DATA COMBINING USING MIXTURES OF G-PRIORS WITH APPLICATION ON COUNTY-LEVEL FEMALE BREAST CANCER PREVALENCE

A Dissertation presented to
the Faculty of the Graduate School
at the University of Missouri

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

---

by

QIAO WANG

Drs. Dongchu Sun, Chong (Zhuoqiong) He, Dissertation Supervisors

May 2022

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

DATA COMBINING USING MIXTURES OF *G*-PRIORS WITH APPLICATION ON COUNTY-LEVEL FEMALE BREAST CANCER PREVALENCE

presented by Qiao Wang,

a candidate for the degree of Doctor of Philosophy and hereby certify that, in their opinion, it is worthy of acceptance.

Dr. Dongchu Sun

Dr. Chong (Zhuoqiong) He

Dr. Athanasios (Sakis) Christou Micheas

Dr. Yushu Shi

Dr. Jeannette Jackson-Thompson

ACKNOWLEDGMENTS

First and foremost, I want to express my sincere gratitude towards my advisors, Drs. Dongchu Sun and Chong (Zhuoqiong) He, for their guidance and patience over the years. This work is only possible because of their expertise and insights. They also showed and taught me the attitude of conducting rigorous research, the ability to identify research questions, and the skill to find innovative solutions.

I would also love to thank my committee members, Drs. Sakis Micheas, Yushu Shi and Jeannette Jackson-Thompson. They introduced me to so many interesting statistical courses as well as a colorful world where statistical knowledge can be applied to improve people's lives. They offered valuable suggestions to improve my research, and prepared me to be a qualified statistician.

Specially, I am very thankful to my wonderful mentors, Drs. Chester Lee Schmaltz and Jeannette Jackson-Thompson at the Missouri Cancer Registry and Research Center, for their constant support and encouragement. They not only generously shared with me their professional experience and wisdom, but also provided many opportunities to help me become a better researcher and collaborator.

Finally, I would love to thank my family and friends. I am forever in debt to my parents and brother, Zhibiao Wang, Ruilin Guo, and Qing Wang, for their unconditional love and unwavering support. I am very grateful for all of my friends for their enjoyable companionship and help. I am highly appreciative of my boyfriend, Zhongheng Cai, for his understanding and support. Despite ups and downs, he stands by me, and his care and love inspire me to move forward and become a better version of myself.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

ABSTRACT

As more and more data are available, data synthesis has become an indispensable task for researchers. From a Bayesian perspective, this dissertation includes three related projects and aims at quantifying the benefits of combining data under various scenarios in terms of the theoretical properties including biases, frequentist variances, and mean squared errors.

In the first project, data combining of linear models with the classical mixtures of g-priors is investigated. We calculate and compare the posterior estimates and the frequentist properties of the Bayesian estimator from the model with individual and combined data.

To resolve the newly identified conditional Lindley paradox and relax constraints on design matrix, data combining with independent mixtures of g-prior is explored, where a different scale is used for each group of coefficients. We not only perform a posterior variance analysis, but also offer a conditional asymptotic analysis of the Bayesian estimators. We also apply the corresponding results in the comparison of models for individual and combined data. Furthermore, to reflect how the use of sample size impact the estimates in a data combining context, we compare the Zellner-Siow prior to its adjustment with the effective sample size.

At last, an application on data combining of the 2016 county-level female breast cancer prevalence is presented using data from the Missouri Cancer Registry and Research Center, and the Missouri County-level Study. To provide a broader scope of the data combining framework, we study the linear mixed model and generalized linear mixed model with a conditional autoregressive prior serving as random effects.

Chapter 1

Introduction

1.1 Research on Data Combining Strategies

Even though a lot of data are available for analysis in the big data era, complex issues inherited from the data availability, collection and preparation still exist. Combining data from multiple data sources is a challenge for researchers in many fields. In general, combining data is carried out through either a direct linkage of databases or statistical methods (Chen et al., 2020; Thomsen and Holmøy, 1998). For the first, it is not frequently conducted by researchers due to discrepancies in case definitions, data qualities, data availabilities, data sharing regulations, etc. Therefore, a large number of statistical methods are developed to combine data according to available data sources and potential research questions.

