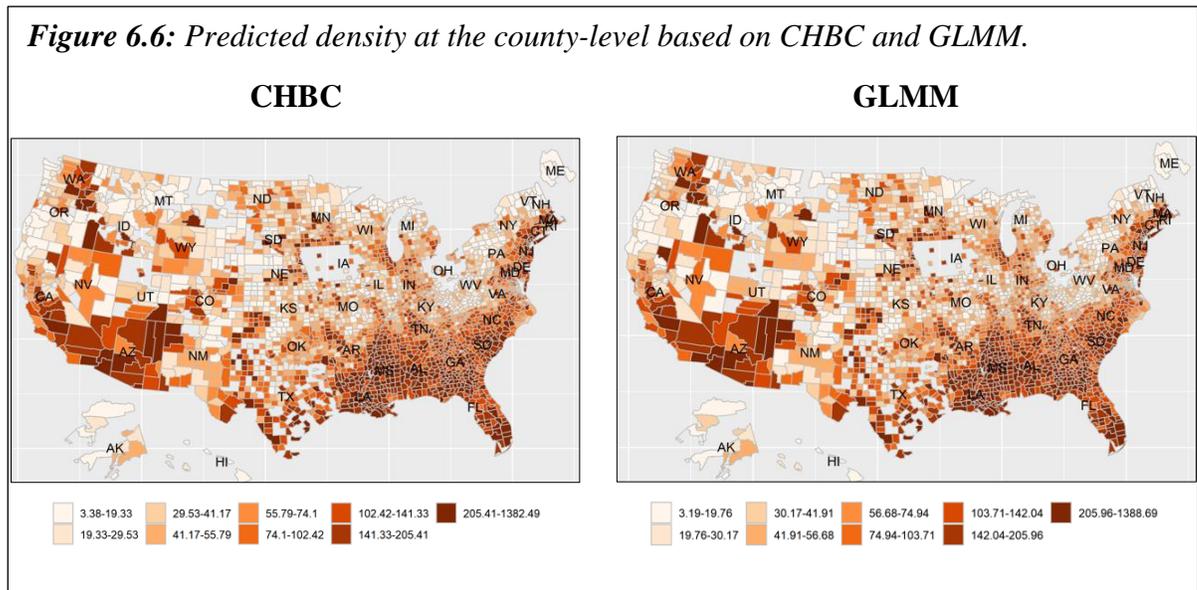
Finally, we obtained the predicted density ($\hat{\lambda}_i^{pred}$) at the county-level for a new variable.

The predicted counts at the county-level can be estimated using

$$\hat{\lambda}_i = \exp(\mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{Z}\hat{\mathbf{u}} + \log(\text{Population}_i)), i = 1, \dots, 2992.$$

$$\hat{\lambda}_i^{pred} = \frac{\hat{\lambda}_i}{\text{Pop}_i} \times 10,000 \tag{6.4}$$

Figure 6.6. displays the predicted density for COVID-19 cases per 10,000. Compared to Figure E.1, the predicted density indicates that some of the high-density count regions have decreased the impact. However, some counties are still in high-density areas.



6.2.1 Model Selection

The model selection is based on the AIC value of the Poisson model evaluated using the estimated model parameters. The lower the AIC value, the better the model performance. First, consider Stirling's approximation (Namias, 1986), for large y ,

$$y! \sim \sqrt{2\pi y} \left(\frac{y}{e}\right)^y.$$

Now, log-likelihood of $y|u$ is obtained

$$\begin{aligned} \ell_{y|u} &= \sum_{i=1}^N y_i \log(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u} + \text{offset}_i) - (\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u} + \text{offset}_i) - \log\left(\sqrt{2\pi y_i} \left(\frac{y_i}{e}\right)^{y_i}\right) \\ \ell_{y|u} &= \sum_{i=1}^N y_i \log(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u} + \text{offset}_i) - (\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{u} + \text{offset}_i) - \frac{1}{2} \log(2\pi y_i) - y_i \log y_i + y_i \end{aligned}$$

Now the AIC is obtained using

$$\text{AIC} = -2 \left(\left(h + \frac{1}{2} \log \det \mathbf{J}^{-1} \right) + s_d \right),$$

where,

$$h = \ell_{y|u} + \ell_{\mathbf{u}},$$

p is the number of fixed effects, m is the number of random effects, s_d is the number of model parameters estimated using p and m , and $\ell_{\mathbf{u}}$ is the log-likelihood of \mathbf{u} .

Table 6.5: Model selection criteria based on CHBC and GLMM

Model	AIC	logLik
CHBC	25201	-12599
GLMM	27948	-13932

Table 6.5 displays the AIC and log-likelihood (logLik) values from both CHBC and GLMM methods. However, AIC has been calculated slightly differently in the GLMM and CHBC method. Thus, it is not comparable to model selection in this scenario. Still, the proposed method

has a low AIC value of a 25201, and GLMM with a value of 27948. The total likelihood from the CHBC is -12599 and 13932 from the GLMM.

6.3 Multivariate Joint Modeling Using COVID-19 Data

In this section, we illustrated the proposed joint modeling approach through multivariate random effects using the COVID-19 data for cases and deaths at the county-level in the US, which was downloaded from the data repository created by the John Hopkins University Center for Systems Science and Engineering (JHU CSSE) (Dong et al., 2020). In section 6.2, we applied the univariate CHBC method to estimate COVID-19 cases. However, it is vital to consider the association between COVID-19 cases and deaths. Therefore, we adopted the bivariate joint model with two variables of interest; confirmed COVID-19 cases and COVID-19 deaths based on data as of July 31, 2020, in the US. The same auxiliary data used in section 6.2 is considered for the multivariate joint model as well. We considered two additional variables, “Number of days since the first recorded death (log)”, and “population density per square mile” in the bivariate joint model.

Table 6.6 shows the MHLEs for fixed effects from the multivariate joint model for COVID-19 cases, and Table 6.7 similarly gives the MHLEs related to COVID-19 deaths. The results indicate that there is a significant impact on the percentage of people not staying at home and the number of different mileage trips. It shows that a strong positive association with people not staying at home ($\beta = 2.994, p < 0.001$), indicating higher COVID-19 cases than people staying at home after adjusting for other covariate information in the model. From Figure 6.2 (b), it is clear that the between 2-3 weeks lagged values of percentage of not staying home are significantly associated with the increase in cases in the following weeks.

Table 6.6: Maximum hierarchical likelihood estimates (MHLEs) of fixed effects based on the multivariate joint model for COVID-19 cases.

		Estimate	SE	Z Value	P(> Z)	95% CI	
						Lower	Upper
Intercept		-9.653	0.135	-71.679	< 0.001	-10.372	-8.934
Age (%)	<= 17 years	0.000					
	18-29 years	-2.272	0.031	-72.428	< 0.001	-2.619	-1.925
	30-44 years	-3.393	0.049	-69.322	< 0.001	-3.827	-2.960
	45-64 years	-2.924	0.048	-61.073	< 0.001	-3.352	-2.495
65 years or above living alone (%)		0.805	0.044	18.358	< 0.001	0.394	1.215
Gender (%)	Female	0.000					
	Male	8.557	0.054	158.920	< 0.001	8.102	9.012
Number active cases (log)		0.279	0.001	394.651	< 0.001	0.227	0.331
Number days cases (log)		-0.516	0.007	-74.463	< 0.001	-0.679	-0.353
Number of trips (%)		0.000					
	less than 1 mile						
	1-4 miles	1.082	0.041	26.543	< 0.001	0.686	1.478
	5-9 miles	-1.945	0.043	-45.225	< 0.001	-2.351	-1.538
	10-24 miles	-1.168	0.041	-28.652	< 0.001	-1.564	-0.772
	25-49 miles	2.276	0.056	40.910	< 0.001	1.814	2.739
	50-99 miles	2.140	0.101	21.110	< 0.001	1.516	2.764
	100 miles above	6.221	0.176	35.262	< 0.001	5.398	7.045
Population per square mile (%)		0.002	0.000	87.547	< 0.001	-0.008	0.013
Population not stay home (%)		2.994	0.029	103.588	< 0.001	2.661	3.327
Poverty level (%)		0.621	0.021	29.511	< 0.001	0.337	0.905
Race (%)	white	0.000					
	African American	0.628	0.007	86.314	< 0.001	0.461	0.795
	Hispanic	0.924	0.007	125.975	< 0.001	0.756	1.091
	Other	0.351	0.006	58.316	< 0.001	0.199	0.503
Unemployment rate (%)		2.389	0.054	44.630	< 0.001	1.936	2.842

The coefficient estimates for “the population per square mile” indicate a positive association with COVID-19 cases ($\beta = 0.002, p < 0.001$) and also with deaths ($\beta = 0.003, p < 0.001$). Furthermore, there is a significant positive association of people not staying at home ($\beta =$

1.824, $p < 0.001$) with COVID-19 deaths compared to staying home, which might be explained based on the total number of deaths by age group ("Provisional COVID-19 Death Counts by Sex, Age, and State," 2020). The majority of older or retired people are more likely to not stay at home; having group gatherings, eat-outs, etc. is also more likely to get infected with the virus. The number of deaths among them is high compared to younger people.

Nevertheless, older people with comorbidity problems or severe health conditions are more likely to stay home; however, they may infect the virus by their family visitors, which eventually increases the mortality. The results indicate that the 65 years or above living alone has a strong positive association with COVID-19 cases ($\beta = 2.994, p < 0.001$) as well as with COVID-19 deaths ($\beta = 4.042, p < 0.001$). Based on ACS, 2016, 65 years or above, people are more likely to live alone, more than 20% older population, in the US than elsewhere in the world (Roberts, Ogunwole, Blakeslee, & Rabe, 2018). The majority among them are living independently, doing errands alone, such as medical appointments, groceries, etc. might lead to exposing and getting infected with the virus, which could also be a risk factor for the increase in COVID-19 cases among older people. Worldwide, the health officials strongly recommended for the population at high-risk to practice extra preventive measures as they potentially experience severe illness than other humans if they get exposed to the virus ((CDC), 2020; Razzaghi et al., 2020). However, the risk for the age groups 18-29 and 30-44 is not significant compared to the individuals 17 years or below.

Besides the impact of not staying at home proportion and 65 years and above living alone, it also shows that the number of various mileage of trips is also a significant risk factor for seeing a higher number of cases. It is expected to see more cases from the population that made 100 miles or above mileage trips compared to less than 1-mile trips. From Figure 6.3, the top ten states with a higher number of 100 miles or above trips, North Dakota, Pennsylvania, Michigan, Virginia, Minnesota, Connecticut, Iowa, Nebraska, and Kentucky are still an increasing trend of cases. The majority of states have a large number of 100 miles or above trips in March, a slight decrease in April, and again a significant increase in June and July, which explains for most of the states to see

the peak of cases in June and July as presented in BTS ("Daily Travel during the COVID-19 Public Health Emergency," 2020).

Table 6.7: Maximum hierarchical likelihood estimates (MHLEs) of fixed effects based on the multivariate joint model for COVID-19 deaths.

	Estimate	SE	Z Value	P(> Z)	95% CI	
					Lower	Upper
Intercept	-14.802	0.326	-45.390	< 0.001	-15.921	-13.682
Age (%)						
<= 17 years	0.000					
18-29 years	-4.219	0.206	-20.511	< 0.001	-5.108	-3.330
30-44 years	-7.726	0.310	-24.941	< 0.001	-8.817	-6.635
45-64 years	1.555	0.291	5.338	< 0.001	0.497	2.612
65 years or above living alone (%)	4.042	0.273	14.820	< 0.001	3.018	5.065
Gender (%)						
Female	0.000					
Male	4.728	0.414	11.430	< 0.001	3.467	5.989
No active cases (log)	0.314	0.004	79.459	< 0.001	0.191	0.437
Number of days deaths (log)	0.137	0.013	10.765	< 0.001	-0.084	0.358
Number of trips (%)						
less than 1 mile	0.000					
1-4 miles	0.634	0.258	2.452	< 0.001	-0.363	1.630
5-9 miles	-2.193	0.269	-8.147	< 0.001	-3.210	-1.176
10-24 miles	-1.186	0.254	-4.666	< 0.001	-2.175	-0.198
25-49 miles	2.737	0.347	7.899	< 0.001	1.584	3.891
50-99 miles	4.306	0.620	6.947	< 0.001	2.763	5.849
100 miles above	3.360	1.205	2.790	< 0.001	1.209	5.512
Population per square mile (%)	0.003	0.000	15.629	< 0.001	-0.023	0.028
Population not stay home (%)	1.824	0.176	10.361	< 0.001	1.001	2.646
Poverty level (%)	1.520	0.128	11.919	< 0.001	0.820	2.220
Race (%)						
white	0.000					
African American	1.120	0.044	25.418	< 0.001	0.709	1.532
Hispanic	1.457	0.048	30.500	< 0.001	1.029	1.886
Other	1.245	0.031	39.945	< 0.001	0.899	1.591
Unemployment rate (%)	3.431	0.373	9.199	< 0.001	2.234	4.629

Similarly, the trips of 25-50, 50-100, 1-5 mileage also have a significant positive association with an increase in cases compared to less than 1-mile trips with multiplicative factors of 47.58, 19.19, and 4.60, sequentially, after controlling for other risk factors. Likewise, a strong positive association with the number of deaths as well. Based on 2020 travel trends, it shows that both Gen X (born between 1965 and 1980) and Boomers (born between 1946 and 1964) have more travel compared to Millennials, similar to annual studies over the last six years (Fry, 2016; Levy, January 2020). There are many types of trips, such as international and domestic, among Gen X and Boomers than the Millennials. Thus, it was evident to see more cases and deaths among them. California, Texas, Florida, New York, Illinois, Georgia, Ohio, North Dakota, Pennsylvania, and Michigan are among the top ten states for these three mileage trip groups, which are also the states with high COVID-19 cases. However, the impact from the trips within 5 miles to 25 miles is lower compared to less than 1-mile trips, which mainly depends on the state, such as in some states, 5-25 mileage trips might be more frequent while in other states it might not.

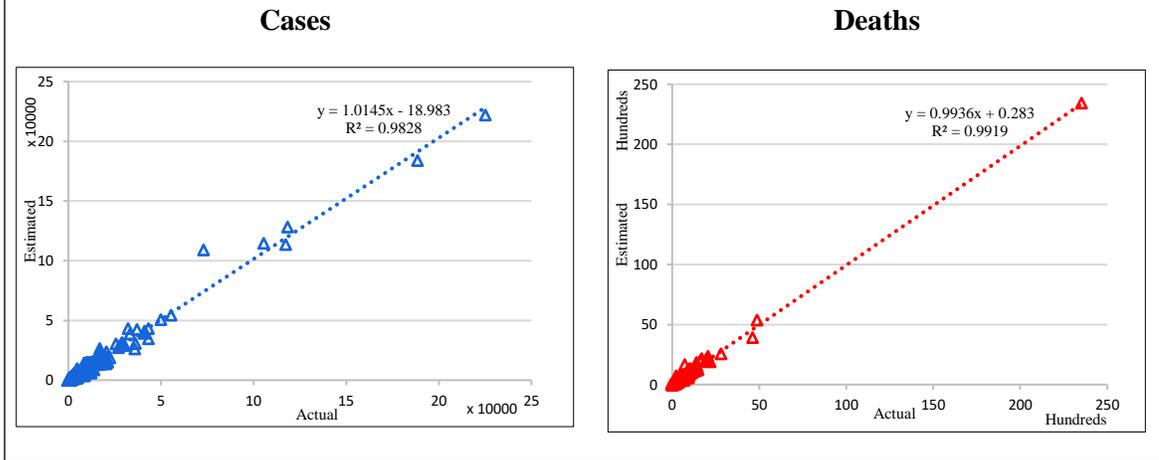
Furthermore, the estimated coefficients for the unemployment rate ($\beta = 2.389, p < 0.001$) and poverty rate ($\beta = 0.621, p < 0.001$) show strong positive associations with COVID-19 cases. Similarly, the coefficient effects $\beta = 3.431, p < 0.001$, and $\beta = 1.520, p < 0.001$ for both the unemployment rate and poverty level also indicate that they are high-risk factors for deaths due to COVID-19. In section 6.2, univariate modeling with the county-level random effect model shows a negative correlation of unemployment rate with COVID-19 cases, which explains with a low unemployment rate in high COVID-19 counties. However, the majority of the states with high COVID-19 also have a high unemployment rate, which expects to observe a positive correlation.

Also, Hispanics, African Americans, and other races are positively correlated with COVID-19 cases than Whites with estimated coefficients $\beta = 0.924, p < 0.001$, $\beta = 0.628, p < 0.001$, and $\beta = 0.351, p < 0.001$, respectively. Similarly, we observed a strong positive correlation of Hispanics ($\beta = 1.457, p < 0.001$), African Americans ($\beta = 1.120, p < 0.001$),

and other races ($\beta = 1.245, p < 0.001$) with COVID-19 deaths than Whites. However, as described in section 6.2, there might be some confounding effects in the variables, which shows in Figure E.2 (a) and (b) based on the univariate mixed model fit. From Figure E.2(b), VIF values for Whites (11.95), African Americans (5.79), age 30-44 (5.55), age 45-74 (5.36), and Hispanics (5.24) are somewhat large, indicating the existence of a moderate correlation with one or more variables in the model. It could result in some biasedness in the regression coefficients. But, we did not account for the multicollinearity in this analysis.

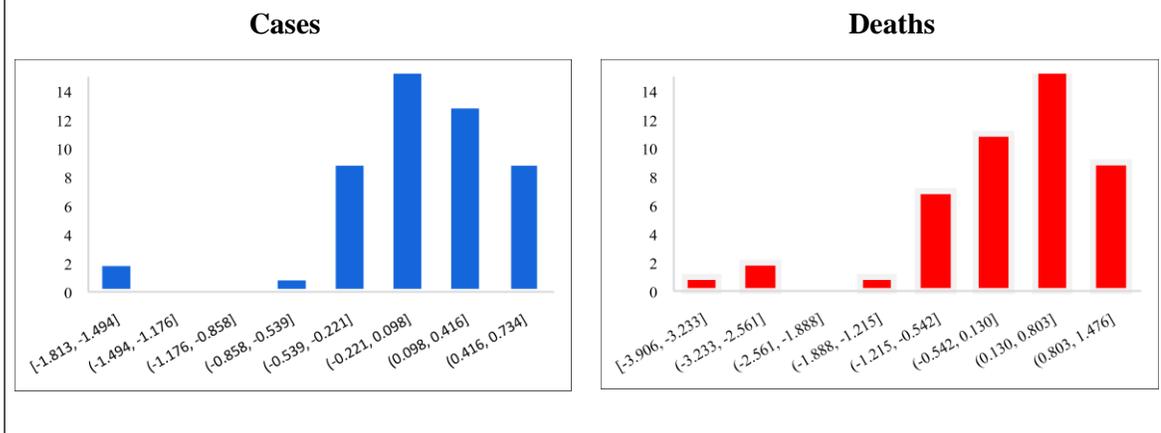
Furthermore, there is a higher risk of dying from COVID-19 for the age group 45-64 years ($\beta = 1.555, p < 0.001$) compared to less than 18 years old. It also shows that a strong association with both cases and deaths among males compared to females with multiplicative factors of $\beta = 8.557, p < 0.001$, and $\beta = 4.728, p < 0.001$ by controlling for other risk factors. The number of active cases is also positively associated with COVID-19 cases and with deaths with coefficient estimates $\beta = 0.279, p < 0.001$, and $\beta = 0.314, p < 0.001$, respectively. Additionally, the number of days since the first reported case (log scale) is positively associated with COVID-19 cases ($\beta = -0.516, p < 0.001$), and also the number of days since the first reported death is positively associated with deaths ($\beta = 0.137, p < 0.001$). To conclude the model estimates, the coefficient effects from each variable related to both outcomes are valid with the current COVID-19 prevalence in the US.