One of the most popular data syntheses tool is meta-analysis (Brockwell and Gordon, 2001; Burke et al., 2017; Jackson et al., 2011; Riley et al., 2007, 2008) and a typical task is the inference for an overall effect or quantifying variabilities within or across multiple data sources. It has widespread applications in many fields including clinical trails (Moreno et al., 2018; Verde et al., 2016), psychology (Williams et al., 2018), medical science (Jahan et al., 2020; Lin and Chu, 2018), etc. Bayesian meta-analysis receives tremendous attention due to its sound performance in some challenging situations such as a small heterogeneity across studies (Chung et al., 2013; Hong et al., 2021) and incomplete outcomes from some data sources (Wei and Higgins, 2013). When the primary focus is comparing multiple treatments, Bayesian network is more frequently used since it integrates direct evidence such as data from arm-based method and indirect evidence such as contrast-based method (Li et al., 2021; Siegel et al., 2020; Zhang et al., 2014).

Besides meta-analysis framework, other data combining methods have also been developed to suit different practical considerations. A common method to combine models is model averaging, which aims at combining different distributions and offering better model selection and prediction (Fragoso et al., 2018; Hoeting et al., 1999; Yuan and Yang, 2005). When the major difficulty lies in a small sample size, one may incorporate information using a larger data set such as the administrative data to obtain a reasonable weight to improve the estimation. The small area estimation techique (Mercer et al., 2014; Pfeffermann et al., 2013) is a common method to deal with such situation. Additionally, Jackson et al. (2009) studied the Bayesian graphical model and imputed missing covariates utilizing other data sources. Zellner (1962) combined seemingly unrelated linear regression models with correlated random errors, and studied general properties of the estimates from a frequentist perspective.

This dissertation intends to generalize the classical model in Zellner (1962) from a

Bayesian perspective. Compared with their unrelated regressions, we allow different regression models to share common covariates and model specific covariates with a simpler random error assumption. Our framework describes a common situation, where some covariates are available to all data sources while some covariates are only available to some sources. Instead of turning to imputation for covariates that are not collected for some sources, we focus on a direct synthesis of available data in its original form. Besides, we target at quantifying the differences in the Bayesian estimators between using the individual data and combined data.

1.2 Research on Prior Specifications

Prior specification plays a key role in the Bayesian framework. This can be greatly reflected in the context of data combining due to the flexibility in prior constructions. For example, one can employ the informative prior eliciting from external data sources or special structure assumptions among multiple outcomes (Bujkiewicz et al., 2016, 2013; Wei and Higgins, 2013). Alternatively, to avoid the subjectivity in prior specification, one may specify non-informative prior such as reference prior (Bodnar et al., 2017). Hurtado Rúa et al. (2015) investigated how the choice of prior distributions impact estimates for coefficients and covariates through extensive simulation studies under the multivariate Bayesian meta-analysis framework. Despite the informative or non-informative version of prior, it is without doubt that a multivariate normal prior, conditional on other parameters, is one of the most used priors for regression coefficients in a linear model for its simple structure and efficient computation regarding posterior distributions.

One classical option is the G-prior or mixtures of g-prior, which is frequently known for its desirable model selection properties. G-prior or Zellner's g-prior refers to Goel and Zellner (1986), where they calculated the corresponding sampling distributions for the Bayesian estimator along with Bayes factor. Due to its convenience in obtaining a closed form of marginal likelihood and Bayes factor, many literature put efforts in finding a suitable value for the scale parameter so that some classical model selection criteria can be met (George, 2000; Kass and Wasserman, 1995). Mixtures of g-prior refers to the case where the scale parameter in g-prior is considered random rather than fixed. The earlist work is Zellner and Siow (1980) and they proposed to employ an inverse gamma distribution for g, which is equivalent to marginally applying a Cauchy prior for coefficients. Later on, many variants have been developed and one of the most influential work is Liang et al. (2008). Besides the Bartlett-Lindley paradox associated with a fixed choice of g, they proved that Zellner's g-prior may lead to information paradox. Specifically, they proposed hyper-g prior as a solution, and showed that both hyper-g prior and Zellner-Siow prior are not only free from the information paradox but also hold other desired model selection properties. However, to the best of our knowledge, rare literature explores the g-prior and mixtures of g-prior from the estimation perspective except its connection with ridge regression and its comparison with least squares estimates. Therefore, we not only intend to bring the g-prior or the mixtures of g-prior into the estimation scope but also compare its related properties between the individual and combined data.

1.3 Proposed Data Combining Framework

This section describes the model and notations throughout the dissertation unless stated otherwise.

1.3.1 Model Specification

In this section, models for the individual and combined data are specified. Notice that, although only two data sources are considered in the formal analysis, our framework and theoretical results can be extended to multiple data sources. Let $\boldsymbol{y}_i$ be a n_i-dimension vector of observations in Source $i, i = 1, 2$, and assume that the model for an individual data source i, denoted as M_i, is defined as:

$$\boldsymbol{y}_i = \boldsymbol{X}_{0i}\boldsymbol{\beta}_0 + \boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i, \tag{1.1}$$

where $\boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2\boldsymbol{I}_{n_i})$, σ^2 is the common error variance, $\boldsymbol{\beta}_0 \in \mathbb{R}^{p_0}$ and $\boldsymbol{\beta}_i \in \mathbb{R}^{p_i}$ are vectors of unknown regression coefficients, and $\boldsymbol{X}_{0i}$ with dimension $n_i \times p_0$ and $\boldsymbol{X}_i$ with dimension $n_i \times p_i$ are the corresponding design matrices. For the combined data from Sources 1 and 2, the model, denoted as M_c, is defined as:

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{01} \\ \boldsymbol{X}_{02} \end{pmatrix} \boldsymbol{\beta}_0 + \begin{pmatrix} \boldsymbol{X}_1 & 0 \\ 0 & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}, \tag{1.2}$$

where $\boldsymbol{y}_i, \boldsymbol{X}_{0i}, \boldsymbol{X}_i, \boldsymbol{\beta}_i,$ and $\boldsymbol{\epsilon}_i$ are defined the same as Source i. Thus, $\boldsymbol{\beta}_0$ is common regression coefficient shared and collected by two sources and $\boldsymbol{\beta}_i$ is the Source i specific coefficient.

1.3.2 Notations

For simplicity, we use the following notations for regression coefficients and design matrices in M_i and M_c.

- $\boldsymbol{\beta}_0 = (\beta_{01}, \cdots, \beta_{0p_0})'$ denotes common coefficients;
- $\boldsymbol{\beta}_i = (\beta_{i1}, \cdots, \beta_{ip_i})'$ denotes Source i specific coefficients;
- $\tilde{\boldsymbol{X}}_i = (\boldsymbol{X}_{0i}\ \boldsymbol{X}_i)$ denotes the design matrix in M_i;
- $\tilde{\boldsymbol{\beta}}_i = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_i')'$ of dimension $p_I = p_0 + p_i$ denotes all the regression coefficients in M_i;
- $\tilde{\boldsymbol{\beta}}_i^B$ denotes the posterior mean for $\tilde{\boldsymbol{\beta}}_i$ in M_i;
- $\boldsymbol{y} = (\boldsymbol{y}_1', \boldsymbol{y}_2')'$ is a n_T-dimension vector of the observations in M_c with $n_T = n_1 + n_2$;
- $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ of dimension $p_T = p_0 + p_1 + p_2$ denotes all the regression coefficients in M_c;
- $\tilde{\boldsymbol{\beta}}^B$ denotes the posterior mean for $\tilde{\boldsymbol{\beta}}$ in M_c;
- $\boldsymbol{X}_0 = (\boldsymbol{X}_{01}', \boldsymbol{X}_{02}')'$ denotes the design matrix for $\boldsymbol{\beta}_0$ in M_c;
- $\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_{01} & \boldsymbol{X}_1 & 0 \\ \boldsymbol{X}_{02} & 0 & \boldsymbol{X}_2 \end{pmatrix}$ denotes all the design matrices in M_c.
- $\tilde{\boldsymbol{X}} = (\boldsymbol{X}_0, diag(\boldsymbol{X}_1, \boldsymbol{X}_2))$, where $diag$ is a diagonal operator, denotes all the design matrices in M_c.