Figure 6.7: Estimated vs. actual COVID-19 cases at county-level based on joint CHBC model through multivariate random effects.



We obtained the observed vs. estimated county-level prevalence, which is displayed in Figure 6.7. It shows that the model estimates are accurate on both COVID-19 cases and deaths with large R^2 (0.98, 0.99) values for both. Additionally, the estimated cases and deaths are consistently accurate in the lower and higher end of counts. The distributions of the estimated state level random effects for both COVID-19 cases and deaths displayed in Figure 6.8 indicate slightly left (negatively) skewed distributions.

Figure 6.8: Distribution of estimated random effects based on joint CHBC model through multivariate random effects.



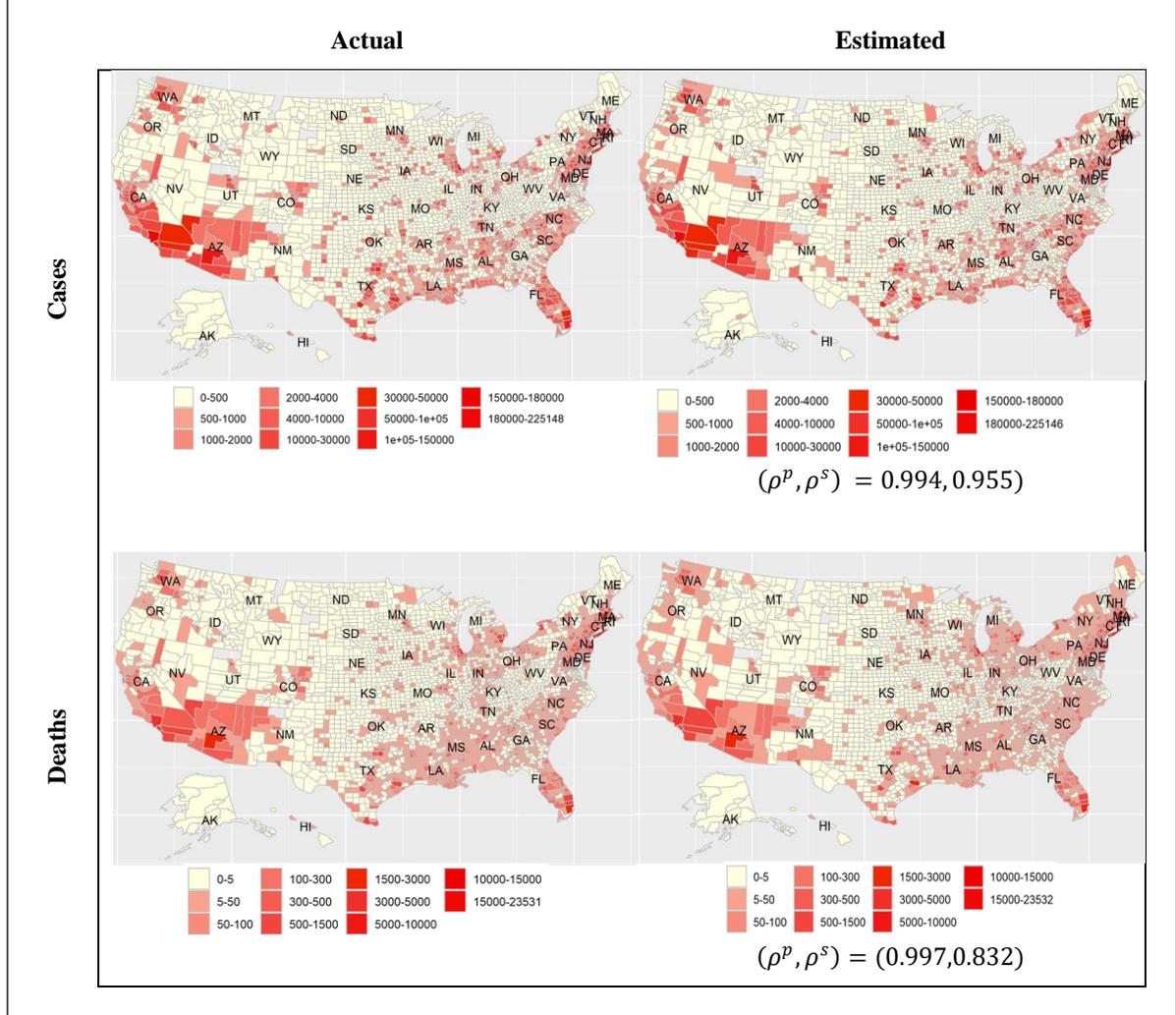
Based on the estimated variance-covariance matrix of bivariate normally distributed random effects, it shows that there exists a negative correlation between and within county variation with the number of cases and deaths. The estimated variances for both random effects ($\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$) are 1.262, and the correlation among them is 0.262. It indicates the existence of a positive association between COVID-19 cases and deaths. Thus, the joint modeling approach is more appropriate for modeling the correlated outcomes. Conducting separate analyses might lose vital information in the estimations.

$$\hat{\Sigma} = \begin{pmatrix} 1.262 & 0.262 \\ 0.262 & 1.262 \end{pmatrix}$$

Next, we calculated the county-level estimates for COVID-19 cases and deaths using the county-level covariate information. The estimated counts are mapped for each county, as displayed in Figure 6.9. We grouped the cases and deaths into 11 categories to have a better understanding of distribution within the US counties. It shows that the majority of counties have cases within the 0-500 range, specifically in northern areas and also some in the Midwest areas. It also shows that the number of deaths in most counties stays within the 0-5 range. Overall, the range of the actual number of confirmed cases is 0 – 225,148, and from the CHBC model estimates 0 – 225,146. Furthermore, the actual deaths range from 0 – 23,531, and the model estimates 1 – 23,532. Pearson's (Spearman's) correlation coefficient(s) for observed vs. estimated cases are 0.994 (0.955) and for deaths 0.997 (0.832). The most counties in Southwest, Northeast, and Southeast are the hot spots with more COVID-19 cases and deaths.

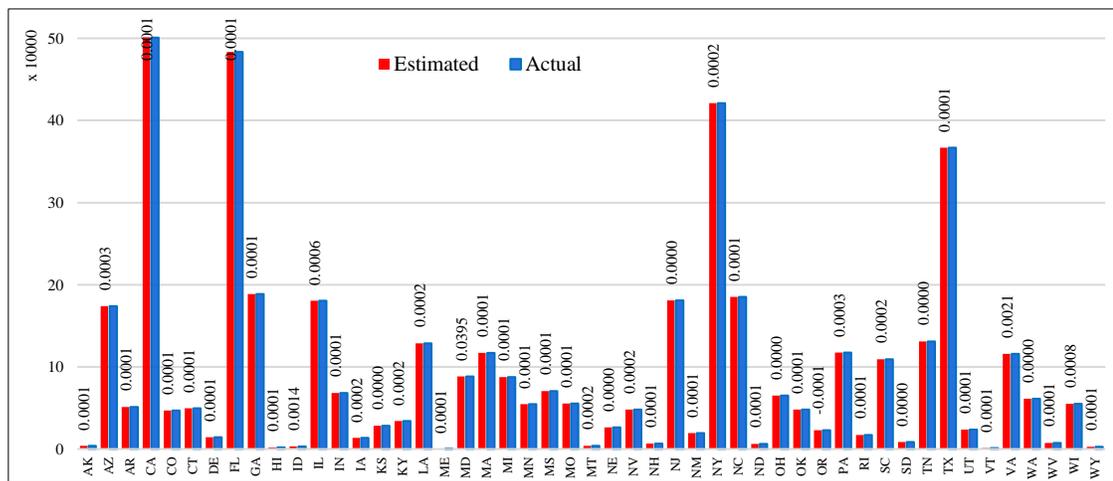
However, the model estimates are accurate for most counties, staying within the same range except for very few counties. Thus, we calculated the total number of cases, and deaths for each state are displayed in Figure 6.10 and 6.11. Based on Figure 6.10, California, Florida, New York, and Texas states have a significantly higher number of cases compared to other US states.

Figure 6.9: Estimated vs. actual COVID-19 at county-level based on joint CHBC model through multivariate random effects.



The next top states are Arizona, Georgia, Illinois, New Jersey, and North Carolina. The lowest COVID-19 cases are in Alaska, Hawaii, Idaho, Maine, Montana, Vermont, Wyoming, New Hampshire, North Dakota, West Virginia, and South Dakota.

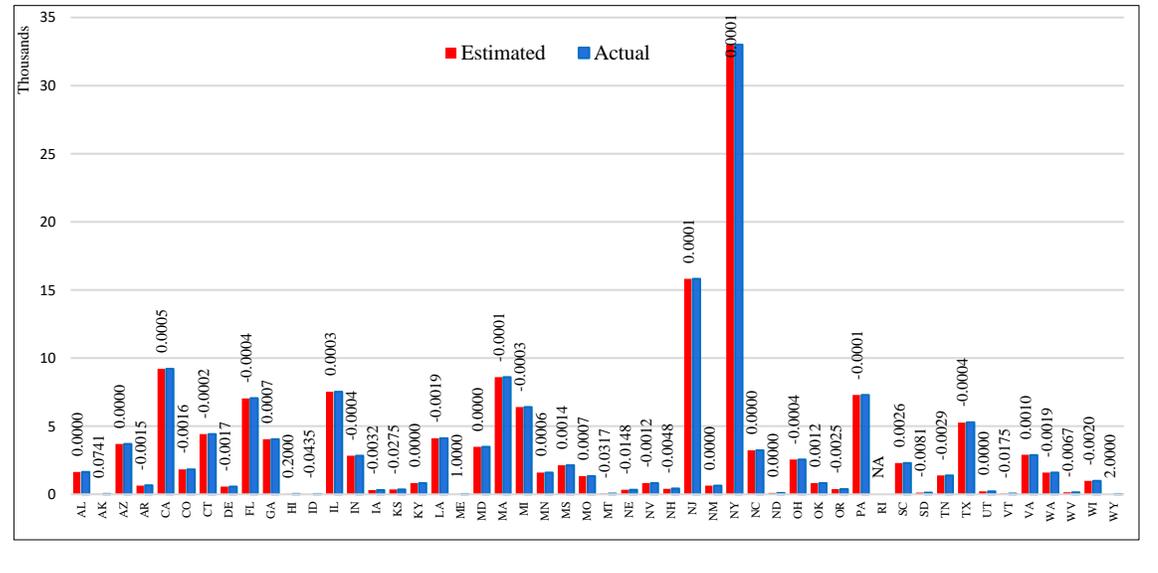
Figure 6.10: Estimated vs. actual COVID-19 cases at state-level based on multivariate joint CHBC model.



The values in each bar represent the relative bias for each state. Ex. Relative bias of Alaska is 0.0001, Arizona is 0.0003.

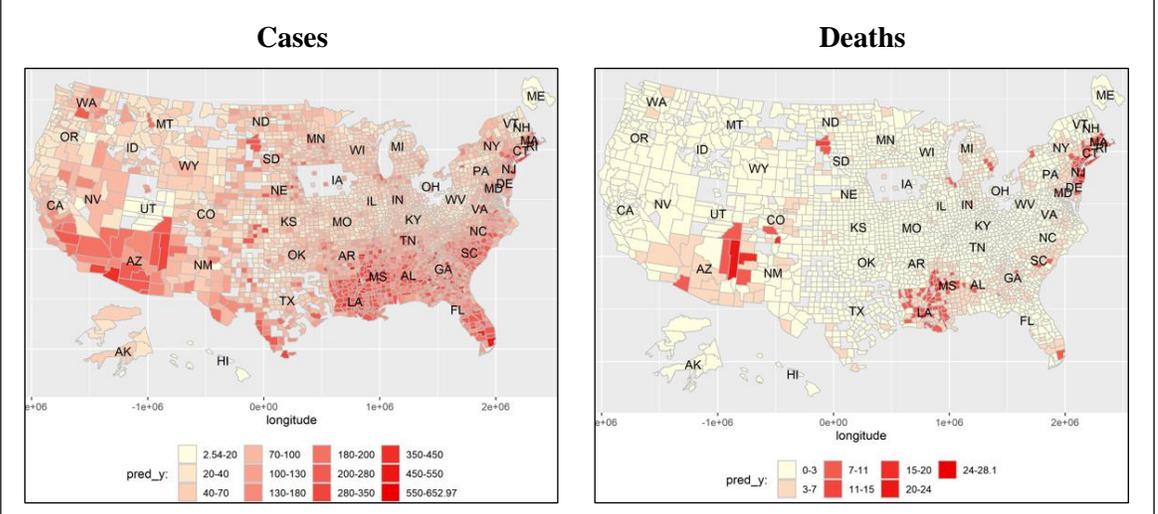
Figure 6.11 shows the observed vs. the estimated deaths, including the relative bias at the state-level. The relative bias is very low for each state estimate. The highest relative bias is 7.41%, and most of the values are significantly lower than the most elevated amount, indicating that the model performance is better assessing the state-level mortalities. New York state has the highest number of deaths, followed by New Jersey. Alaska, Arizona, Louisiana, Massachusetts, New York, North Dakota, Ohio, and Vermont states have a 100% match of the estimated number of deaths. In conclusion, the model performance is very accurate, estimating the number of cases and deaths at the state-level compared to the county-level.

Figure 6.11: Estimated vs. actual COVID-19 deaths at state-level based on multivariate joint CHBC model.



Finally, we obtained the predicted density for the number of cases and deaths per 10,000 population using equation (6.4). Figure 6.12 shows that the county-level predicted density for the cases indicates that the risk in most US counties is still in a considerable range. Mostly, southeast, southwest regions, and also some counties in the midwest are more likely to have a higher number

Figure 6.12: Predicted density of COVID-19 at county-level based on multivariate joint model using the CHBC method (per 10,000).

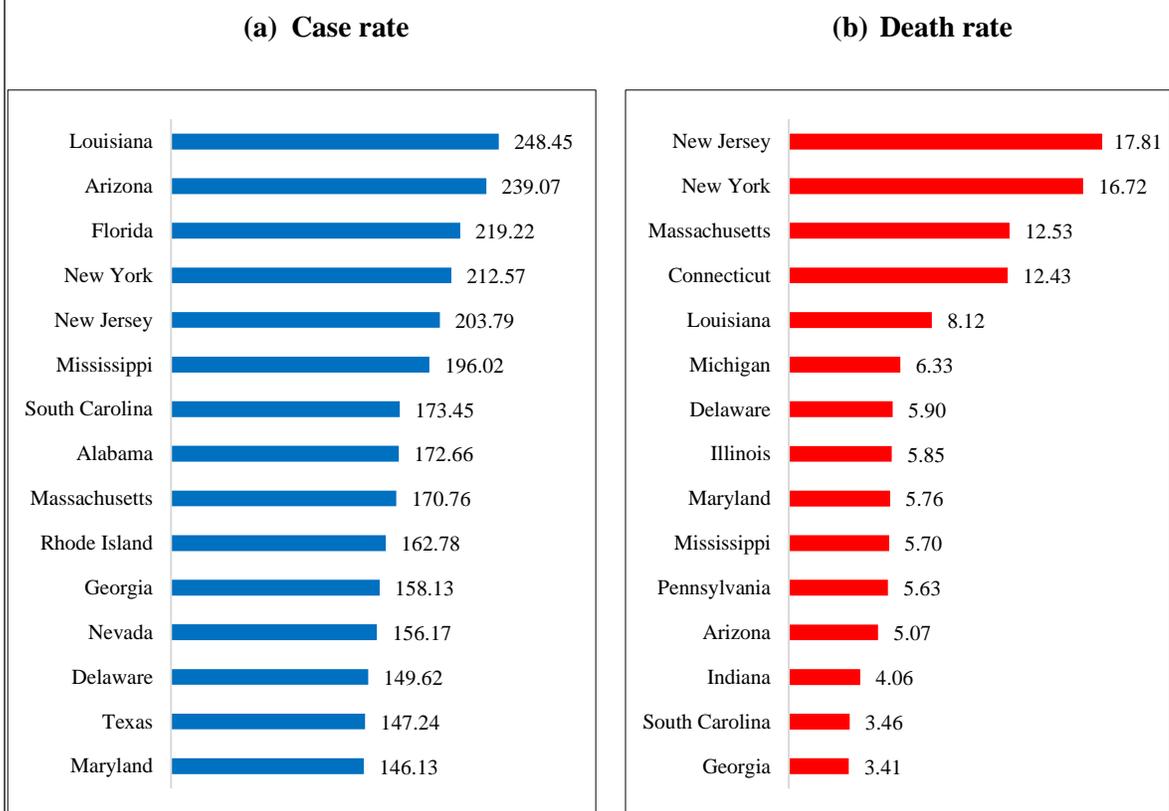


of cases. Based on the predicted density for deaths, it shows that most counties are in 0-3 deaths per 10,000, while some counties in Arizona, Louisiana, Mississippi, and in northeast areas have a higher range of deaths. We obtained the case rate and the death rate each state using the formula given below

$$\text{Case (Death) rate}_i = \frac{\text{Total cases}_i(\text{Deaths}_i)}{\text{Pop}_i} \times 10,000, i = 1, \dots, 50.$$

As shown in Figure 6.13, some states have a lower death rate compared to case rate, such as case rate is highest in Louisiana (248.45), and the death rate is highest in New Jersey (17.81) as

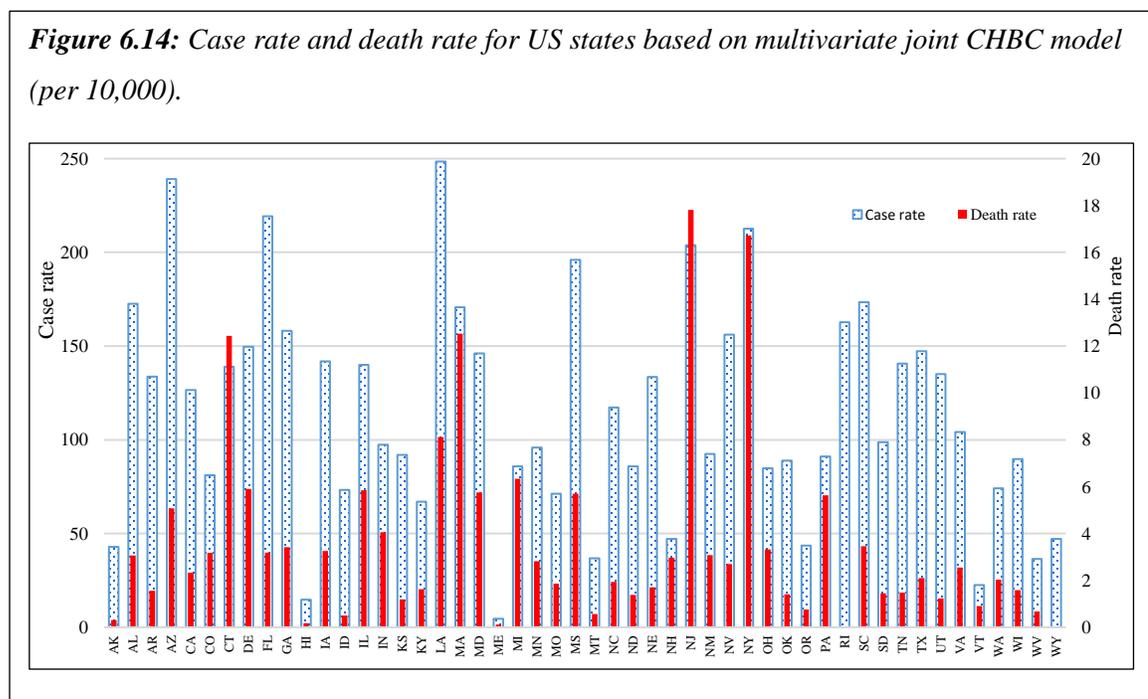
Figure 6.13: Case rate and death rate for top 15 states based on multivariate joint CHBC model (per 10,000).



COVID-19 Case rate and death rate are obtained using the county-level predicted density obtained using equation (6.4).

of July 31, 2020. The next four states with higher case rates are Arizona (239.07), Florida (219.22), New York (212.57), and New Jersey (203.79), and the states with higher death rates are New York (16.72), Massachusetts (12.53), Connecticut (12.43), and Louisiana (8.12).

Figure 6.14 represents the case rate and death rate per 10,000 population for all US states, which shows that some states with higher COVID-19 cases have lower death rates, while some have considerably high death rates. Though New Jersey has the highest death rate, followed by New York, Louisiana and Arizona have the highest case rates with lower death rates. The five states that have the lowest death rates are Rhode Island (0.00), Wyoming (0.02), Maine (0.12), Hawaii (0.18), and Alaska (0.31). The five states with the lowest care rates are Maine, Hawaii, Vermont, West Virginia, and Montana, with values 4.58, 14.75, 22.55, 36.54, and 36.82, respectively.



6.4 Joint Modeling Through Shared Random Effects Using Tobacco Smoking Data

We evaluated the proposed joint modeling approach through shared random effects using the same tobacco smoking data set described in section 6.1, considering the current-use of E-Cigarettes and the ever-use of E-Cigarettes. In section 6.1, we observed that the counties with a higher (lower) estimated prevalence of current-use, also seem to have a higher (lower) prevalence of the ever-use of E-Cigarettes. Accordingly, it suggests that the existence of the correlation between these two outcomes, hence it would be beneficial to model them jointly to study the association, as well as to obtain accurate estimations for both. We considered the joint model through shared random effects to examine the association between the current-use and ever-use of E-Cigarettes, which explained by the shared parameter $\hat{\gamma}$.

The MHLEs from the joint model for fixed effects of both outcomes using the same tobacco smoking data set are displayed in Table 6.8. Compared to the coefficient effects in Table 6.1, the values for each variable are different in the joint model output in Table 6.8, even though they are directionally the same. However, the estimated shared parameter $\hat{\gamma} = 2.966 (> 1)$ indicates the existence of a positive association between the current-use and the ever-use of E-Cigarettes. Hence, it is critical to consider joint modeling to make inferences on these two variables of interest.

Based on the joint model, African Americans have 1.38 times higher current-use of E-Cigarettes prevalence compared to other races after adjusting for other variables. Still, it is 0.78 times higher based on the univariate analysis. The estimated prevalence of the ever-use of E-Cigarettes is 1.08 times higher among African Americans than other races based on the joint model, indicating that slightly higher than the effect from the univariate analysis with a value of 0.93. Furthermore, the poverty rate is a significant variable, with a p-value being <0.001 from the joint model (National Cancer Institute, 2017; Prevention). However, the p-value is not significant ($>$

0.05) based on the univariate analysis for both current-use and ever-use models. The joint model results indicate that adolescents in lower poverty levels tend to smoke more than a higher poverty level.

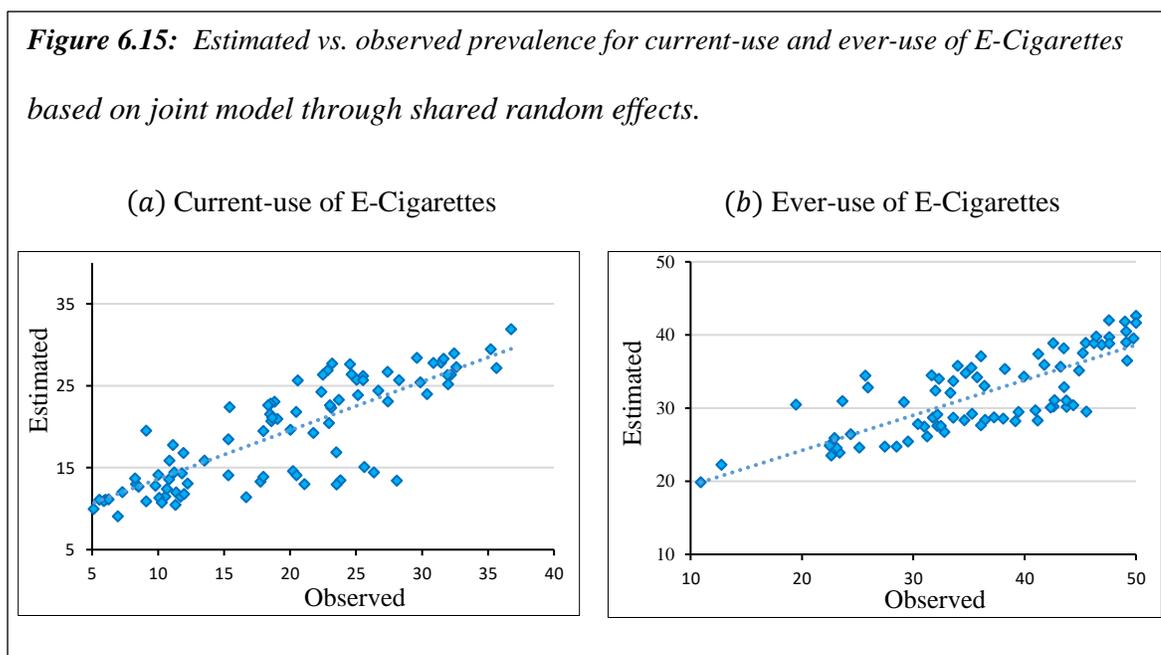
Table 6.8: Fixed effect estimates for current-use and ever-use of E-Cigarettes based on the joint model through shared random effects.

	Current-use of E-cigarettes				Ever-use of E-cigarettes			
	Estimate	SE	Z Value	P(> Z)	Estimate	SE	Z Value	P(> Z)
Intercept	-1.709	0.087	-13.344	<0.001	-0.908	0.068	-13.344	<0.001
<= 14 yrs	0.000				0.000			
Age 15-17 yrs	0.227	0.054	8.823	<0.001	0.374	0.042	8.823	<0.001
>= 18 yrs	0.475	0.055	16.271	<0.001	0.706	0.043	16.271	<0.001
Race White	0.042	0.323	1.906	0.0567	0.477	0.250	1.906	0.0567
African American	0.325	0.054	1.778	0.0754	0.076	0.043	1.778	0.0754
Hispanic	-0.304	0.071	-1.460	0.1444	-0.076	0.052	-1.460	0.1444
Others	0.000				0.000			
Sex Male	0.000				0.000			
Female	0.113	0.057	6.892	<0.001	0.309	0.045	6.892	<0.001
Year 2015	0.000				0.000			
2017	0.283	0.031	6.095	<0.001	0.155	0.025	6.095	<0.001
Poverty Rate (%)	-0.827	0.034	-19.425	<0.001	-0.509	0.026	-19.425	<0.001

Furthermore, compared to 2015, it shows that the prevalence of both current-use and ever-use of E-Cigarettes is 1.33 and 1.17 times higher in 2017, which is contrary to univariate model outcomes. From the joint model it shows that the current-use and ever-use of E-Cigarettes are more popular among youth females (1.12 and 1.36 times higher) compared to youth males. Besides, the

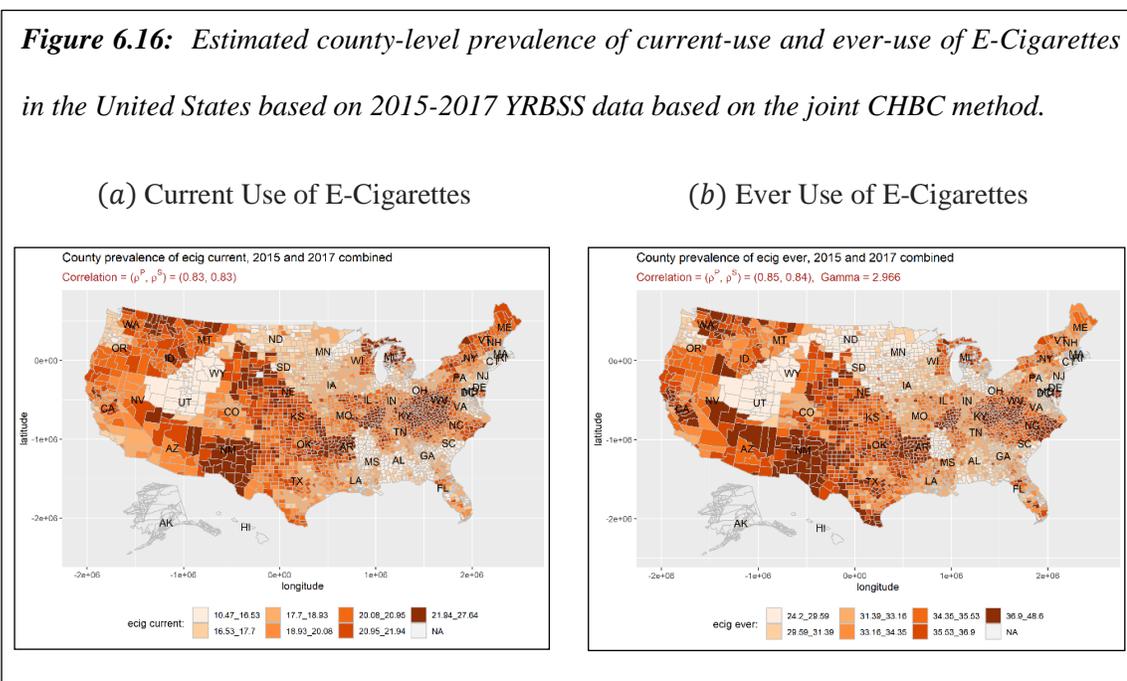
18 years or above are more likely to use E-Cigarettes currently compared to 14 years or less with a multiplicative factor of 1.61. Similarly, the prevalence of ever-use also two times higher in 18 years or above more than 14 years or younger. In conclusion, it shows that those who have a higher prevalence for current-use also have a higher prevalence for every-use of E-Cigarettes.

The observed vs. estimated prevalence for current-use and ever-use of E-Cigarettes from Figure 6.14 implies that the model performs well, determining the prevalence at both higher and also at the lower end. We used the estimated county-level random effects and obtained the random effects for missing counties as described in section 6.1 using equations (6.1), then calculated the county-level prevalence using (6.2), and (6.3).



The observed vs. estimated prevalence for current-use and ever-use of E-Cigarettes from Figure 6.15 implies that the model performs well, determining the prevalence at both higher and also at the lower end. We used the estimated county-level random effects and obtained the random effects for missing counties as described in section 6.1 using equations (6.1), then calculated the county-level prevalence using (6.2), and (6.3).

Figure 6.16 shows the estimated county-level prevalence of current-use and the ever-use of E-Cigarettes. The areas with a higher prevalence of current-use also seem to have a higher value for ever-use, which is evident from the model estimates. The current-use of E-Cigarettes in the US ranges from 10.47% - 27.64%, indicating that it is more prevalent in the Midwest and Northwest regions compared to the Southeast and North regions. It is also higher in the Southwest than the Southeast and North regions. Similarly, the prevalence of ever-use E-Cigarettes too more prevalent in the Southwest, Midwest, Northwest, and Northeast compared to Southeast and North regions in the US based on 2015 and 2017 data.



Both current-use and ever-use prevalence are lowest in the northern areas. Both plots (a) and (b) shows that the estimated prevalence for both outcomes in most counties in Utah stays in the lowest quantile. The data was only available in Salt Lake County, Utah, in 2017, which had a prevalence of 2% for current-use and 13% for ever-use. We estimated the random effects for all missing counties through the Nearest Neighboring method and obtained the prevalence for all other counties in Utah. The estimated values reveal the state’s lower range prevalence of E-Cigarette

usage based on the Public Health Indicator Based Information System, Utah (IBIS). Also, the other counties with lower and higher values consistent with the county-level prevalence, indicating that the proposed CHBC joint model performs well.

Chapter 7. Discussion, Limitations, and Future Work

7.1 Discussion

In this work, we proposed a statistical methodology in SAE based on the Hierarchical (h) likelihood with a calibrated parameter estimation procedure, named the CHBC approach. The proposed CHBC method does not require a closed form of the log-likelihood function, does not involve any computationally intensive integral approximations, hence it is computationally efficient. The proposed CHBC approach provides accurate parameter estimates for fixed effects, random effects, and dispersion parameters using h -likelihood through iterative approximation techniques with bias correction. Unlike in the standard linear mixed model, the model (2.2) does not have a closed-form for the joint log-likelihood. Hence it is very challenging to estimate the BLUP or EBLUP of model parameters. In such situations, h -likelihood plays a significant role in simplifying the parameter estimation procedure. The MHLEs are often obtained via numerical approximation methods due to intractable integrals in the joint log-likelihood function.

First, in chapter 1, a brief introduction to SAE and SAE methods were discussed. Then, specifically in chapter 2, we illustrated the widely-used GLMM and parameter estimation of the fixed effects, random effects, and dispersion parameters. The last section in chapter 2 covers the literature review of the research work based on current models, their limitations, and motivation to the proposed technique. Then, in chapter 3, we described the proposed CHBC method for univariate models, which can be applied to a general class of exponential family distributions. The maximum hierarchical likelihood estimation process through iterative approximation for model parameters is discussed in detail. The asymptotic properties of MHLEs and Wald confidence intervals are discussed in section 3.5.

We extended the proposed CHBC method to joint modeling of multiple outcomes based on two ways; 1. through multivariate random effects and 2. through shared random effects to

account for the association between several outcomes. The empirical performance of the proposed univariate and multivariate CHBC methods is assessed through three extensive simulation studies varying the number of small areas and the sample size. Furthermore, we applied the proposed methods on two public health-related datasets; a novel coronavirus (COVID-19) data extracted from the John Hopkins University (JHU) COVID-19 repository and tobacco smoking dataset downloaded from the Youth Risk Behavior Surveillance System (BRFSS). The univariate CHBC method results were compared with the GLMM, which showed similar performance in both.

For illustration, we considered $u_i \sim \mathcal{N}(0, \sigma^2)$, $i = 1, \dots, m$, but random effects may come from any conjugate of exponential family distributions in HGLMs and can directly estimate model parameters using the proposed CHBC method. The h -likelihood approach avoids computationally expensive integration by taking partial derivatives of the logarithm of the joint density function, namely h -likelihood, which simplifies the differentiation. Moreover, the proposed h -likelihood approach is a computationally efficient method that provides reliably accurate results through a single algorithm.

The proposed h -likelihood method with bias correction of estimates provides consistently better estimates compared to the h -likelihood method without bias correction. The simulation results demonstrate that the univariate CHBC method performs well, providing reliable estimates even for the areas with small sample sizes, as shown in Figures 5.1 and 5.2. Even though the MHLEs of the variance parameter σ^2 provide better estimates with a large number of small areas; it also offers reliable estimates with small sample sizes and with a small number of areas with slightly underestimated values. Similarly, the MHLEs of fixed effects are reasonably accurate in any combination of m and n . Moreover, they are more precise with large sample sizes than small sample sizes and a few small areas. Overall, the proposed CHBC method results are consistent, the process of parameter estimation is less complicated, and sometimes slightly better in some cases compared with the related historical work done by the other researchers (Breslow & Clayton, 1993;

Shun & McCullagh, 1995). As we expected, the mean squared error of fixed effects estimates gets smaller when increasing the sample size as well as the number of small areas. Overall, the h -likelihood produces reasonable MHLEs. In some situations, there might be convergence issues with very small sample sizes with a small number of areas, especially with binary response data.

The mean squared errors for mixed logit and Poisson models obtained from MC simulation for the average random effects are displayed in Figure 5.3 for each combination of the number of small areas and sample sizes. In both scenarios, the MSEs are reliably low, and it shows that MSE decreases when the sample size increases. In this dissertation, in the univariate model, $\hat{\sigma}^2$ was obtained using partial derivatives assuming that it does not depend on $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{u}}$. Practically, it might be the case where σ^2 is a function of $\boldsymbol{\beta}$ and \mathbf{u} , so estimation of σ^2 expected to be more accurate based on the total derivative of h_A with respect to some σ^2 to avoid the indirect interdependencies between the estimators.

Furthermore, as stated above, the CHBC method can be applied to a generalized class of exponential family distributions, since it does not require the closed form of the log-likelihood function. However, the process is more straightforward to adopt for canonical GLM family distributions, since the score function and the Hessian matrix can easily be obtained using the score function of $\ell_{\mathbf{y}|\mathbf{u}}(\boldsymbol{\theta}; \mathbf{y})$, the weight matrix (\mathbf{W}), the first derivative ($\nabla_{\boldsymbol{\theta}}^1$), and the second derivative of the $\ell_{\mathbf{u}}(\nabla_{\boldsymbol{\theta}}^2)$, as described in section 3.1. Additionally, the real data analyses based on the COVID-19 data to illustrate the univariate Poisson model and the YRBSS data to demonstrate the mixed logit model indicate that the proposed CHBC method performs equally or reliably better compared to the widely used GLMM.

Moreover, the simulation results in section 5.3 show that the proposed multivariate joint model based on the CHBC method performs well. The MHLEs for fixed effects for \mathbf{y}_1 and \mathbf{y}_2 displayed in Figures 5.5 and 5.6, respectively, indicate the accurate results in most scenarios. The RMSE and the relative bias is small, and it improves as m and n increase. Additionally, the

distributions of estimated random effects for both outcomes lie close to the actual distribution. We also illustrated both joint modeling approaches using the COVID-19 data for cases and deaths for the multivariate case, and the YRBSS data for current-use and ever-use of E-Cigarettes for the shared random effects case. Based on the multivariate joint model on the COVID-19 information, it shows that the cases and deaths are positively associated with a correlation factor of 0.262. Hence, it is vital to model such outcomes jointly, accounting for the association among them, which leads to more precise model estimations. Furthermore, the joint model based on the shared random effects on the tobacco data set to evaluate the current-use and ever-use of prevalence shows that they are positively correlated. The estimated shared parameter $\hat{\gamma} = 2.97$ implies that the existence of a strong association between current-use and ever-use of E-Cigarettes. Finally, we conclude the proposed CHBC approach performs better in both univariate and multivariate models.