1.3.3 Overview of Chapter 2

With M_i and M_c in Section 1.3.1, first, we adopt Zellner g-prior and a classical mixtures of g-prior, Zellner-Siow (ZS) prior (Zellner and Siow, 1980), and examine the posterior distributions together with frequentist properties of the Bayesian estimator. Second, we theoretically compare posterior variances and frequentist properties of the Bayesian estimator from M_i and M_c. Third, we evaluate reasonability of M_c compared with golden standard, where data could be fully observed through extensive simulation studies. At last, we conduct simulation studies and real data analysis to offer an overall relative performance of M_i and M_c.

1.3.4 Overview of Chapter 3

This chapter considers a more flexible version of Zellner's g-prior and ZS prior. Specifically, the prior is applied to each parameter $\boldsymbol{\beta}_i$ independently rather than $\tilde{\boldsymbol{\beta}}_i$ or $\tilde{\boldsymbol{\beta}}$. This specification releases the full rank assumption on the whole design matrix in Chapter 2 and accommodates different shrinkage for different parameters. It also avoids the newly defined conditional information paradox. With independent version of Zellner's g-prior and ZS prior, we formally investigate posterior distributions and frequentist properties of the Bayesian estimator, and compare their performances under M_i and M_c. To enhance the frequentist properties of the Bayesian estimator, we further incorporate the effective sample size (TESS) (Berger et al., 2014) in the prior. We close this chapter by extensive simulation studies and one real data example to evaluate the relative performance of the Bayesian estimator from the data combining perspective.

1.3.5 Overview of Chapter 4

M_i and M_c provide a fundamental framework of data combining with original data. We may need to model more sophisticated issues in practice. For example, the variability in different data sources may differ, and random effects may exist. It is also common to have counts as outcomes. To demonstrate and explore the potential of our data combining framework, we focus on an application on the prevalence of female breast cancer using data from the Missouri Cancer Registry and 2016 Missouri County-level Study. We incorporate spatial effects as random effects and investigate data combining strategies under various assumptions. The corresponding data analyses are carried out in both linear model and generalized linear model. We conclude with comments on their relative performances.

Chapter 2

Standard Mixtures of G-Priors

2.1 Introduction

Over the last century, the linear model has been by far the most popular and appealing statistical model in both the frequentist and Bayesian literature. The use of Bayesian approaches, which combine information from the data likelihood with reasonable prior distributions placed on the unknown model parameters to carry out inference, is a valuable direction taken to broaden the linear model. Consider a linear regression problem $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$, and suppose there is some information about the regression coefficients and little information about the σ^2. A normal conjugate prior is naturally of interest for computational tractability, and one popular option is the g-prior deduced by Goel and Zellner (1986) with $\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \sigma^2 g(\boldsymbol{X}'\boldsymbol{X})^{-1})$, where g is a scale parameter and can be considered either known or unknown.

Zellner's g-prior and Zellner-Siow prior have received tremendous attention in model selection and many variants have been developed. One of the most influential works is Liang et al. (2008). They proposed hyper-g-prior and examined mixtures of g-prior, Zellner's g-prior and empirical g-prior by evaluating some desirable theoretical properties in model selection such as model selection consistency and prediction consistency. They proved that these properties are reserved only when a random g is adopted. Maruyama and George (2011) investigated Bayes factor by generalizing Zellner's g-prior in high-dimensional setting and utilized a Pearson Type VI distribution for the hyperparameter g including Cui and George (2008) and Liang et al. (2008) as special cases. Zhang et al. (2016) extended Zellner's g-prior to a two-component g-prior by introducing a tuning parameter and enabled a different estimate for each coefficient. Li and Clyde (2018) further explored mixtures of g-prior in the generalized linear model and proposed confluent hypergeometric prior on g to ensure the asymptotic consistency in terms of model selection and Bayesian model average (BMA) estimation.