7.2 Limitations

One of the well-known limitations using iterative approximation is the convergence issue in specific scenarios, which might not be very often. However, this is hard to avoid in such an iterative approximation process (Nocedal & Wright, 2006). In some situations where the Hessian matrix is singular, it is probable to consider the modified Hessian based on the eigenvalue decomposition of the Hessian, which is numerically positive definite. Thus, the modified Hessian is invertible. However, this process is computationally expensive with large dimensions.

Additionally, we observed that the MHLEs of the variance parameters are underestimated by the CHBC method, while overestimated the GLMM. It has been challenging to obtain very accurate estimations for the unobserved random effects; however, the proposed approach provides small RMSE for the MHLE of σ^2 . Furthermore, we could not find a benchmark analysis to compare the results from the proposed multivariate joint model based on the CHBC method. However, the accuracy of the simulation results and the real data analysis eliminate this issue. Additionally, as in

most data analysis problems, it was challenging to find a perfect data set to illustrate the proposed CHBC method.

7.3 Future Work

In terms of future work, the proposed method has opportunities to extend in other applications in SAE. With this, we will also consider extending this methodology to a high dimensional case in SAE, where the number of small areas is sufficiently large. However, in this scenario, not all clusters contribute to the model; some areas might not be informative and predictive enough in obtaining a model. Hence, it is vital to consider dimensionality reduction to eliminate the model complexity by removing the unimportant clusters or small areas. So, the proposed CHBC method can be extended, introducing a penalty term to reduce the dimensionality of random effects.

Additionally, the joint modeling approach could grow in different ways, which are not considered yet. The joint modeling through shared random effects is considered only using one shared parameter in this work. However, it is open to considering multiple outcomes with different options for the shared random effect component, such as a function of a single parameter, additional parameters, or in terms of a piecewise function. Additionally, $\widehat{\sigma^2}$ can be obtained based on the total derivative h_A .

References

- (CDC), C. f. D. C. a. P. (2020). People with Certain Medical Conditions. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html#copd>. Retrieved 07/30/2020
<https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html#copd>
- Akaike, H. (1973). Information theory and the maximum likelihood principle in 2nd International Symposium on Information Theory (BN Petrov and F. Cs ä ki, eds.). *Akademiai Ki à do, Budapest*.
- Ambrose, B. K., Day, H. R., Rostron, B., Conway, K. P., Borek, N., Hyland, A., & Villanti, A. C. (2015). Flavored tobacco product use among US youth aged 12-17 years, 2013-2014. *Jama*, *314*(17), 1871-1873.
- Banerjee, T., & Nayak, A. (2020). US county level analysis to determine If social distancing slowed the spread of COVID-19. *Revista Panamericana de Salud Pública*, *44*, e90.
- Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, *70*(2), 343-365.
- Barrington-Trimis, J. L., Urman, R., Berhane, K., Unger, J. B., Cruz, T. B., Pentz, M. A., . . . McConnell, R. (2016). E-Cigarettes and Future Cigarette Use. *Pediatrics*. doi:10.1542/peds.2016-0379
- Battese, G. E., & Fuller, W. A. (1981). *Prediction of county crop areas using survey and satellite data*. Paper presented at the Proceedings of the section on survey research methods, American Statistical Association.
- Behavioral Risk Factor Surveillance System Survey Data (BRFSS)*. (2015).
- Benavent, R., & Morales, D. (2016). Multivariate Fay–Herriot models for small area estimation. *Computational Statistics & Data Analysis*, *94*, 372-390.
- Berkowitz, Z., Zhang, X., Richards, T. B., Peipins, L., Henley, S. J., & Holt, J. (2016a). Multilevel small-area estimation of multiple cigarette smoking status categories using the 2012 behavioral risk factor surveillance system. *Cancer Epidemiol Biomarkers Prev* *25*(10), 1402-1410.
- Berkowitz, Z., Zhang, X., Richards, T. B., Peipins, L., Henley, S. J., & Holt, J. (2016b). Multilevel small-area estimation of multiple cigarette smoking status categories using the 2012 behavioral risk factor surveillance system. *Cancer Epidemiology and Prevention Biomarkers*, *25*(10), 1402-1410.

- Berkowitz, Z., Zhang, X., Richards, T. B., Sabatino, S. A., Peipins, L. A., Holt, J., & White, M. C. (2019). Multilevel Regression for Small-Area Estimation of Mammography Use in the United States, 2014. *Cancer Epidemiol Biomarkers Prev*, 28(1), 32-40. doi:10.1158/1055-9965.EPI-18-0367
- Booth, J. G., & Hobert, J. P. (1998). Standard errors of prediction in generalized linear mixed models. *Journal of the American statistical association*, 93(441), 262-272.
- Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical association*, 88(421), 9-25.
- Brian King, T. P., Peter Mariolis. (2014). Best Practices for comprehensive tobacco control programs-2014.
- Burgard, J. P., Esteban, M. D., Morales, D., & Pérez, A. (2020). Small area estimation under a measurement error bivariate Fay–Herriot model. *Statistical Methods & Applications*, 1-30.
- Cameron, J. M., Howell, D. N., White, J. R., Andrenyak, D. M., Layton, M. E., & Roll, J. M. (2014). Variable and potentially fatal amounts of nicotine in e-cigarette nicotine solutions. *Tob Control*, 23(1), 77-78. doi:10.1136/tobaccocontrol-2012-050604
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*: Chapman and Hall/CRC.
- Casella, G., & Berger, R. L. (2002). *Statistical inference* (Vol. 2): Duxbury Pacific Grove, CA.
- Centers for Disease Control and Prevention. (2013). QuickStats: Number of Deaths from 10 Leading Causes—National Vital Statistics System, United States, 2010. . *Morbidity and Mortality Weekly Report*, http://www.cdc.gov/mmwr/preview/mmwrhtml/mm6325a3.htm?s_cid=mm6325a3_w, 62(08), 155.
- Cressie, N. (1991). *Small-area prediction of undercount using the general linear model*. Paper presented at the Proceedings of statistics symposium 90: measurement and improvement of data quality.
- Crouch, E. A., & Spiegelman, D. (1990). The Evaluation of Integrals of the form $\int_{-\infty}^{+\infty} f(t) \exp(-t^2) dt$: Application to Logistic-Normal Models. *Journal of the American statistical Association*, 85(410), 464-469.
- Dahab, M., van Zandvoort, K., Flasche, S., Warsame, A., Ratnayake, R., Favas, C., . . . Checchi, F. (2020). COVID-19 control in low-income settings and displaced populations: what can realistically be done? *Conflict and Health*, 14(1), 1-6.
- Dai, H. (2018). Single, Dual, and Poly Use of Flavored Tobacco Products among Youth in the United States: 2014. *Preventing Chronic Disease*. *In press*.

- Dai, H. (2019). Changes in Flavored Tobacco Product Use Among Current Youth Tobacco Users in the United States, 2014-2017. *JAMA pediatrics*.
- Dai, H. (2019). Changes in Flavored Tobacco Product Use among Current Youth Tobacco Users in the United States: 2014 – 2017. *JAMA Pediatrics*.
- Dai, H., Catley, D., Richter, K. P., Goggin, K., & Ellerbeck, E. F. (2018). Electronic Cigarettes and Future Marijuana Use: A Longitudinal Study. *Pediatrics*, *141*(5), e20173787.
- Dai, H., Catley, D., Richter, K. P., Goggin, K., & Ellerbeck, E. F. (2018). Electronic Cigarettes and Future Marijuana Use: A Longitudinal Study. *Pediatrics*. doi:10.1542/peds.2017-3787
- Dai, H., & Hao, J. (2016). Exposure to advertisements and susceptibility to electronic cigarette use among youth. *Journal of Adolescent Health*, *59*(6), 620-626.
- Dai, H., & Hao, J. (2018). Flavored Tobacco Use Among US Adults by Age Group: 2013–2014. *Substance use & misuse*, 1-9.
- Daily Travel during the COVID-19 Public Health Emergency. (2020). Retrieved from <https://www.bts.gov/browse-statistical-products-and-data/trips-distance/explore-us-mobility-during-covid-19-pandemic>
- Daniels, M. J., & Gatsonis, C. (1999). Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of the American statistical association*, *94*(445), 29-42.
- Datta, G. S., & Ghosh, M. (1991). Bayesian prediction in linear models: Applications to small area estimation. *The annals of statistics*, 1748-1770.
- Dempster, A., & Tomberlin, T. (1980). *The analysis of census undercount from a postenumeration survey*. Paper presented at the Proceedings of the Conference on Census Undercount.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.
- Dick, P. (1995). Modelling net undercoverage in the 1991 Canadian Census. *Survey Methodology*, *21*(1), 45-54.
- Dinakar, C., & O'Connor, G. T. (2016). The Health Effects of Electronic Cigarettes. *N Engl J Med*, *375*(14), 1372-1381. doi:10.1056/NEJMra1502466
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, *20*(5), 533-534.
- Duke, J. C., Lee, Y. O., Kim, A. E., Watson, K. A., Arnold, K. Y., Nonnemaker, J. M., & Porter, L. (2014). Exposure to electronic cigarette television advertisements among youth and young adults. *Pediatrics*, *134*(1), e29-e36.

- Dwyer-Lindgren, L., Mokdad, A. H., Srebotnjak, T., Flaxman, A. D., Hansen, G. M., & Murray, C. J. (2014). Cigarette smoking prevalence in US counties: 1996-2012. *Population health metrics*, 12(1), 5.
- E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General*. (2016). Retrieved from US Department of Health and Human Services: https://e-cigarettes.surgeongeneral.gov/documents/2016_SGR_Full_Report_non-508.pdf
- Effects of COVID-19 Pandemic on Employment and Unemployment Statistics. (2020). *U.S. Bureau of Labor Statistics*. Retrieved from <https://www.bls.gov/covid19/effects-of-covid-19-pandemic-on-employment-and-unemployment-statistics.htm>
- Farrell, P. J., MacGibbon, B., & Tomberlin, T. J. (1997). Empirical Bayes estimators of small area proportions in multistage designs. *Statistica Sinica*, 1065-1083.
- Fay III, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American statistical association*, 74(366a), 269-277.
- Fay III, R. E., & Herriot, R. A. J. J. o. t. A. S. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. 74(366a), 269-277.
- Fraser, G. E., & Stram, D. O. (2012). Regression calibration when foods (measured with error) are the variables of interest: markedly non-Gaussian data with many zeroes. *American journal of epidemiology*, 175(4), 325-331.
- Freedman, L. S., Midthune, D., Carroll, R. J., & Kipnis, V. (2008). A comparison of regression calibration, moment reconstruction and imputation for adjusting for covariate measurement error in regression. *Statistics in Medicine*, 27(25), 5195-5216.
- Fry, R. (2016). Millennials overtake Baby Boomers as America's largest generation. *Pew Research Center*, 25.
- Fuller, W. A., & Battese, G. E. (1973). Transformations for estimation of linear models with nested-error structure. *Journal of the American statistical association*, 68(343), 626-632.
- Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical Association*, 85(410), 398-409.
- Ghosh, M., Natarajan, K., Stroud, T., & Carlin, B. P. (1998). Generalized linear models for small-area estimation. *Journal of the American statistical association*, 93(441), 273-282.
- Ghosh, M., & Rao, J. (1994). Small area estimation: an appraisal. *Statistical Science*, 9(1), 55-76.
- Glasser, A., Abudayyeh, H., Cantrell, J., & Niaura, R. (2019). Patterns of e-cigarette use among youth and young adults: review of the impact of e-cigarettes on cigarette smoking. *Nicotine and Tobacco Research*, 21(10), 1320-1330.

- Global research on coronavirus disease (COVID-19). Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>
- Gómez-Rubio, V., Best, N., Richardson, S., Li, G., & Clarke, P. (2010). Bayesian statistics small area estimation.
- Goniewicz, M. L., Hajek, P., & McRobbie, H. (2014). Nicotine content of electronic cigarettes, its release in vapour and its consistency across batches: regulatory implications. *Addiction*, *109*(3), 500-507. doi:10.1111/add.12410
- González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D., & Santamaría, L. (2008). Analytic and bootstrap approximations of prediction errors under a multivariate Fay–Herriot model. *Computational Statistics & Data Analysis*, *52*(12), 5242-5252.
- Gonzalez, M. E., & Hoza, C. (1978). Small-area estimation with application to unemployment and housing estimates. *Journal of the American statistical association*, *73*(361), 7-15.
- Gueorguieva, R. (2001). A multivariate generalized linear mixed model for joint modelling of clustered outcomes in the exponential family. *Statistical Modelling*, *1*(3), 177-193.
- Ha, I. D., Lee, Y., & Song, J. k. (2001). Hierarchical likelihood approach for frailty models. *Biometrika*, *88*(1), 233-233.
- Ha, I. D., Noh, M., & Lee, Y. (2017). H-likelihood approach for joint modeling of longitudinal outcomes and time-to-event data. *Biometrical Journal*, *59*(6), 1122-1143.
- Hartley, H. O., & Rao, J. N. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*, *54*(1-2), 93-108.
- Harville, D. A. (1991). [That BLUP is a Good Thing: The Estimation of Random Effects]: Comment. *Statistical Science*, *6*(1), 35-39.
- Hausman, J. A., Newey, W. K., & Powell, J. L. (1995). Nonlinear errors in variables estimation of some Engel curves. *Journal of Econometrics*, *65*(1), 205-233.
- Henderson, C. R. (1950). *Estimation of genetic parameters*. Paper presented at the Biometrics.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics*, *9*(2), 226-252.
- Henderson, C. R., Kempthorne, O., Searle, S. R., & Von Krosigk, C. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, *15*(2), 192-218.
- Hobza, T., & Morales, D. (2016). Empirical best prediction under unit-level logit mixed models. *Journal of official statistics*, *32*(3), 661-692.

- How COVID-19 Spreads. (06/16/2020). Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html>
- Hu, S. (2016). *2015 National Youth Tobacco Survey: Methodology Report*. Retrieved from https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/zip_files/2015/2015-nyts-methodology-report.zip
- Huang, S. Y. (2005). Regression calibration using response variables in linear models. *Statistica Sinica*, 685-696.
- Huang, X., & Wolfe, R. A. (2002). A frailty model for informative censoring. *Biometrics*, 58(3), 510-520.
- IBIS. Complete Health Indicator Report of Electronic Cigarettes / Vape Products. Retrieved from https://ibis.health.utah.gov/ibisph-view/indicator/complete_profile/ECig.html
- Institute, N. C. (1998). *Smoking and Tobacco Control Monograph 9: Cigars: Health Effects and Trends*. Bethesda, MD. Available at <https://cancercontrol.cancer.gov/BRP/tcrb/monographs/9/index.html>. Accessed on July 16, 2017.
- Isaki, C., Tsay, J. H., & Fuller, W. A. (2000). Estimation of census adjustment factors. *Survey Methodology*, 26(1), 31-42.
- Jamal, A., Gentzke, A., Hu, S. S., Cullen, K. A., Apelberg, B. J., Homa, D. M., & King, B. A. (2017). Tobacco Use Among Middle and High School Students - United States, 2011-2016. *MMWR Morb Mortal Wkly Rep*, 66(23), 597-603. doi:10.15585/mmwr.mm6623a1
- Jensen, R. P., Luo, W., Pankow, J. F., Strongin, R. M., & Peyton, D. H. (2015). Hidden formaldehyde in e-cigarette aerosols. *N Engl J Med*, 372(4), 392-394. doi:10.1056/NEJMc1413069
- Jiang, J., & Lahiri, P. (2006). Mixed model prediction and small area estimation. *Test*, 15(1), 1.
- Jiang, J., Lahiri, P., & Wan, S.-M. (2002). A unified jackknife theory for empirical best prediction with M-estimation. *The annals of statistics*, 30(6), 1782-1810.
- Kochhar, R. (2020). Unemployment rose higher in three months of COVID-19 than it did in two years of the Great Recession. *Pew Research Center*. Retrieved from <https://www.pewresearch.org/fact-tank/2020/06/11/unemployment-rose-higher-in-three-months-of-covid-19-than-it-did-in-two-years-of-the-great-recession/>
- Kostygina, G., Glantz, S. A., & Ling, P. M. (2016). Tobacco industry use of flavours to recruit new users of little cigars and cigarillos. *Tob Control*, 25(1), 66-74. doi:10.1136/tobaccocontrol-2014-051830

- Laplace, P. S. (1986). Memoir on the probability of the causes of events. *Statistical Science*, 1(3), 364-378.
- Lee, Y., Alam, M. M., Noh, M., Rönnegård, L., & Skarin, A. (2016). Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts. *Ecology and evolution*, 6(19), 7047-7056.
- Lee, Y., Jang, M., & Lee, W. (2011). Prediction interval for disease mapping using hierarchical likelihood. *Computational Statistics*, 26(1), 159-179.
- Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B*, 58(4), 619-656.
- Lee, Y., & Nelder, J. A. (1996). Hierarchical generalized linear models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(4), 619-656.
- Lee, Y., & Nelder, J. A. (2005). Likelihood for random-effect models. *Sort*, 29(2), 141-164.
- Lee, Y., Nelder, J. A., & Pawitan, Y. (2006). *Generalized linear models with random effects: unified analysis via H-likelihood*: Chapman and Hall/CRC.
- Lee, Y., Nelder, J. A., & Pawitan, Y. (2018). *Generalized linear models with random effects: unified analysis via H-likelihood* (Vol. 153): CRC Press.
- Lee, Y., Ronnegard, L., & Noh, M. (2017). *Data analysis using hierarchical generalized linear models with R*: Chapman and Hall/CRC.
- Lehmann, E. L., & Romano, J. P. (2006). *Testing statistical hypotheses*: Springer Science & Business Media.
- Leventhal, A. M., Strong, D. R., Kirkpatrick, M. G., Unger, J. B., Sussman, S., Riggs, N. R., . . . Audrain-McGovern, J. (2015). Association of Electronic Cigarette Use With Initiation of Combustible Tobacco Product Smoking in Early Adolescence. *JAMA*, 314(7), 700-707. doi:10.1001/jama.2015.8950 2428954 [pii]
- Levy, V. (January 2020). 2020 Travel Trends, Boomers Have Big Travel Plans in 2020. *AARP Research*. Retrieved from https://www.aarp.org/content/dam/aarp/research/surveys_statistics/life-leisure/2019/2020-travel-trends.doi.10.26419-2Fres.00359.001.pdf
- Liu, Q., & Pierce, D. A. (1993). Heterogeneity in Mantel-Haenszel-type models. *Biometrika*, 80(3), 543-556.
- MacDonald, G., Starr, G., Schooley, M., Yee, S. L., & Klimowski, K. (2001). Introduction to program evaluation for comprehensive tobacco control programs.
- MacGibbon, B., & Tomberlin, T. J. (1987). *Small area estimates of proportions via empirical Bayes techniques*: Faculty of Commerce and Administration, Concordia University.

- Martuzzi, M., & Elliott, P. (1996). Empirical Bayes estimation of small area prevalence of non-rare conditions. *Statistics in Medicine*, 15(17-18), 1867-1873.
- Mauro, F., Monleon, V. J., Temesgen, H., & Ford, K. R. (2017). Analysis of area level and unit level models for small area estimation in forest inventories assisted with LiDAR auxiliary information. *PloS one*, 12(12), e0189401.
- McCullagh, P. (2018). *Generalized linear models*: Routledge.
- McCulloch, C. E., & Searle, S. R. (2004). *Generalized, linear, and mixed models*: John Wiley & Sons.
- Molenberghs, G., Verbeke, G., Demétrio, C. G., & Vieira, A. M. (2010). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, 25(3), 325-347.
- Molina, I., & Marhuenda, Y. J. T. R. J. (2015). sae: An R package for small area estimation. 7(1), 81-98.
- Morris, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *Journal of the American statistical association*, 78(381), 47-55.
- Namias, V. (1986). A simple derivation of Stirling's asymptotic series. *The American Mathematical Monthly*, 93(1), 25-29.
- National Cancer Institute. (2017). A Socioecological Approach to Addressing Tobacco-Related Health Disparities. National Cancer Institute Tobacco Control Monograph 22. Available at <https://cancercontrol.cancer.gov/brp/tcrb/monographs/22/monograph22.html>. Accessed on October 20, 2017.
- Nocedal, J., & Wright, S. (2006). *Numerical optimization*: Springer Science & Business Media.
- Noh, M., & Lee, Y. (2007). REML estimation for binary data in GLMMs. *Journal of Multivariate Analysis*, 98(5), 896-915. doi:10.1016/j.jmva.2006.11.009
- Patterson, H. D., & Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika*, 58(3), 545-554.
- Petrucci, A., & Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment. *Journal of agricultural, biological, and environmental statistics*, 11(2), 169.
- Pfeffermann, D. (2013). New important developments in small area estimation. *Statistical Science*, 28(1), 40-68.
- Prasad, N., & Rao, J. (1986). *On the estimation of mean square error of small area predictors*. Paper presented at the Proceedings of the Section on Survey Research Methods, American Statistical Association.

- Prasad, N. N., & Rao, J. N. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association*, 85(409), 163-171.
- Prevention, C. f. D. C. a. Cigarette Smoking and Tobacco Use Among People of Low Socioeconomic Status. Retrieved from <https://www.cdc.gov/tobacco/disparities/low-ses/index.htm>
- Primack, B. A., Soneji, S., Stoolmiller, M., Fine, M. J., & Sargent, J. D. (2015). Progression to Traditional Cigarette Smoking After Electronic Cigarette Use Among US Adolescents and Young Adults. *JAMA Pediatr*, 169(11), 1018-1023. doi:10.1001/jamapediatrics.2015.1742
- Provisional COVID-19 Death Counts by Sex, Age, and State. (2020, 08/19/2020). Retrieved from <https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku>
- Rahman, A., & Harding, A. (2016). *Small area estimation and microsimulation modeling*: CRC Press.
- Rao, J. N., & Molina, I. (2015). *Small area estimation*: John Wiley & Sons.
- Rao, J. N., & Yu, M. (1994). Small-area estimation by combining time-series and cross-sectional data. *Canadian Journal of Statistics*, 22(4), 511-528.
- Razzaghi, H., Wang, Y., Lu, H., Marshall, K. E., Dowling, N. F., Paz-Bailey, G., . . . Greenlund, K. J. (2020). Estimated County-Level Prevalence of Selected Underlying Medical Conditions Associated with Increased Risk for Severe COVID-19 Illness—United States, 2018. *Morbidity and Mortality Weekly Report*, 69(29), 945.
- Reducing Vaping Among Youth and Young Adults* (SAMHSA Publication No. PEP20-06-01-003). (2020). Retrieved from Rockville, MD: National Mental Health and Substance Use Policy Laboratory, Substance Abuse and Mental Health Services Administration: https://store.samhsa.gov/sites/default/files/SAMHSA_Digital_Download/PEP20-06-01-003_508.pdf
- Rigotti, N. A., & Kalkhoran, S. (2017). Reducing Health Disparities by Tackling Tobacco Use. *Journal of General Internal Medicine*, 32, 961-962. doi:10.1007/s11606-017-4098-7
- Rigotti, N. A., & Kalkhoran, S. (2017). Reducing Health Disparities by Tackling Tobacco Use. *J Gen Intern Med*, 32(9), 961-962. doi:10.1007/s11606-017-4098-7
- Roberts, A. W., Ogunwole, S. U., Blakeslee, L., & Rabe, M. A. (2018). *The population 65 years and older in the United States: 2016*: US Department of Commerce, Economics and Statistics Administration, US . . .
- Saei, A., & Chambers, R. (2003). Small area estimation under linear and generalized linear mixed models with time and area effects.

- Schabenberger, O. (2005). Introducing the GLIMMIX procedure for generalized linear mixed models. *SUGI 30 Proceedings*, 196.
- Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Searle, S. R., Casella, G., & McCulloch, C. E. (2009). *Variance components* (Vol. 391): John Wiley & Sons.
- Shun, Z., & McCullagh, P. (1995). Laplace approximation of high dimensional integrals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(4), 749-760.
- Soneji, S., Barrington-Trimis, J. L., Wills, T. A., Leventhal, A. M., Unger, J. B., Gibson, L. A., . . . Sargent, J. D. (2017). Association Between Initial Use of e-Cigarettes and Subsequent Cigarette Smoking Among Adolescents and Young Adults: A Systematic Review and Meta-analysis. *JAMA Pediatr*, 171(8), 788-797. doi:10.1001/jamapediatrics.2017.1488
- Spiegelman, D., Logan, R., & Grove, D. (2011). Regression calibration with heteroscedastic error variance. *The international journal of biostatistics*, 7(1).
- Spiegelman, D., McDermott, A., & Rosner, B. (1997). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *The American journal of clinical nutrition*, 65(4), 1179S-1186S.
- Stroud, T. (1994). Bayesian analysis of binary survey data. *Canadian Journal of Statistics*, 22(1), 33-45.
- Taioli, E., & Wynder, E. L. (1991). Effect of the age at which smoking begins on frequency of smoking in adulthood. *N Engl J Med*, 325(13), 968-969. doi:10.1056/NEJM199109263251318
- Team, R. (2015). RStudio: integrated development for R. PBC, Boston, MA Retrieved from <https://www.rstudio.com/>
- Torabi, M., & Rao, J. (2014). On small area estimation under a sub-area level model. *Journal of Multivariate Analysis*, 127, 36-55.
- Tsiatis, A., & Davidian, M. (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, 14, 793-818.
- U.S. Department of Health and Human Services. (2012). Preventing Tobacco Use Among Youth and Young Adults: A Report of the Surgeon General. *Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health*.
- U.S. Department of Health and Human Services. (2014). The Health Consequences of Smoking—50 Years of Progress: A Report of the Surgeon General. Atlanta: U.S. Department of Health

- and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- U.S. Department of Health and Human Services. (2016). E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General. *Atlanta: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.*
- Ubaidillah, A., Notodiputro, K. A., Kurnia, A., & Mangku, I. W. (2019). Multivariate Fay-Herriot models for small area estimation with application to household consumption per capita expenditure in Indonesia. *Journal of Applied Statistics, 46*(15), 2845-2861.
- Vaida, F., Meng, X. L., & Xu, R. (2004). Mixed effects models and the EM algorithm. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, 253-264.
- Villanti, A. C., Richardson, A., Vallone, D. M., & Rath, J. M. (2013). Flavored tobacco product use among US young adults. *American journal of preventive medicine, 44*(4), 388-391.
- Walker, K. (2018). Tidycensus: Load us census boundary and attribute data as 'tidyverse' and 'sf'-ready data frames. In: R package version 0.4.
- Walker, P., Whittaker, C., Watson, O., Baguelin, M., Ainslie, K., Bhatia, S., . . . Cattarino, L. (2020). *Report 12: The global impact of COVID-19 and strategies for mitigation and suppression*. Retrieved from
- Wang, C., Hsu, L., Feng, Z., & Prentice, R. L. (1997). Regression calibration in failure time regression. *Biometrics, 131*-145.
- Wang, T. W., Gentzke, A., Sharapova, S., Cullen, K. A., Ambrose, B. K., & Jamal, A. (2018). Tobacco Product Use Among Middle and High School Students—United States, 2011–2017. *Morbidity and Mortality Weekly Report, 67*(22), 629.
- Whiteside, D. T. (1967). *The mathematical papers of Isaac Newton* (Vol. 1982): Cambridge University Press Cambridge.
- Wills, T. A., Knight, R., Sargent, J. D., Gibbons, F. X., Pagano, I., & Williams, R. J. (2016). Longitudinal study of e-cigarette use and onset of cigarette smoking among high school students in Hawaii. *Tob Control*. doi:10.1136/tobaccocontrol-2015-052705
- Wills, T. A., & Soneji, S. S. (2018). Individual-Level and Ecological Studies. *J Adolesc Health, 62*(5), 507-508. doi:10.1016/j.jadohealth.2018.03.002
- Yasui, Y., Liu, H., Benach, J., & Winget, M. (2000). An empirical evaluation of various priors in the empirical Bayes estimation of small area disease risks. *Statistics in Medicine, 19*(17-18), 2409-2420. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/10960862>

- You, Y. (2008). Small area estimation using area level models with model checking and applications. *J Proceedings of Survey Methods Section, Statistical Society of Canada*.
- You, Y., & Chapman, B. (2006). Small area estimation using area level models and estimated sampling variances. *Survey Methodology*, 32(1), 97.
- Yun, S., & Lee, Y. (2004). Comparison of hierarchical and marginal likelihood estimators for binary outcomes. *Computational Statistics & Data Analysis*, 45(3), 639-650.
- Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American journal of epidemiology*, 179(8), 1025-1033.

Appendix

Appendix A. Data Analysis Based on GLMM Using NYTS Tobacco Smoking Data

Table A.1: Ten US states with the highest and lowest estimated prevalence of current-use and ever-use of E-Cigarettes based on GLMM.

Current-use of E-Cigarettes							
Highest rankings				Lowest rankings			
State	County	True	Estimated	State	County	True	Estimated
Oklahoma	Oklahoma	44.31	57.89	Louisiana	Orleans	0.03	0.89
Pennsylvania	Lawrence	24.56	36.11	New York	Kings	0.65	1.47
Texas	Lubbock	16.97	27.08	Minnesota	Dakota	2.19	1.52
Hawaii	Hawaii	32.09	26.76	Virginia	Dickenson	1.6	1.59
Massachusetts	Bristol	18.92	23.2	New Jersey	Hudson	1.97	1.67
Texas	El Paso	22.03	23.17	Ohio	Williams	3.33	1.75
Wisconsin	Washington	18.83	21.36	Utah	Davis	4.95	2.15
Ohio	Jefferson	19.96	19.48	New York	Queens	5.24	2.42
Mississippi	Lincoln	18.66	18.88	Florida	Osceola	2.27	2.46
Illinois	Lake	18.16	18.67	Ohio	Cuyahoga	3.36	2.54
Ever-use of E-Cigarettes							
Highest rankings				Lowest rankings			
State	County	True	Estimated	State	County	True	Estimated
Oklahoma	Oklahoma	72.97	63.46	Pennsylvania	Cumberland	2.6	6.51
Hawaii	Hawaii	43.66	53.40	Utah	Davis	4.26	8.59
Texas	El Paso	42.20	42.87	Wisconsin	Walworth	4.29	7.98
Ohio	Jefferson	40.00	40.29	Maryland	Baltimore	4.52	8.92
New Jersey	Passaic	29.90	38.93	Wisconsin	Trempealeau	4.62	8.76
New York	Suffolk	33.33	38.00	Michigan	Eaton	5.75	6.52
Florida	Brevard	26.50	36.96	Louisiana	Orleans	6.31	12.48
Oklahoma	Custer	42.19	36.66	Florida	Osceola	7.72	7.33
Arkansas	Washington	41.42	36.08	New Jersey	Hudson	7.98	7.7
Ohio	Butler	20.41	35.55	Florida	St. Lucie	8.4	12.73

Appendix B. Benchmark Analysis Using Binomial HGLM and Poisson HGLM

Table B.1: MHLEs of variance parameter from CHBC and GLMM for mixed logit model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).

m	n	HMLE	RMSE	Bias	MLE	RMSE	Bias
5	5	0.003	0.097	-0.974	0.323	0.050	2.235
	10	0.002	0.099	-0.979	0.319	0.050	2.190
	30	0.002	0.099	-0.983	0.304	0.047	2.043
	100	0.002	0.098	-0.975	0.249	0.045	1.488
	500	0.016	0.086	-0.842	0.203	0.050	1.031
10	5	0.003	0.097	-0.974	0.277	0.032	1.773
	10	0.002	0.098	-0.984	0.276	0.032	1.761
	30	0.001	0.099	-0.986	0.276	0.034	1.757
	100	0.003	0.098	-0.969	0.243	0.035	1.429
	500	0.054	0.049	-0.458	0.208	0.041	1.084
20	5	0.003	0.097	-0.974	0.272	0.030	1.716
	10	0.002	0.098	-0.984	0.266	0.028	1.664
	30	0.001	0.099	-0.987	0.251	0.025	1.515
	100	0.001	0.099	-0.986	0.226	0.023	1.257
	500	0.052	0.050	-0.481	0.199	0.027	0.993
30	5	0.003	0.097	-0.974	0.238	0.019	1.383
	10	0.002	0.098	-0.984	0.237	0.019	1.374
	30	0.001	0.099	-0.986	0.233	0.019	1.329
	100	0.002	0.098	-0.981	0.221	0.020	1.212
	500	0.067	0.035	-0.330	0.195	0.023	0.952

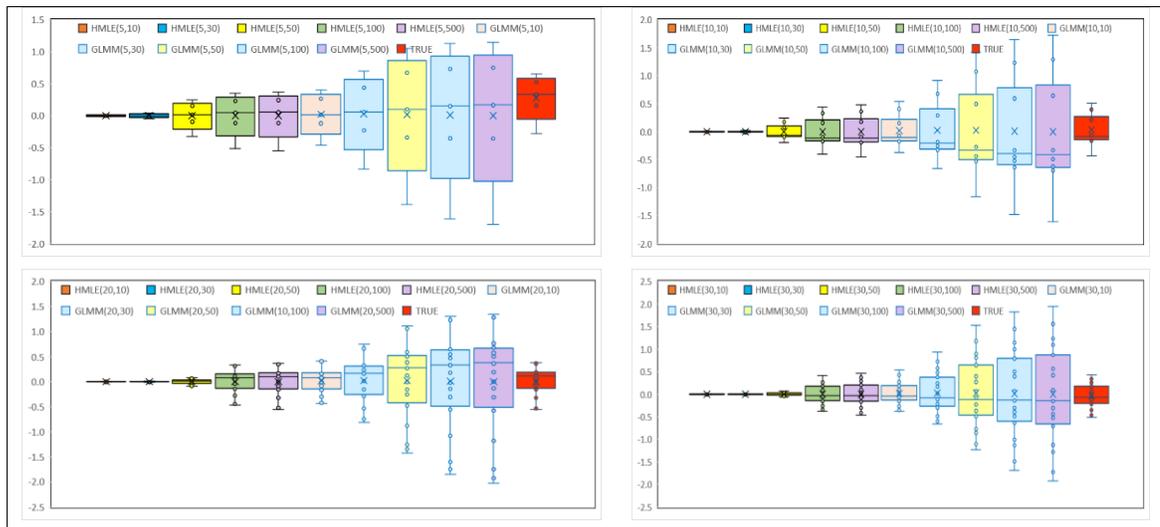
Note: True value of variance is 0.1 for each combination of m and n .

Table B.2: MHLEs of variance parameter from CHBC and GLMM for Poisson mixed model ($m = 5, 10, 20, 30, n = 10, 30, 50, 100, \text{ and } 500$).

m	n	HMLE	RMSE	Bias	MLE	RMSE	Bias
5	5	0.0039	0.0097	-0.9606	0.323	0.050	2.235
	10	0.0096	0.0093	-0.9035	0.319	0.050	2.190
	30	0.0500	0.0042	-0.4997	0.304	0.047	2.043
	100	0.0892	0.0007	-0.1079	0.249	0.045	1.488
	500	0.1021	0.0001	0.0211	0.203	0.050	1.031
10	5	0.0017	0.0097	-0.9827	0.277	0.032	1.773
	10	0.0038	0.0095	-0.9619	0.276	0.032	1.761
	30	0.0285	0.0059	-0.7152	0.276	0.034	1.757
	100	0.0630	0.0016	-0.3700	0.243	0.035	1.429
	500	0.0743	0.0007	-0.2572	0.208	0.041	1.084
20	5	0.0014	0.0097	-0.9859	0.272	0.030	1.716
	10	0.0011	0.0098	-0.9892	0.266	0.028	1.664
	30	0.0074	0.0087	-0.9263	0.251	0.025	1.515
	100	0.0526	0.0024	-0.4741	0.226	0.023	1.257
	500	0.0701	0.0009	-0.2991	0.199	0.027	0.993
30	5	0.0015	0.0097	-0.9853	0.238	0.019	1.383
	10	0.0011	0.0098	-0.9888	0.237	0.019	1.374
	30	0.0043	0.0092	-0.9566	0.233	0.019	1.329
	100	0.0406	0.0036	-0.5942	0.221	0.020	1.212
	500	0.0537	0.0022	-0.4629	0.195	0.023	0.952

Note: True value of variance is 0.1 for each combination of m and n .

Figure B.1: Distribution of estimated random effects in Poisson model¹.



¹ HMLE (m, n), GLMM (m, n): the boxplot for hierarchical MLE and MLE of estimated random effects for m small areas and n observations for each small area based on CHBC with bias correction method and GLMM respectively.

Appendix C. Simulation Results for the Multivariate Joint Model Using Count Data

Table C.1: Estimates of variance parameters of random effects.

Parameter	Sample size (m, n)	HMLE	Abs Error	Bias	Sample size (m, n)	HMLE	Abs Error	Bias
σ_{11}	(10, 10)	1.009	0.291	-0.224	(30, 10)	1.121	0.179	-0.137
σ_{12}		0.209	0.291	-0.582		0.321	0.179	-0.357
σ_{22}		1.209	0.291	-0.194		1.321	0.179	-0.119
σ_{11}	(10, 30)	1.386	0.086	0.066	(30, 30)	1.058	0.242	-0.186
σ_{12}		0.586	0.086	0.172		0.258	0.242	-0.484
σ_{22}		1.586	0.086	0.057		1.258	0.242	-0.161
σ_{11}	(10, 100)	1.080	0.220	-0.169	(30, 100)	1.059	0.241	-0.185
σ_{12}		0.280	0.220	-0.439		0.259	0.241	-0.482
σ_{22}		1.280	0.220	-0.146		1.259	0.241	-0.161
σ_{11}	(10, 500)	1.136	0.164	-0.126	(50, 10)	1.017	0.283	-0.218
σ_{12}		0.336	0.164	-0.327		0.217	0.283	-0.566
σ_{22}		1.336	0.164	-0.109		1.217	0.283	-0.189
σ_{11}	(20, 10)	1.974	0.674	0.519	(50, 30)	1.062	0.238	-0.183
σ_{12}		1.174	0.674	1.349		0.262	0.238	-0.477
σ_{22}		2.174	0.674	0.450		1.262	0.238	-0.159
σ_{11}	(20, 100)	1.726	0.426	0.328	(50, 100)	1.000	0.300	-0.230
σ_{12}		0.926	0.426	0.853		0.200	0.300	-0.599
σ_{22}		1.926	0.426	0.284		1.200	0.300	-0.200
σ_{11}	(20, 500)	0.882	0.418	-0.322	(50, 500)	1.000	0.300	-0.230
σ_{12}		0.082	0.418	-0.836		0.200	0.300	-0.599
σ_{22}		1.082	0.418	-0.279		1.200	0.300	-0.200

Note: True values of σ_{11} , σ_{12} , and σ_{22} are 1.3, 0.5, and 1.5.