However, g-prior has been less studied from the estimation perspective since Goel and Zellner (1986). Agliari and Parisetti (1988) derived A-g reference informative prior to incorporate prior knowledge of different independent variables by replacing $\boldsymbol{X}$ with $\boldsymbol{AX}$ in g-prior, where $\boldsymbol{A}$ is a diagonal matrix with non-negative elements. Sparks et al. (2015) examined g-priors, including the empirical model in George (2000), the hyper g-prior in Liang et al. (2008) and the classical Zellner-Siow prior, and provided the corresponding necessary and sufficient conditions for posterior consistency under certain defined sequences. Beyond the parametric framework model, in an analogy to Zellner's g-prior, Zhang et al. (2009) introduced Silverman's g-prior (Silverman, 1985) to capture the regularization in kernel supervised learning methods and studied the posterior consistency under the corresponding Bayesian model. In addition, it is well noticed that g-prior is connected to the ridge regression, whose penalty term is L_2 norm, and offers shrinkage estimation for coefficients compared with least squares estimates. Shrinkage estimation generally offers many good properties such as smaller sampling variance and mean squared error compared with the least squares estimate. Therefore, it is interesting to consider the classic shrinkage prior (Berger et al., 2005) and evaluate its behaviors in our data combining context.

This chapter mainly researches on the classical Zellner's g-prior, Zellner-Siow (ZS) prior and shrinkage prior from the estimation perspective, with a particular emphasis on their relative performances in M_i and M_c. The remainder of this chapter is organized as follows. In Section 2.2, we study Zellner's g-prior in two cases according to whether σ^2 is known or unknown. The sufficient and necessary conditions are established for smaller posterior variances in M_c compared with M_i. In Section 2.3, we focus on ZS prior and performed analyses on posterior variance and frequentist properties of the Bayesian estimator. In Section 2.3.3, we conduct extensive simulation studies including sensitivity analysis of M_c compared to the golden data-combining standard and comparison between M_i and M_c based on some frequentist properties and posterior variances. Finally, we discuss some key findings as well as some issues related to g-prior and potential future works.

2.2 Conventional G-priors

Conventional g-prior refers to the situation, where the scale parameter g is known, and this section focuses on its application in M_i and M_c. We begin with the standard g-

prior by Goel and Zellner (1986) in Section 2.2.1, where both the scale parameter and σ^2 are known. Then, with the scale parameter fixed, we consider σ^2 to be unknown for a practical consideration in 2.2.2.

2.2.1 Case 1. Known (σ^2, g)

Priors, posterior distributions and some frequentist properties of regression coefficients for M_i and M_c are given in Facts 2.1 and 2.2, respectively.

Fact 2.1. *For M_i in (1.1), assume the joint conventional g prior for $\tilde{\boldsymbol{\beta}}_i$ is:*

$$\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_i \sim N_{p_I}(\mathbf{0}, \sigma^2 g_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}), i = 1, 2. \tag{2.1}$$

(a) The posterior distribution for $(\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i)$ is $N_{p_I}(\tilde{\boldsymbol{\beta}}_i^B, \boldsymbol{\Sigma}_i^B)$, where

$$\tilde{\boldsymbol{\beta}}_i^B = \frac{g_i}{1+g_i}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i\boldsymbol{y}_i \text{ and } \boldsymbol{\Sigma}_i^B = \frac{g_i\sigma^2}{1+g_i}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}. \tag{2.2}$$

(b) The posterior variances for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ are:

$$\begin{aligned} VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) &= \frac{g_i\sigma^2}{1+g_i}[\boldsymbol{X}_{0i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}]^{-1}, \\ VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) &= \frac{g_i\sigma^2}{1+g_i}\{(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} \\ &\quad + (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{X}_{0i}[\boldsymbol{X}_{0i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}]^{-1}\boldsymbol{X}_{0i}'\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\}, \end{aligned}$$

where $\boldsymbol{P}_i = \boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'$. Notice that $[\boldsymbol{X}_{0i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}]^{-1}$ exists since $(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}$ exists.

(c) *The frequentist distribution for the Bayesian estimator* $\tilde{\boldsymbol{\beta}}_i^B$ *is* $N_{p_I}(\boldsymbol{m}_i, \boldsymbol{V}_i)$*, where*

$$\boldsymbol{m}_i = \frac{g_i}{1+g_i}\tilde{\boldsymbol{\beta}}_i \text{ and } \boldsymbol{V}_i = (\frac{g_i}{1+g_i})^2\sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}. \tag{2.3}$$

Fact 2.2. *For* M_c *in* (1.2)*, assume the dependent conventional g-prior for* $\tilde{\boldsymbol{\beta}}$*,*

$$\tilde{\boldsymbol{\beta}}|\sigma^2, g \sim N_{p_T}(\boldsymbol{0}, \sigma^2 g(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}). \tag{2.4}$$

(a) *The posterior distribution for* $\tilde{\boldsymbol{\beta}}$ *is* $N_{p_T}(\tilde{\boldsymbol{\beta}}^B, \boldsymbol{\Sigma}^B)$*, where*

$$\tilde{\boldsymbol{\beta}}^B = \frac{g}{1+g}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y} \text{ and } \boldsymbol{\Sigma}^B = \frac{g\sigma^2}{1+g}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}. \tag{2.5}$$

(b) *The posterior variance for* $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ *and* $\boldsymbol{\beta}_2$ *are:*

$$\begin{aligned}
VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g\sigma^2}{1+g}\{\boldsymbol{X}_{01}'(\boldsymbol{I}_{n_1} - \boldsymbol{P}_1)\boldsymbol{X}_{01} + \boldsymbol{X}_{02}'(\boldsymbol{I}_{n_2} - \boldsymbol{P}_2)\boldsymbol{X}_{02}\}^{-1},\\
VAR(\boldsymbol{\beta}_1|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g\sigma^2}{1+g}\{(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + (\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_{01}[\boldsymbol{X}_{01}'(\boldsymbol{I}_{n_1} - \boldsymbol{P}_1)\boldsymbol{X}_{01}\\
&+ \boldsymbol{X}_{02}'(\boldsymbol{I}_{n_2} - \boldsymbol{P}_2)\boldsymbol{X}_{02}]^{-1}\boldsymbol{X}_{01}'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\},\\
VAR(\boldsymbol{\beta}_2|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g\sigma^2}{1+g}\{(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1} + (\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{X}_{02}[\boldsymbol{X}_{01}'(\boldsymbol{I}_{n_1} - \boldsymbol{P}_1)\boldsymbol{X}_{01}\\
&+ \boldsymbol{X}_{02}'(\boldsymbol{I}_{n_2} - \boldsymbol{P}_2)\boldsymbol{X}_{02}]^{-1}\boldsymbol{X}_{02}'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\}.
\end{aligned}$$

(c) *The frequentist distribution for the Bayesian estimator* $\tilde{\boldsymbol{\beta}}^B$ *is* $N_{p_T}(\boldsymbol{m}, \boldsymbol{V})$*, where*

$$\boldsymbol{m} = \frac{g}{1+g}\tilde{\boldsymbol{\beta}} \text{ and } \boldsymbol{V} = (\frac{g}{1+g})^2\sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}.$$

Let M_i and M_c indicate the estimates obtained from Source i and combined data.

Theorem 2.1. *Consider M_i in (1.1) with conventional g-prior in (2.1), M_c in (1.2) with conventional g-prior in (2.4), and $\boldsymbol{A}_i = \boldsymbol{X}'_{0i}(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}$, $\boldsymbol{B}_j = (\boldsymbol{I}_{n_j} - \boldsymbol{P}_j)\boldsymbol{X}_{0j}$, $\boldsymbol{Q}_i = (\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}\boldsymbol{X}'_i\boldsymbol{X}_{0i}$, $\boldsymbol{M}_i = [(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1} + \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{Q}'_i] \in \mathbb{R}^{p_i \times p_i}$, $\boldsymbol{N} = \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{B}'_j[\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}'_j]^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{Q}'_i \in \mathbb{R}^{p_i \times p_i}$, where i or j indicates data from Source (i) or (j), and $i, j = 1, 2$ with $i + j = 3$.*

(a) The comparison of posterior variances in M_i and M_c for $\boldsymbol{\beta}_0$ is:

$$VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

if and only if

$$1 - \frac{g_i(1+g)}{g(1+g_i)} \leq \frac{\lambda_1}{1+\lambda_1}, \tag{2.6}$$

where λ_1 is the smallest eigenvalue of $\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}'_j$. Since $\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}'_j$ depends on the rank of $\boldsymbol{B}_j$, $\lambda_1 = 0$ if and only if $\boldsymbol{B}_j$ is not of full column rank.