Appendix D. Benchmark Analysis for Mixed Logistic Model Using YRBSS Data

Table D.1: Summary statistics of current-use and ever-use of E-Cigarettes, current-smoke, and ever-smoke by gender, race, and age.

Variable	# of Respondents	Prevalence (%)	95% CI		# of Respondents	Prevalence (%)	95% CI	
			Lower	Upper			Lower	Upper
	Ecig_Current				Ecig_Ever			
Total	27955	19.16	17.80	20.53	29194	43.62	41.55	45.69
Gender								
Male	13737	21.28	19.61	22.95	14332	45.52	43.45	47.58
Female	14218	17.02	15.57	18.47	14862	41.70	38.97	44.43
Race	27542	19.17	17.79	20.54	28783	43.69	41.60	45.77
White	12235	20.82	18.82	22.83	12831	42.52	39.48	45.57
African American	4080	13.72	11.47	15.97	4190	39.29	36.26	42.32
Hispanic	8112	19.72	17.80	21.64	8521	50.33	48.02	52.65
Others	3115	16.29	13.45	19.14	3241	40.66	37.22	44.09
Age	28033	19.17	17.82	20.53	29264	43.61	41.54	45.67
12yrs or younger	77	57.69	39.67	75.72	63	81.91	66.99	96.83
13 yrs old	34	19.00	0.92	37.07	35	34.20	11.60	56.81
14 yrs old	3367	12.05	10.22	13.87	3491	31.23	28.21	34.25
15 yrs old	6877	16.23	14.65	17.81	7188	37.81	35.19	40.42
16 yrs old	7144	18.72	17.06	20.38	7483	44.78	42.02	47.55
17 yrs old	6902	21.67	19.73	23.61	7224	48.75	46.81	50.70
18 yrs or older	3932	25.92	23.28	28.56	3780	52.22	48.97	55.47
	Current-smoke				Ever-smoke			
Total	29176	9.83	8.70	10.91	25523	30.67	28.43	32.90
Gender								
Male	14275	10.85	9.78	11.92	12501	32.35	30.26	34.43
Female	14901	8.76	7.46	10.06	13022	28.99	26.16	31.81
Race	28754	9.79	8.68	10.90	25187	30.69	28.44	32.94
White	12815	11.75	10.08	13.43	3525	31.37	28.10	34.63
African American	4281	5.45	4.31	6.60	3540	25.97	22.59	29.36
Hispanic	8411	8.15	7.16	9.14	8130	32.54	30.12	34.96
Others	3247	8.48	6.76	10.20	2918	28.86	25.13	32.59
Age	29260	9.81	8.71	10.92	25585	30.66	28.43	32.89
12yrs or younger	78	55.86	38.36	73.36	42	76.21	57.29	95.14
13 yrs old	32	20.41	1.85	38.98	17	19.03	0.00	40.28
14 yrs old	3529	4.55	3.55	5.56	2640	19.59	17.27	21.91
15 yrs old	7190	6.93	5.73	8.12	6210	23.95	21.71	26.19
16 yrs old	7473	9.31	7.93	10.70	6563	30.30	27.34	33.27
17 yrs old	7200	11.99	10.53	13.45	6446	36.47	33.97	38.97
18 yrs or older	3758	15.54	13.52	17.57	3667	41.26	37.32	45.20

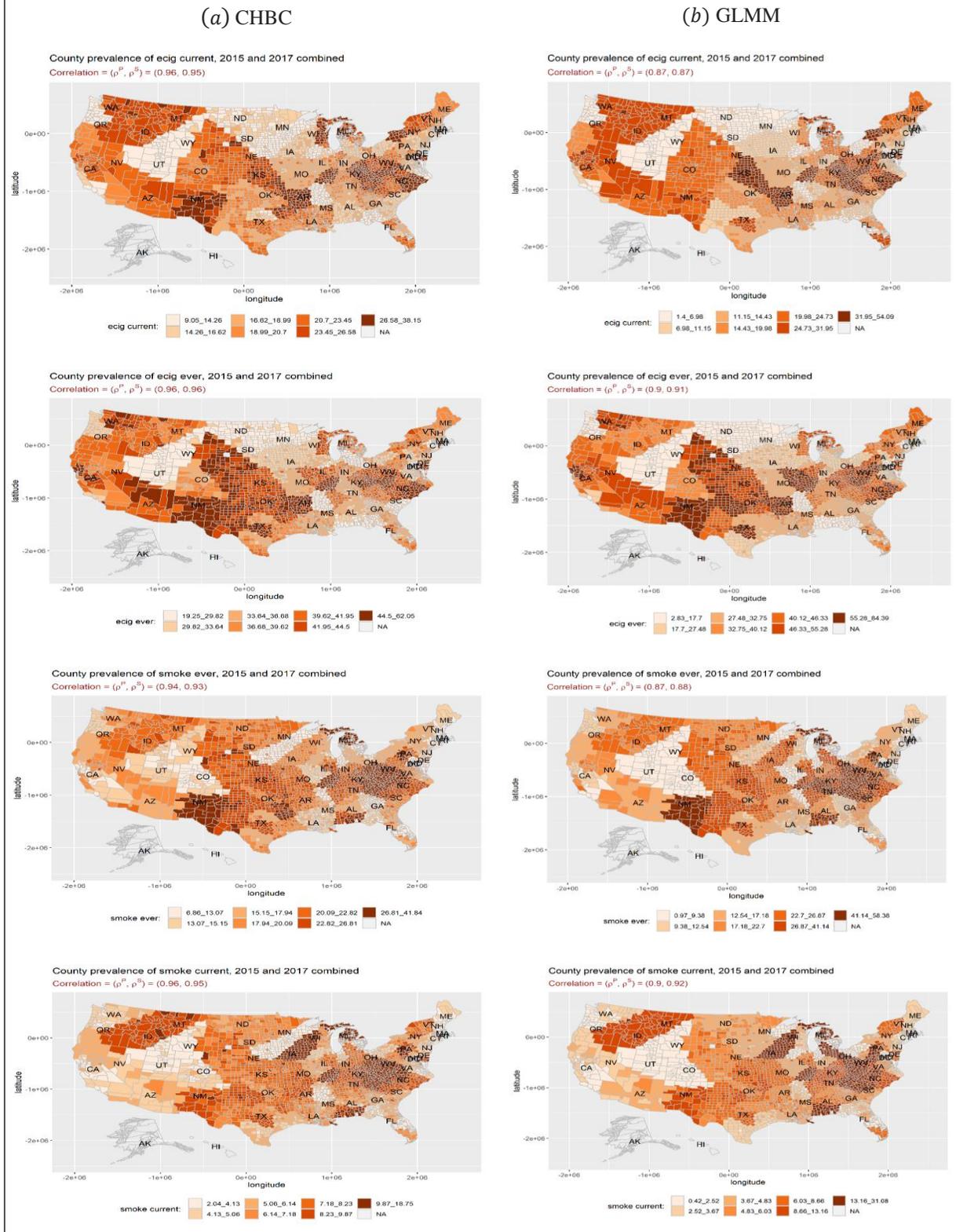
Table D.2: Ten US states with the highest and lowest estimated prevalence of current-use and ever-use of E-Cigarettes based on CHBC and GLMM.

Current-use of E-Cigarettes							
Highest rankings				Lowest rankings			
CHBC		GLMM		CHBC		GLMM	
North Carolina	29.22	North Carolina	34.55	Utah	10.59	Utah	5.99
Kentucky	27.38	Kentucky	31.83	Minnesota	12.52	North Dakota	6.11
New Mexico	27.00	West Virginia	30.64	North Dakota	13.91	Minnesota	6.90
West Virginia	26.09	Arkansas	29.39	Oregon	14.43	Rhode Island	7.08
Arkansas	25.89	South Carolina	26.46	Rhode Island	14.52	Oregon	8.59
Wisconsin	24.99	Nevada	26.16	Connecticut	14.81	Iowa	9.08
Delaware	24.04	Kansas	24.71	Illinois	14.91	Connecticut	10.44
Vermont	24.04	New Mexico	24.10	Virginia	15.08	South Dakota	10.70
South Carolina	23.82	Delaware	23.60	Iowa	15.48	Maryland	11.26
Kansas	23.35	Vermont	22.94	District of Columbia	15.59	Illinois	11.41
Ever-use of E-Cigarettes							
Highest rankings				Lowest rankings			
CHBC		GLMM		CHBC		GLMM	
New Mexico	56.16	New Mexico	75.28	Utah	26.19	Utah	12.66
Oklahoma	44.88	West Virginia	57.36	Connecticut	27.02	North Dakota	13.88
West Virginia	44.37	Oklahoma	55.94	North Dakota	27.77	Connecticut	15.12
North Carolina	44.30	North Carolina	51.07	Minnesota	28.83	Maryland	17.22
Kentucky	42.58	Nevada	49.20	Virginia	29.24	Minnesota	20.66
Nevada	42.06	Kentucky	49.11	Rhode Island	30.55	District of Columbia	23.90
Arizona	41.27	Arkansas	44.59	Maryland	31.80	Mississippi	24.59
Colorado	40.89	Kansas	44.47	Georgia	31.93	Rhode Island	25.08
Arkansas	40.74	Colorado	44.26	Mississippi	32.04	Oregon	25.57
Delaware	40.44	Delaware	44.03	Iowa	32.28	Virginia	25.83

Table D.3: Ten US states with the highest and lowest estimated prevalence of current-smoke and ever-smoke based on CHBC and GLMM.

Current-smoke							
Highest rankings				Lowest rankings			
CHBC		GLMM		CHBC		GLMM	
West Virginia	11.89	Iowa	15.50	Rhode Island	2.12	Rhode Island	0.92
Kentucky	10.28	West Virginia	15.44	Massachusetts	2.95	Utah	1.84
Iowa	9.94	Kentucky	11.47	Utah	3.42	Massachusetts	2.06
North Carolina	9.68	North Carolina	11.45	New Hampshire	3.59	Connecticut	2.14
South Carolina	8.81	South Carolina	11.23	Connecticut	3.69	Washington	2.33
Tennessee	8.25	Nebraska	10.24	Maine	3.70	California	2.65
Nebraska	8.11	Tennessee	10.00	Washington	4.20	New York	3.00
Montana	7.95	Ohio	8.74	New York	4.22	New Hampshire	3.00
New Mexico	7.91	Montana	8.59	California	4.24	Maine	3.13
Kansas	7.71	Alabama	8.04	Minnesota	4.46	Oregon	3.26
Ever-smoke							
Highest rankings				Lowest rankings			
CHBC		GLMM		CHBC		GLMM	
New Mexico	30.71	New Mexico	44.37	Maryland	3.48	District of Columbia	1.21
Kentucky	27.29	Kentucky	41.23	Rhode Island	9.09	Maryland	3.37
North Carolina	25.99	West Virginia	40.65	Connecticut	9.50	Rhode Island	3.97
West Virginia	24.34	North Carolina	37.89	Massachusetts	10.58	Connecticut	5.06
South Carolina	24.06	South Carolina	36.89	Virginia	10.98	Utah	6.48
Kansas	23.51	Tennessee	33.90	New York	11.83	Massachusetts	6.63
Oklahoma	23.06	Kansas	27.04	New Hampshire	12.42	New York	7.91
Tennessee	21.84	Oklahoma	26.45	Maine	12.77	New Hampshire	9.32
Texas	20.56	Ohio	25.77	Utah	13.24	Maine	9.66
Nebraska	20.56	Nebraska	24.84	Delaware	13.35	New Jersey	10.30

Figure D.1: County prevalence of ever-use and current-use of E-Cigarettes in the United States based on 2015-2017 YRBSS data.



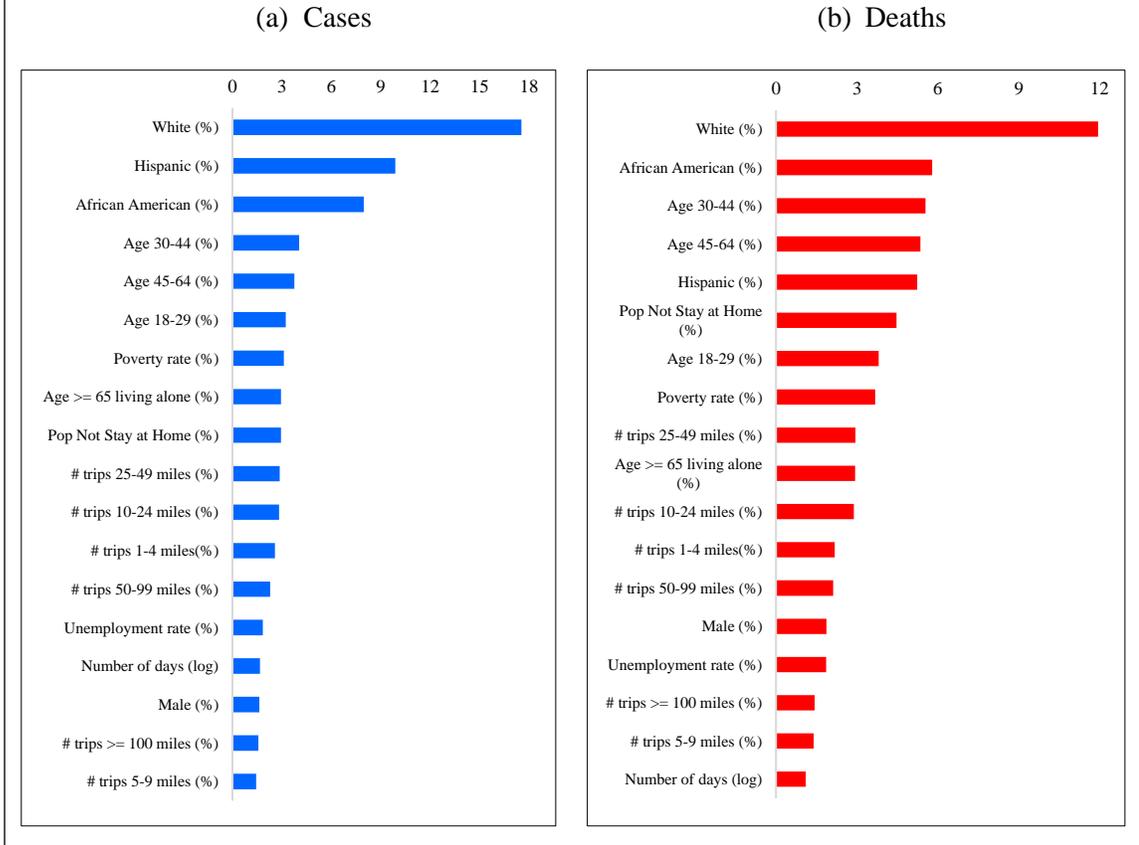
Appendix E. Benchmark Analysis for Poisson Mixed Model Using COVID-19

Data

Table E.1: MLEs for fixed effects and variance parameter using the GLMM for COVID-19 data.

Characteristics	CHBC				95% CI	
	Estimate	SE	Z Value	P(> Z)	Lower	Upper
Intercept	-8.838	0.427	-20.711	<0.001	-9.674	-8.001
Number of days (log)	0.141	0.045	3.166	0.002	0.054	0.228
Number of active cases (log)	0.325	0.008	40.467	<0.001	0.309	0.341
Population not stay home (%)	3.366	0.280	12.005	<0.001	2.816	3.915
Number of trips (%)	3.358	0.265	12.666	<0.001	2.838	3.877
< 1 mile	0.000					
1 - 4 miles	-1.437	0.328	-4.377	<0.001	-2.081	-0.794
5 - 9 miles	-1.602	0.369	-4.340	<0.001	-2.325	-0.878
10 - 24 miles	-1.059	0.383	-2.762	0.006	-1.811	-0.308
25 - 49 miles	2.233	0.518	4.314	<0.001	1.218	3.247
50 - 99 miles	3.470	0.968	3.585	<0.001	1.573	5.367
>= 100 miles	2.048	1.801	1.137	0.256	-1.483	5.578
65 Years or above living alone (%)	-1.322	0.465	-2.844	0.004	-2.234	-0.411
Age (%)						
<= 17 years	0.000					
18_29 years	-4.903	0.403	-12.180	<0.001	-5.692	-4.114
30_44 years	-2.400	0.678	-3.539	<0.001	-3.729	-1.071
45_64 years	-8.375	0.643	-13.016	<0.001	-9.636	-7.114
Race (%)						
White	0.630	0.141	4.460	<0.001	0.353	0.908
African American	2.316	0.153	15.168	<0.001	2.017	2.615
Hispanic	1.177	0.155	7.586	<0.001	0.873	1.482
Other races	0.000					
Gender (%)						
Male	4.058	0.553	7.335	<0.001	2.974	5.143
Female	0.000					
Poverty rate (%)	1.921	0.238	8.086	<0.001	1.455	2.386
Unemployment rate (%)	-4.049	0.796	-5.089	<0.001	-5.608	-2.490
Variance Component						
County-level	0.506	1.382	0.366	<0.001	0.000	3.215

Figure E.2: Variance inflation factor and standard error based on univariate mixed models for COVID-19 cases and deaths.



Appendix F. The Elements of \mathcal{S} and \mathcal{J} of the Binomial-Normal HGLM

Appendix F.1 Partial Derivatives of Model Parameters for General Class of HGLM Family

The elements of the score function \mathcal{S} and Fisher information matrix \mathcal{J} will be obtained by taking the first and second partial derivatives with respect to model parameters. We first illustrate the general case where the CHBC method applies to a broad class of HGLM family distributions taking the partial derivative with respect to model parameters in Appendix H.1. Appendix H.2 covers the CHBC method for Canonical family GLMs where it simplifies obtaining \mathcal{S} and \mathcal{J} elements.

First, consider the mixed logit model: $P(\mathbf{y}|\mathbf{u}) = (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1}$ and $f(\mathbf{u}) = (2\pi)^{-m/2}(\det(\mathbf{G}))^{-1/2} \exp\left(-\frac{1}{2}\mathbf{u}^T \mathbf{G}^{-1}\mathbf{u}\right)$.