(b) The comparison of posterior variances in M_i and M_c for $\boldsymbol{\beta}_i$ is:

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|, \sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

if and only if

$$1 - \frac{g_i(1+g)}{g(1+g_i)} \leq \lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}}\boldsymbol{N}\boldsymbol{M}_i^{-\frac{1}{2}}) \in [0, 1). \tag{2.7}$$

Proof. See Appendix A.1.1. □

Theorem 2.1 gives the necessary and sufficient condition where M_c offers a smaller posterior variance. In fact, the conditions in Theorem 2.1 can be easily verified and achieved. We would present several special cases for a demonstrative purpose according to different choices of g_i or g as well as the design matrix.

Example 1: Since we assume g_i and g are known and need to be chosen, with no further information available, we may set $g_i = g = c$ for a non-informative purpose, where c is a relatively large number. According to conditions in (2.6) and (2.7), in such setting, M_c generates a smaller posterior variance for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$. As an extreme, if we let $c \to \infty$, $\tilde{\boldsymbol{\beta}}_i^B$ and $\tilde{\boldsymbol{\beta}}^B$ reduce to the least squares estimate of $\tilde{\boldsymbol{\beta}}_i$ and $\tilde{\boldsymbol{\beta}}$. This theorem shows that, for least squares estimates, combining the data is always beneficial for providing estimates with better precision.

Example 2: When $\boldsymbol{X}_i'\boldsymbol{X}_{0i} = \boldsymbol{0}_{p_i \times p_0}$ or $p_i > p_0, \boldsymbol{M}_i$ and $\boldsymbol{N}$ are not of full rank, $\lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}) = 0$,

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0,$$

if and only if

$$g \leq g_i.$$

This indicates that when the design matrix for common and specific regression coefficients are orthogonal or when the dimension of specific regression coefficient is larger than that for common, $g_i \geq g$ is a sufficient and necessary condition to achieve a smaller posterior variance in M_c.

Example 3: At last, we consider $\boldsymbol{X}_{0i} = \boldsymbol{1}_{n_i}$, where we only allow two data sources to share the same intercept. Since $\boldsymbol{1}_{n_i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{1}_{n_i} = n_i - s_i$, where s_i is the summation of all elements in $\boldsymbol{P}_i$, we have the following.

1. For M_i,

$$
\begin{aligned}
VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) =& \frac{g_i}{1+g_i}\sigma^2(n_i - s_i)^{-1}, \\
VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) =& \frac{g_i}{1+g_i}\sigma^2[(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} + \\
& (n_i - s_i)^{-1}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{J}_{n_i}\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}].
\end{aligned}
$$

   For M_c,

$$
\begin{aligned}
VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g}{1+g}\sigma^2(n_1 - s_1 + n_2 - s_2)^{-1}, \\
VAR(\boldsymbol{\beta}_1|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g}{1+g}\sigma^2[(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + \\
& (n_1 - s_1 + n_2 - s_2)^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{J}_{n_1}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}], \\
VAR(\boldsymbol{\beta}_2|\sigma^2, g, \boldsymbol{y}, M_c) =& \frac{g}{1+g}\sigma^2[(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1} + \\
& (n_1 - s_1 + n_2 - s_2)^{-1}(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{J}_{n_2}\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}].
\end{aligned}
$$

2. (a) For common regression coefficients,

$$VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

   if and only if

$$\frac{g_i(1+g)}{g(1+g_i)} \geq \frac{n_i - s_i}{n_1 - s_1 + n_2 - s_2}.$$

   (b) For specific regression coefficients, when $\boldsymbol{X}_i'\boldsymbol{1}_{n_i} = \boldsymbol{0}$ or $p_i > p_0 = 1$,

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

is equivalent to

$$g \leq g_i.$$

When $\boldsymbol{X}_i'\mathbf{1}_{n_i} \neq \mathbf{0}$ and $p_i = p_0 = 1$, $\boldsymbol{X}_i$ will reduce to a vector $\boldsymbol{x}_i$ of dimension n_i and $\boldsymbol{Q}_i$ will reduce to a real number q_i. Here,

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

is equivalent to

$$1 - \frac{g_i(1+g)}{g(1+g_i)} \leq \frac{[(n_i - s_i)^{-1} - (n_1 - s_1 + n_2 - s_2)^{-1}]q_i^2}{(\boldsymbol{x}_i'\boldsymbol{x}_i)^{-1} + (n_i - s_i)^{-1}q_i^2} \in (0, 1).$$