From (4.3), the h -likelihood function can be expressed as

$$h = \mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})) - \frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{u} + c,$$

$$h = \mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})) - \frac{1}{2} \mathbf{u}^T \mathbf{G}^{-1} \mathbf{u} - \frac{1}{2} \log(\det(\mathbf{G})) + c,$$

where $\mathbf{1}^T$ is a unit vector with dimension $(1 \times N)$. Take the partial derivative with respect to $\boldsymbol{\beta}$ and \mathbf{u}

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = [\mathbf{X}^T \mathbf{y} - \mathbf{X}^T (1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}))^{-1} \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})]_{p \times 1}$$

$$= \mathbf{X}^T \mathbf{y} - \mathbf{X}^T \boldsymbol{\pi} = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}),$$

$$\frac{\partial h}{\partial \mathbf{u}} = [\mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T (1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}))^{-1} \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{G}^{-1} \mathbf{u}]_{m \times 1}$$

$$= \mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T \boldsymbol{\pi} - \mathbf{G}^{-1} \mathbf{u} = \mathbf{Z}^T (\mathbf{y} - \boldsymbol{\pi}) - \mathbf{G}^{-1} \mathbf{u},$$

where $\boldsymbol{\pi} = 1/(1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))$.

The components of the asymptotic variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$, \mathcal{J} also known as the observed information matrix that is calculated by

$$\mathbf{J} = \begin{bmatrix} -\frac{\partial^2 h}{\partial \boldsymbol{\beta}^2} & -\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} \\ -\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} & -\frac{\partial^2 h}{\partial \mathbf{u}^2} \end{bmatrix},$$

where

$$\begin{aligned} -\frac{\partial^2 h}{\partial \boldsymbol{\beta}^T \partial \boldsymbol{\beta}} &= \frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{X}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}) \mathbf{X}) \\ &= \mathbf{X}^T \frac{1}{(1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))} \left(\frac{1}{1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})} \right) \mathbf{X} \\ &= \mathbf{X}^T \boldsymbol{\pi} (\mathbf{1} - \boldsymbol{\pi}) \mathbf{X} \\ &= \mathbf{X}^T \mathbf{W} \mathbf{X}, \end{aligned}$$

$$\begin{aligned} -\frac{\partial^2 h}{\partial \boldsymbol{\beta} \partial \mathbf{u}} &= \mathbf{X}^T \frac{\partial}{\partial \mathbf{u}} ((1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u})) \\ &= \mathbf{X}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-2} \mathbf{Z} \\ &= \mathbf{X}^T \boldsymbol{\pi} (\mathbf{1} - \boldsymbol{\pi}) \mathbf{Z} \\ &= \mathbf{X}^T \mathbf{W} \mathbf{Z}, \end{aligned}$$

$$\begin{aligned} -\frac{\partial^2 h}{\partial \mathbf{u} \partial \boldsymbol{\beta}} &= -\frac{\partial}{\partial \boldsymbol{\beta}} (\mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} - \mathbf{G}^{-1} \mathbf{u}) \\ &= \mathbf{Z}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-2} \mathbf{X} \\ &= \mathbf{Z}^T \boldsymbol{\pi} (\mathbf{1} - \boldsymbol{\pi}) \mathbf{X} \\ &= \mathbf{Z}^T \mathbf{W} \mathbf{X}, \end{aligned}$$

$$-\frac{\partial^2 h}{\partial \mathbf{u}^T \partial \mathbf{u}} = -\frac{\partial}{\partial \mathbf{u}} (\mathbf{Z}^T \mathbf{y} - \mathbf{Z}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} - \mathbf{G}^{-1} \mathbf{u})$$

$$\begin{aligned} &= \mathbf{Z}^T (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-2} \mathbf{Z} + \mathbf{G}^{-1} \\ &= \mathbf{Z}^T \boldsymbol{\pi}(\mathbf{1} - \boldsymbol{\pi})\mathbf{Z} + \mathbf{G}^{-1} \\ &= \mathbf{Z}^T \mathbf{W}\mathbf{Z} + \mathbf{G}^{-1}, \end{aligned}$$

where \mathbf{W} is a $N \times N$ diagonal matrix with the diagonal elements of each block being $\boldsymbol{\pi}_i(\mathbf{1} - \boldsymbol{\pi}_i)$ for area i .

Appendix F.2 Based on Score Function for Canonical GLM Family

Consider h -likelihood

$$h = \ell_{y|u} + \ell_u,$$

where $\ell_{y|u} = \mathbf{y}^T(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) - \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}))$, and $\ell_u = -\frac{m}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{u} + c$.

Now, in canonical GLM form

$$\boldsymbol{\theta} = (\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}), b(\boldsymbol{\theta}) = \mathbf{1}^T \log(1 + \exp(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})) = \mathbf{1}^T \log(1 + \exp(\boldsymbol{\theta})), \phi = 1.$$

Differentiate $b(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$,

$$b'(\boldsymbol{\theta}) = (1 + \exp \boldsymbol{\theta})^{-1} \exp \boldsymbol{\theta} = (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1} = \boldsymbol{\pi},$$

where $\boldsymbol{\pi} = (1 + \exp(-\mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}))^{-1}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \mathbf{u})$.

Thus, the score function of $\ell_{y|u}$

$$\mathbf{S}(\boldsymbol{\theta}; \mathbf{y}) = \frac{\mathbf{y} - b'(\boldsymbol{\theta})}{\phi} = \mathbf{y} - \boldsymbol{\pi}$$

and

$$\frac{\partial h}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \mathbf{S}(\boldsymbol{\theta}; \mathbf{y}) = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}),$$

$$\frac{\partial h}{\partial \mathbf{u}} = \mathbf{Z}^T \mathbf{S}(\boldsymbol{\theta}; \mathbf{y}) + \nabla_{\ell_u}^1 = \mathbf{Z}^T (\mathbf{y} - \boldsymbol{\pi}) - \frac{1}{\sigma^2} \mathbf{u}.$$

Hence,

$$\mathbf{s} = \begin{pmatrix} \frac{\partial h}{\partial \boldsymbol{\beta}} \\ \frac{\partial h}{\partial \mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}) \\ \mathbf{Z}^T (\mathbf{y} - \boldsymbol{\pi}) - \frac{1}{\sigma^2} \mathbf{u} \end{pmatrix},$$

$$\mathbf{J} = \begin{bmatrix} \mathbf{X}^T \mathbf{W} \mathbf{X} & \mathbf{X}^T \mathbf{W} \mathbf{Z} \\ \mathbf{Z}^T \mathbf{W} \mathbf{X} & \mathbf{Z}^T \mathbf{W} \mathbf{Z} + \nabla_{\ell_u}^2 \end{bmatrix},$$

Where

$$\mathbf{V}_{\ell_{\mathbf{u}}}^2 = -\frac{\partial^2 \ell_{\mathbf{u}}}{\partial \mathbf{u}^T \partial \mathbf{u}} = \frac{1}{\sigma^2},$$

$$\mathbf{W} = \text{Diag} \left(\frac{1}{\text{Var}(\mathbf{y})(g'(\boldsymbol{\mu}))^2} \right),$$

and

$$g(\boldsymbol{\mu}) = \log \boldsymbol{\pi}(\mathbf{1} - \boldsymbol{\pi}) = \log \boldsymbol{\mu} \rightarrow g'(\boldsymbol{\mu}) = \boldsymbol{\mu}^{-1}.$$

Thus,

$$\mathbf{W} = \text{Diag} \left(\frac{1}{\boldsymbol{\mu}(\boldsymbol{\mu}^{-1})^2} \right) = \text{Diag}(\boldsymbol{\mu}) = \text{Diag}(\boldsymbol{\pi}(\mathbf{1} - \boldsymbol{\pi})).$$

Appendix G. R Package for CHBC Approach in SAE

A “hglmbc2” R package is developed for the proposed CHBC approach. The development version of the R package is available to download using devtools <https://niroshar.github.io/hglmbc2/>. The function “hglmbc” function estimates the fixed effects, random effects, and variance parameter based on the proposed method.

```
hglmbc(
  data,
  mformula = NULL,
  dom = NULL,
  y.family = "binomial",
  rand.family = "gaussian",
  tol = 1e-05,
  ...
)
```

Arguments

- data** a data frame.
- mformula** an object of class `myFormula`: a symbolic description of the model to be fitted. The details of the `mformula` is given under the details section.
- dom** a domain/cluster/small area to specify the random effect. e.g. numeric zip code, county, or state code, and also the name of the county or name of the state.
- y.family** a distribution from the exponential family to describe the error distribution. See "[family](#)".
- rand.family** a description of the distribution of random effects.
- tol** predefined tolerance value. Default value is `tol = 1e-5`.
- ...** other arguments, See the details section.

Value

An object of class `hglmbc` consists of the hierarchical maximum likelihood estimates (HMLEs) of fixed effects, random effects, and variance parameters with other values,

est.beta

HMLE of fixed effects.

re

HMLE of the random effects.

var.par

HMLE of the dispersion parameter for the random effects.

fe.cov

the estimated variance-covariance matrix of the fixed effects.

fe.cov

the estimated variance-covariance matrix of the random effects.

iter

the number of iterations at convergence.

AIC

A list of likelihood values for model selection purposes, where AIC is the AIC value, ("AIC"), hLik is the h-likelihood value.

Summary

a summary object of the fitted model.

Examples

```
# 1. Using ever use of smoke data set. Discrete and continuous variables are
defined.
data <- eversmoke
mformula <- "smoke_ever ~ as.factor(age) + as.factor(gender) + as.factor(race)
+ as.factor(year) + povt_rate"
dom <- "county"
y.family <- "binomial"
rand.family <- "gaussian"
hglmfit <- hglmfit(data=eversmoke, mformula, dom = "county",
y.family="binomial")
```

```
# 2. mformula is not defined,
resp <- "smoke_ever"
dom <- "county"
catX <- c("year", "gender", "race", "age")
contX <- "povt_rate"
hglmfit <- hglmfit(data = eversmoke, resp, dom = "county", fe.disc = catX,
fe.cont = contX, y.family = "binomial")
```

```
hglmfit <- function(data, mformula = NULL, dom = NULL, y.family =
"binomial", rand.family = "gaussian", tol=1e-05, ...)
{
fit.hglmfit <- match.call()
namedList <- list()
if(!exists("data") | is.null(dom)){
stop("Error! please define the data frame and random effect component.")
}
}
```

```

data <- data[order(data[,dom]), ]
# If mformula and dom is defined
if(!is.null(mformula) & exists("mformula") & exists("dom")){
  varOut <- getVars(mformula)
  resp <- as.character(varOut[[1]])
  fe.disc <- as.vector(varOut[[2]])
  fe.cont <- as.vector(varOut[[3]])
}
# if mformula is not defined, but resp is defined
if(is.null(mformula) & exists("resp") & exists("dom")){
  myformula <- myFormula(data, resp, dom)
  # mformula <- as.formula(myformula)
  varOut <- getVars(myformula)
  resp <- as.character(varOut[[1]])
  fe.disc <- as.vector(varOut[[2]])
  fe.cont <- as.vector(varOut[[3]])
}
if(is.null(mformula) & !exists("resp")){
  stop("At least mformula or the response variable (resp) with dom need to be
defined!")
}
# cov_data <- data[, !colnames(data) %in% paste0(y, re)]
fe.data <- data[, colnames(data) %in% c(fe.cont, fe.disc)]
## Get the design matrix for user defined referenced group
# Reference group for each categorical variable
if(!exists("ref.group")){
  minVal <- apply(data[, fe.disc], 2, min)
  ref.group <- as.vector(paste0(names(minVal), minVal))
}
X <- DesignM(data=fe.data, fe.cont, fe.disc, ref.group)
y <- data[, paste0(resp)]
uDom <- data[, paste0(dom)]
## Initial parameters
m <- length(unique(uDom))
## Call function to obtain initial parameters
beta0 <- initPar(data, resp, dom)
beta00 <- bInitOrder(beta0, fe.cont)
beta_new <- as.matrix(beta00$est)
theta0 <- 0.1
u_new <- u0 <- as.matrix(rnorm(m, 0, sd=sqrt(theta0)))
X <- as.matrix(X)
N <- nrow(data)
p <- ncol(X)

```

```

zeta_new <- rep(1,m)
cnty <- unique(uDom)
delta <- NULL
beta_new_all <- beta_new ; u_new_all <- u_new;
theta_all <- theta_new; col_names <- c()
theta_new_all <- c(theta_new); delta_final <- conv_iter <- iter <- 0

repeat{
  beta <- as.vector(beta_new)
  u <- as.vector(u_new)
  theta <- theta_new; zeta <- zeta_new
  BU_old <- rbind(as.matrix(beta),as.matrix(u))
  theta_inv <- solve(theta)
  G_inv <- kronecker(theta_inv, diag(m))
  corr_ZU <- Z%*(zeta*u_new + (theta*(1-zeta))/2)
  J <- as.matrix(HessianM(X, Z, beta = beta_new, u = u_new, theta, y.family,
rand.family))
  # J_inv <- solve(J, tol=10^-20)
  # For non-singular matrices the pseudoinverse is equivalent to the standard
inverse.
  J_inv <- pseudoinverse(J)
  D1h_B <- t(X)%*(y-P)
  D1h_u <- t(Z)%*(y-P) - G_inv%*u
  S <- as.vector(rbind(D1h_B, D1h_u))
  BU_new = BU_old + (J_inv%*S) # Newton Raphson
  beta_new <- as.vector(BU_new[1:length(beta)])
  u_new <- as.vector(BU_new[(length(beta)+1):length(BU_new)])
  convergence_beta <- abs(beta_new-beta)
  max(convergence_beta)
  beta_new_all <- cbind.data.frame(beta_new_all,value=round(beta_new,6))
  u_new_all <- cbind.data.frame(u_new_all,value=round(u_new,6))
  rm(P); rm(W); rm(J);
  # ++++++ #
  # Bias Correction #
  # ++++++ #
  # update J with estimated beta and u
  J <- as.matrix(HessianM(X, Z, beta = beta_new, u = u_new, theta, y.family,
rand.family))
  Jhat_Inv <- pseudoinverse(J)
  Tau <- Jhat_Inv[(p+1):(p+m),(p+1):(p+m)]
  Taut <- as.vector(diag(Tau))
  zeta <- as.vector(theta)/(as.vector(theta)+Taut)
  zeta <- as.vector(zeta)

```

```

corr_u <- zeta*u_new      # Corrected u_new
rm(P); rm(W); rm(J);
iter <- iter + 1
##### -----#####
#####      Estimate theta using Dh/Dtheta=0      #####
##### -----#####
D2h_UU <- t(Z)%*%W%*%Z + G_inv
J_inv22 <- pseudoinverse(D2h_UU)
theta_new <- 1/m*(t(u_new)%*%u_new)+1/m*func_trace(J_inv22)
### To get the variance of u based on J
J <- as.matrix(HessianM(X, Z, beta = beta_new, u = u_new, theta = theta_new,
y.family, rand.family))
J_inv <- pseudoinverse(J)
fe.cov <- J_inv[1:p,1:p]      ## Var-Cov matrix of fixed effects
re.cov <- J_inv[(p+1):(p+m),(p+1):(p+m)]  ## Var-Cov matrix of random effects
fe.var <- diag(fe.cov)      ### Var of fixed effects
fe.var <- round(fe.var,10)
fe.var <- fe.var[fe.var != 0]  ## Remove coefs of ref group(which has very small
values)
fe.std.err <- round(sqrt(fe.var), 5)
# ++++++#####
convergence_theta <- abs(theta_new-theta)
delta <- max(convergence_beta,convergence_theta)
delta_final <- cbind.data.frame(delta_final,value=delta)
col_names <- c(col_names,paste0("iter",iter))
# cat(paste0("delta: ",delta,"\n \n"))
if(delta <= tol){
break
}
}
est.beta <- round(beta_new, 5)
est.theta <- round(theta_new, 5)
est.re <- data.frame(Domain = cnty, est=round(u_new, 5))
z1 <- est.beta/fe.std.err
est.fe <- data.frame(est=est.beta, std.err=fe.std.err,Z0 = round(z1,5),
                    P = round(2*pnorm(-abs(z1)),5))
colnames(est.fe) <- c("Estimate","Std.Error","Z Value","P(>|Z|)")
rownames(est.fe) <- beta00[ ,1]
fit.hglm$est.fe <- est.fe
fit.hglm$iter <- iter
fit.hglm$est.beta <- est.beta
fit.hglm$re <- est.re
fit.hglm$var.par <- est.theta

```

```

fit.hglmabc$fe.cov <- fe.cov
fit.hglmabc$re.cov <- re.cov
fit.hglmabc$model.sel <- ModelSel(X, Z, est.beta, u = est.re, theta = est.theta,
y.family, rand.family)
  namedList <- list(`Model formula` = mformula, `random effect` = dom, `fixed
effects estimates` = est.fe, `dispersion parameter` = est.theta, `hglm model
inference` = model.sel, iter = iter, ` ` = paste0("Converged in ",iter," iterations
with tol = ",delta, "."))
  class(namedList) <- "summary.hglm.fit"
  fit.hglmabc$summary <- namedList
  return(fit.hglmabc)
  rm(mformula); rm(fe.cont); rm(fe.disc)
  # cat(paste0("Converged in ",iter," iterations with tol = ",delta, ". \n \n"))
}
HessianM <- function(X, Z, beta = NULL, u = NULL, theta = NULL, y.family, rand.family,
...)
{
  if(y.family = "binomial"){
    P <- 1/(1+exp(-(X%%beta+ Z%%u)))
    W <- Diagonal(P*(1-P))
  }else (y.family = "Poisson"){
    P <- exp(X%%beta+Z%%u)
    W <- Diagonal(P)
  }
  if(rand.family %in% c("gaussian","normal")){
    theta_inv <- solve(theta)
    D2Lu <- G_inv <- kronecker(theta_inv, diag(m))
  }
  D2h_BB <- t(X)%%W%%X # or crossprod(t(crossprod(X,W)),X)
  D2h_BU <- t(X)%%W%%Z
  D2h_UB <- t(Z)%%W%%X
  D2h_UU <- t(Z)%%W%%Z + D2Lu
  J <- rbind(cbind(D2h_BB,D2h_BU),cbind(D2h_UB,D2h_UU))
  J <- as.matrix(J)
  return(J)
}

## Calculate AIC and adjusted h-likelihood
ModelSel <- function(X, Z, beta = NULL, u = NULL, theta = NULL, offset = NULL,
y.family, rand.family, ...)
{
  if(tolower(y.family) = "binomial"){
    P <- 1/(1+exp(-(X%%beta+ Z%%u)))
    yt <- 1-y
    logLik_y <- y*log(P) + yt*log(1-P)
  }else if(tolower(y.family) = "poisson"){

```

```

        If(!is.null(offset)){
            thetaGLM <- exp(X%%beta+Z%%u + offset)
        }else{
            thetaGLM <- exp(X%%beta+Z%%u)
        }
        y <- ifelse(y==0, 0.0000001, y)
        log_facty <- 1/2*log(2*pi*y) + y*log(y) - y
        logLik_y <- y*log(lambda) - thetaGLM - log_facty
    }
    if(rand.family %in% c("gaussian","normal")){
        theta_inv <- solve(theta)
        G_inv <- kronecker(theta_inv, diag(m))
        logLik_u <- -m/2*log(theta) -1/2*t(u_new)%%G_inv%%u_new - m/2*log(2*pi)
    }

    ## Calculate AIC
    sumlogylik <- sum(logLik_y)
    sumlogulik <- sum(logLik_u)
    hlik <- sumlogylik + sumlogulik
    AIC <- -2*(sumlogylik + sumlogulik) + 2*(p + length(theta)+1)
    model.sel <- cbind.data.frame(AIC, hLik = hlik)
    return(model.sel)
}

## Get variable names
getVars <- function(mformula = NULL)
{
    if(is.null(mformula)){
        stop("mformula needs to be defined!")
    }else if(!is.null(mformula)){
        mform <- as.formula(mformula)
        y_var <- as.character(gsub("\\~.*", "", mform)[2])
        mform <- as.character(unlist(splitFormula(mform, sep = "+")))
        mform <- as.character(gsub("~", "", mform))
        x_disc <- as.vector(vars_select(mform, starts_with("as.factor")))
        x_cont <- mform[!mform %in% x_disc]
        x_disc <- gsub("as.factor\\(", "", x_disc)
        x_disc <- as.vector(gsub("\\)", "", x_disc))
        varOut <- list(y_var,x_disc,x_cont)
        return(varOut)
    }
}

```

```

## Get model formula
myFormula <- function(data,resp=NULL,dom=NULL, ...){
  if(!exists("resp") | !exists("dom")){
    stop("Please define the response variable and random effect!")
  }
  IntParFomular0 <- paste0(resp,"~")
  dataF <- data[ ,!colnames(data) %in% c(resp,dom)]
  if(ncol(dataF) > 0){
    if(!exists("fe.cont") & !exists("fe.disc")){
      # If discrete and continous variables are not defined, consider based on
variable type
      cat("Fixed effects and random effects are not defined, selected by variable
type !!!!", "\n\n\n")
      # Get discrete variables by variable type
      unqV <- lapply(dataF, unique)
      unqVC <- unlist(lapply(unqV, length))
      fe.disc <- c(names(unqVC[unqVC<=5]))
      fe.cont <- c(names(unqVC)[!names(unqVC) %in% fe.disc])
      fe.cont <- c(fe.cont[fe.cont != c(dom) & fe.cont != c(resp)])
      discFomular <- IntParFomular0
      if(length(fe.disc) != 0){
        for(i in 1:length(fe.disc)){
          temp <- paste0("as.factor(", fe.disc[i],")+")
          discFomular <- paste0(discFomular,temp)
        }
      }
      if(length(fe.cont) != 0){
        for(j in 1:length(fe.cont)){
          temp <- paste0(fe.cont[j],"+") # paste0(fe.cont[j])
          discFomular <- paste0(discFomular,temp)
        }
      }
      IntParFomular <- discFomular
    }else if(!exists("fe.cont") & exists("fe.disc")){
      fe.cont <- c(names(which(lapply(dataF, is.character)==FALSE)))
      fe.cont <- c(fe.cont[fe.cont != c(dom) & fe.cont != c(resp)])
      discFomular <- IntParFomular0
      for(i in 1:length(fe.disc)){
        temp <- paste0("as.factor(", fe.disc[i],")+")
        discFomular <- paste0(discFomular,temp)
      }
      if(length(fe.cont) != 0){
        for(j in 1:length(fe.cont)){
          temp <- paste0(fe.cont[j],"+")

```

```

        discFomular <- paste0(discFomular,temp)
    }
}
IntParFomular <- discFomular
}else if(exists("fe.cont") & !exists("fe.disc")){
    fe.disc <- c(names(which(lapply(dataF, is.character)==TRUE)))
    discFomular <- IntParFomular0
    if(length(fe.disc) != 0){
        for(i in 1:length(fe.disc)){
            temp <- paste0("as.factor(", fe.disc[i], "+")
            discFomular <- paste0(discFomular,temp)
        }
    }
    for(j in 1:length(fe.cont)){
        temp <- paste0(fe.cont[j], "+")
        discFomular <- paste0(discFomular,temp)
    }
    IntParFomular <- discFomular
}else{
    fe.disc <- fe.disc
    fe.cont <- fe.cont
    discFomular <- IntParFomular0
    for(i in 1:length(fe.disc)){
        temp <- paste0("as.factor(", fe.disc[i], "+")
        discFomular <- paste0(discFomular,temp)
    }
    for(j in 1:length(fe.cont)){
        temp <- paste0(fe.cont[j], "+")
        contFomular <- paste0(discFomular,temp)
    }
    IntParFomular <- contFomular
}
if(sapply(strsplit(as.character(IntParFomular), ""), tail, 1)==""){
    rmP <- sapply(strsplit(as.character(IntParFomular), ""), tail, 1)
    IntParFomular <- stri_replace_last(IntParFomular, fixed = "+", "")
}
mformula <- IntParFomular
}else{
    stop("Error! No other variables exist except response response variable and
random effect")
}
return(mformula)
}

```

```

## Get the design matrix
DesignM <- function(data, fe.cont, fe.disc, ref.group){
  DX_out <- data[, c(fe.cont)]
  for(i in 1:length(fe.disc)){
    # i <- 1
    var_temp <- fe.disc[i]
    DX_temp <- data[, var_temp]
    DX_temp <- as.factor(DX_temp)
    DX_out <- cbind.data.frame(DX_out, DX_temp)
  }
  colnames(DX_out) <- c(fe.cont, fe.disc)
  DX11 <- dummy_cols(DX_out, select_columns = paste0(fe.disc)) %>%
    select(-c(paste0(fe.disc)))
  # DX11 <- data.frame(Intercept=rep(1,nrow(DX_out)), DX11)
  colnames(DX11) <- gsub("_", "", colnames(DX11))
  DX11 <- DX11[ , !colnames(DX11) %in% ref.group] ### Remove ref group
  DX <- DX11[ , order(colnames(DX11))]
  X <- data.frame(Intercept=rep(1,nrow(DX)), DX)
  return(X)
}

## Get trace of a matrix
func_trace <- function(X){
  n <- dim(X)[1]
  tr <- 0 ### initialize trace
  for (j in 1:n){
    k <- X[j,j]
    tr <- tr + k
  }
  return(tr[[1]])
}

## Get initial parameters
initPar <- function(data, resp, dom, fe.disc = NULL, fe.cont = NULL, y.family =
NULL, ...){
  myFormula <- myFormula(data, resp, dom)
  mformula <- as.formula(myFormula)
  # Returns the dist of y if not specified
  if(is.null(y.family)){
    yVal <- data[, resp]
    if(length(unique(yVal))==2){
      y.family <- "binomial"
      cat(paste0("Distribution of response variable is taken as", y.family))
    }else if(length(unique(yVal)) > 2 & is.numeric(yVal)){
      y.family <- "gaussian"
    }
  }
}

```

```

    cat(paste0("Distribution of response variable is taken as", y.family))
  }else{
    stop("Error: Family is not defined !!!")
  }
}
}else{
  y.family <- y.family
}
if(y.family == "binomial"){
  mdlFit <- glm(formula = mformula, family = binomial(link=logit), data=data)
}else if(y.family == "Poisson"){
  mdlFit <- glm(formula = mformula, family = poisson(link=log), data=data)
}else if(y.family == "gaussian"){
  mdlFit <- glm(formula = myFormula, family = gaussian(link="identity"), data=data)
}
beta0 <- data.frame(summary(mdlFit)$coefficients[,1])
beta0 <- data.frame(parameter=rownames(beta0), est=beta0[,1])
rownames(beta0) <- NULL
return(beta0)
}
## Get initial parameters ordered by variable names
bInitOrder <- function(beta0, fe.cont){
  beta0_cont <- beta0[grepl("^as.factor", beta0$parameter)==FALSE, ] ## beta for
non-discrete parameters(including intercept)
  beta0_Int <- beta0_cont[!beta0_cont$parameter %in% fe.cont, ] ## Intercept
  beta0_cont <- beta0_cont[beta0_cont$parameter %in% fe.cont, ]
  beta0_cont$parameter <- gsub("_", "", beta0_cont$parameter)
  beta0_Int$parameter <- gsub("\\(Intercept\\)", "Intercept", beta0_Int$parameter)
  beta0_dis <- beta0[grepl("^as.factor", beta0$parameter)==TRUE, ] # beta for disc
vars
  beta0_dis$parameter <- gsub("as.factor\\(", "", beta0_dis$parameter)
  beta0_dis$parameter <- gsub("\\)", "_", beta0_dis$parameter)
  beta0_dis$col <- gsub("_.*", "", beta0_dis$parameter)
  unq <- unique(beta0_dis$col)
  beta0_dis$parameter <- gsub("_", "", beta0_dis$parameter)
  beta0_dis$col <- NULL
  beta0 <- rbind(beta0_cont, beta0_dis)
  if(!is.character(beta0$parameter)){
    beta0$parameter <- as.character(beta0$parameter)
  }
  beta0 <- beta0[order(beta0$parameter), ]
  beta0$est <- round(as.numeric(beta0$est), 5)
  beta00 <- rbind.data.frame(beta0_Int, beta0)
  return(beta00)
}

```

```
}
```

`hglmbcJM` model fit function in the multivariate joint model. The MHLEs are obtained using the CHBC method.

```
hglmbcJM(
  data,
  JMformula = NULL,
  dom = NULL,
  JMy.family = c("Poisson", "Poisson")
  rand.family = "gaussian",
  tol = 1e-05,
  ...
)
```

Arguments

- data** a data frame to be used for the multivariate joint model.
 - JMformula** A list with model formulas for each outcome variable, [myFormula](#): a symbolic description of the model to be fitted. The details of the `mformula` is given under the details section. Ex. For two outcomes (y1,y2), `mformula1 = y1~x1+x2+x3`, `mformula2 = y2~x1+x2`, `mvformula = list(mformula1, mformula2)`.
 - dom** A domain/cluster/small area to specify the random effect. e.g. numeric zip code, county, or state code, and also the name of the county or name of the state.
 - JMy.family** A vector of distributions for each response from the exponential family to describe the error distribution. `multiY.family = c("Poisson", "Poisson")`. See ["family"](#).
 - rand.family** A description of the distribution of random effects. Default is $u \sim \mathcal{N}(0, \sigma^2)$.
 - tol** Predefined tolerance value. Default value is `tol = 1e-5`.
 - ... Other arguments, See the details section.
- ```
Obtain MHLEs of the bivariate joint model
X <- addiag(X1, X2)
Z <- addiag(Z1, Z2)
y <- rbind(as.matrix(y1), as.matrix(y2))
hglmbcJM <- function(data, JMformula = NULL, dom = NULL, JMy.family = NULL, rand.family =
"gaussian", tol=1e-05, ...)
```

```

{
 repeat{
 beta1 <- as.vector(beta1_new) ### fixed effects for resp 1
 beta1 <- round(beta1,10)
 beta2 <- as.vector(beta2_new) ### fixed effects for resp 2
 beta2 <- round(beta2,10)
 ## Bias correction
 ## bias correction- replace u_hat by mu_tilde = sigma(sigma + v_i)^-1*u_hat
 ## and exp(u_hat) by exp(mu_tilde + 1/2 diag(sigma_tilde)), where
 ## sigma_tilde = (sigma - sigma (sigma + v_i)^-1 sigma)
 est_U <- rbind(as.matrix(u1_new), as.matrix(u2_new))
 V <- V_new
 sigmaBigM <- kronecker(sigmaNew, diag(m))
 sigmaV_inv <- (sigmaBigM + V)
 if(!is.na(det(sigmaV_inv))){
 mu_tilde <- round(sigmaBigM%%pseudoinverse(sigmaV_inv),10)%%est_U
 }else if(is.na(det(sigmaV_inv))){
 next
 }
 # mu_tilde <- round(sigmaBigM%%pseudoinverse(sigmaV_inv),10)%%est_U
 corr_U <- mu_tilde # ----- **
 corr_u1 <- corr_U[1:m]
 corr_u2 <- corr_U[(m+1):nrow(corr_U)]
 sigmaV_inv <- (sigmaBigM + V)
 # sigma_tilde <- (sigmaBigM - sigmaBigM%%pseudoinverse(sigmaV_inv)%%sigmaBigM)
 if(!is.na(det(sigmaV_inv))){
 sigma_tilde <- (sigmaBigM - sigmaBigM%%pseudoinverse(sigmaV_inv)%%sigmaBigM)
 }else if(is.na(det(sigmaV_inv))){
 next
 }
 corr_ZU <- Z%%(mu_tilde + 1/2*diag(sigma_tilde)) # not corrected exp(ZU)
 corr_Zu1 <- corr_ZU[1:N1]
 corr_Zu2 <- corr_ZU[(N1+1):nrow(corr_ZU)]
 ## Bias corrected u
 est_u1 <- as.vector(corr_u1)
 est_u2 <- as.vector(corr_u2)
 est_U <- corr_U # rbind(as.matrix(est_u1), as.matrix(est_u2))
 beta_old <- rbind(as.matrix(beta1),as.matrix(beta2))
 BDU_old <- rbind(as.matrix(beta_old), as.matrix(est_U))
 sigma_inv <- solve(sigmaNew)
 Q <- kronecker(sigma_inv, diag(m)) ## create 2m x 2m of sigma inverse
 out1 <- funcW(yfamily1,X1,beta1,corr_Zu1)
 W1 <- out1[[1]]
 }
}

```

```

mu1 <- out1[[2]]
out2 <- funcW(yfamily2,X2,beta2,corr_Zu2)
W2 <- out2[[1]]
mu2 <- out2[[2]]
mu <- as.matrix(rbind(mu1,mu2))
D1Lu <- -dot(est_U,Q) # -U(sigma_inv)I_mm
D2Lu <- Q ## create 2m x 2m of sigma inverse
D1h_B <- t(X)%*(y-mu)
D1h_U <- t(Z)%*(y-mu) + D1Lu #- dot(Ur,sigmaInvBlockDiag)
S <- as.vector(rbind(D1h_B, D1h_U))
Hessian matrix
W <- addiag(as.matrix(W1), as.matrix(W2))
Q <- sigmaInvBlockDiag
D2h_BB <- t(X)%*W*X
D2h_BU <- t(X)%*W*Z
D2h_UB <- t(Z)%*W*X
D2h_UU <- t(Z)%*W*Z + D2Lu
J <- rbind(cbind(D2h_BB,D2h_BU),cbind(D2h_UB,D2h_UU))
J <- as.matrix(J)
if(is.finite(det(J)) & det(J) != 0){
J_inv <- pseudoinverse(J)
if(!is.na(det(J))){
 J_inv <- pseudoinverse(J)
}else if(is.na(det(J))){
 next
}
Newton Raphson
BDU_new = BDU_old + (J_inv*S)
p <- p1 + p2
beta_new <- as.vector(BDU_new[1:p])
u_new <- as.vector(BDU_new[(p+1):length(BDU_new)])
beta1_new <- beta_new[1:p1]
beta2_new <- beta_new[(p1+1):p]
u1_new <- u_new[1:m]
u2_new <- u_new[(m+1):(2*m)]
convergence_beta <- abs(beta_new - beta_old)
rm(P1); rm(P2); rm(W);
beta_new_all <- cbind.data.frame(beta_new_all,value=round(beta_new,6))
u_new_all <- cbind.data.frame(u_new_all,value=round(u_new,6))

Estimate sigma - var-cov matrix of U

score function of h_A = d(h_A)/d(theta)

```

```

U_new <- rbind(as.matrix(u1_new),as.matrix(u2_new))
Q <- kronecker(sigma_inv, diag(m))
QPrime <- -1*Q**Q
Q2Prime <- 2*kronecker((sigma_inv**sigma_inv**sigma_inv), diag(m))
corr_Zu1 <- Z1**u1_new
out1 <- funcW(yfamily1,X1,beta1_new,corr_Zu1)
W1 <- out1[[1]]
mu1 <- out1[[2]]
corr_Zu2 <- Z2**u2_new
out2 <- funcW(yfamily2,X2,beta2_new,corr_Zu2)
W2 <- out2[[1]]
mu2 <- out2[[2]]
W <- addiag(as.matrix(W1), as.matrix(W2))
D2h_UU <- t(Z)**W**Z + Q
if(!is.na(det(D2h_UU))){
 J22_inv <- pseudoinverse(D2h_UU)
}else if(is.na(det(D2h_UU))){
 next
}
J22_inv <- pseudoinverse(D2h_UU)
SA1 <- -m/2*func_trace(sigma_inv)
Bias correction for u - start -----
V <- J22_inv
sigmaBigM <- kronecker(sigmaNew, diag(m))
sigmaV_inv <- sigmaBigM + V
if(!is.na(det(sigmaV_inv))){
 mu_tilde <- sigmaBigM**pseudoinverse(sigmaV_inv)**u_new
}else if(is.na(det(sigmaV_inv))){
 next
}
mu_tilde <- sigmaBigM**pseudoinverse(sigmaV_inv)**u_new
corr_U <- mu_tilde # ----- **
Bias correction for u - end -----
SA2 <- -1/2*t(corr_U)**QPrime**corr_U
SA30 <- J22_inv**QPrime
SA3 <- -1/2*func_trace(SA30)
D1hA_Sigma <- SA1 + SA2 + SA3 # ----- ***
components of 2nd derivative wrt to sigma
JA1 <- m/2*func_trace(sigma_inv**sigma_inv)
JA2 <- -1/2*(t(corr_U)**Q2Prime**corr_U)
JA30 <- J22_inv**Q2Prime - QPrime**J22_inv**QPrime**J22_inv
JA3 <- -1/2*func_trace(JA30)
JA <- JA1 + JA2 + JA3

```

```

D2hA_Sigma <- -1*JA
D2hASigmaInv <- solve(D2hA_Sigma)
sigmaOld <- sigmaNew
D2D1 <- D2hASigmaInv%*%D1hA_Sigma
sigmaNew <- sigmaOld + as.numeric(D2D1)
rm(D2D1);
Update V with corrected u for next iteration
sigma_inv <- solve(sigmaNew)
Q <- kronecker(sigma_inv, diag(m))
for poisson
corr_Zu1 <- Z1%*%u1_new
out1 <- funcW(yfamily1,X1,beta1_new,corr_Zu1)
W1 <- out1[[1]]
mu1 <- out1[[2]]
corr_Zu2 <- Z2%*%u2_new
out2 <- funcW(yfamily2,X2,beta2_new,corr_Zu2)
W2 <- out2[[1]]
mu2 <- out2[[2]]
W <- adiaq(as.matrix(W1), as.matrix(W2))
D2h_UU <- t(Z)%*%W%*%Z + Q
if(!is.na(det(D2h_UU))){
 J22_inv <- pseudoinverse(D2h_UU)
}else if(is.na(det(D2h_UU))){
 Next
}
J22_inv <- pseudoinverse(D2h_UU)
V_new <- J22_inv ## ----- **
convergence_sigma <- abs(sigmaNew - sigmaOld)
delta <- max(convergence_beta, convergence_sigma)
print(paste0("tolerance: ", delta))
print(paste0("No of iterations: ", iter))
iter <- iter + 1
iter_all <- c(iter_all,paste0("iter",iter))
iter_all <- c(iter_all, iter)
if(delta <= 1e-5){
 break
}else if(iter >= 100){
 Break
}
}
}

Function to get the weight matrix

```

```
funcW(X,beta,Z,u)
funcW <- function(yfamily,X,beta,corr_zu){
 if(tolower(yfamily) %in% "poisson"){
 P <- exp(X%%beta + corr_zu)
 mu <- P
 W <- Diagonal(x=as.vector(P))
 out <- list(W,mu)
 }
 return(out)
}
```