Remark 2.1. *According to Facts 2.1-2.2, given g_i and g, the frequentist variance and posterior variance for $\tilde{\boldsymbol{\beta}}_i^B$ and $\tilde{\boldsymbol{\beta}}^B$ are related through*

$$\boldsymbol{V}_i = \frac{g_i}{1+g_i}\boldsymbol{\Sigma}_i^B \text{ and } \boldsymbol{V} = \frac{g}{1+g}\boldsymbol{\Sigma}^B.$$

Hence, the comparison of posterior variance in Theorem 2.1 can be applied to frequentist variances and we only need to replace $g_i(1+g)/[g(g_i+1)]$ in Theorem 2.1 with $g_i^2(1+g)^2/[g^2(g_i+1)^2]$.

Remark 2.2. *If $\boldsymbol{\beta}_0, \boldsymbol{\beta}_i$ is estimated jointly, combining data have better estimates in terms of a smaller covariance matrix.*

Proof. See Appendix A.1.2. □

2.2.2 Unknown σ^2 and Known g

Case 1 studies the ideal situation where both σ^2 and g are known, which is unlikely in practice. Hence, with the conventional g-priors for regression coefficients in equations (2.1) and (2.4), we further assume σ^2 is unknown and a Jeffrey prior is utilized:

$$\pi(\sigma^2) \propto 1/\sigma^2. \tag{2.8}$$

Then, posterior distributions for $\boldsymbol{\beta}_i$ and frequentist properties of Bayesian estimators for M_i and M_c are given in Facts 2.3 and 2.4, respectively.

Fact 2.3. *For M_i in (1.1), with priors in (2.1) and (2.8), we have:*

(a) The posterior distribution of $\tilde{\boldsymbol{\beta}}_i$ is t-distribution with $t_{n_i}(\tilde{\boldsymbol{\beta}}_i^B, \boldsymbol{\Sigma}_i^B)$, where

$$\tilde{\boldsymbol{\beta}}_i^B = \frac{g_i}{g_i+1}(\tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i{}'\boldsymbol{y}_i \text{ and } \boldsymbol{\Sigma}_i^B = \frac{\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i}{n_i(\frac{1}{g_i}+1)}(\tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i)^{-1}, \tag{2.9}$$

where $\tilde{\boldsymbol{P}}_i = \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i{}'$ and the posterior covariance is $n_i\boldsymbol{\Sigma}_i^B/(n_i-2)$.

(b) The frequentist distribution for $\tilde{\boldsymbol{\beta}}_i^B$ is $N_{p_I}(\boldsymbol{m}_i, \boldsymbol{V}_i)$, where

$$\boldsymbol{m}_i = \frac{g_i}{1+g_i}\tilde{\boldsymbol{\beta}}_i \text{ and } \boldsymbol{V}_i = \left(\frac{g_i\sigma}{1+g_i}\right)^2(\tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i)^{-1}.$$

Proof. See Appendix A.1.3. □

Fact 2.4. *For M_c in (1.2), with priors in (2.4) and (2.8), we have:*

(a) The posterior distribution of $\tilde{\boldsymbol{\beta}}$ is t-distribution with $t_{n_T}(\tilde{\boldsymbol{\beta}}^B, \boldsymbol{\Sigma}^B)$, where

$$\tilde{\boldsymbol{\beta}}^B = \frac{g}{g+1}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y} \text{ and } \boldsymbol{\Sigma}^B = \frac{\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \tilde{\boldsymbol{P}})\boldsymbol{y}}{n_T(\frac{1}{g}+1)}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}. \tag{2.10}$$

where $\tilde{\boldsymbol{P}} = \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'$ *and the posterior covariance is* $n_T\boldsymbol{\Sigma}^B/(n_T-2)$.

(b) *The frequentist distribution for* $\tilde{\boldsymbol{\beta}}^B$ *is* $N_{p_T}(\boldsymbol{m}, \boldsymbol{V})$, *where*

$$\boldsymbol{m} = \frac{g}{1+g}\tilde{\boldsymbol{\beta}} \text{ and } \boldsymbol{V} = \left(\frac{g\sigma}{1+g}\right)^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}.$$

With the same notations as Theorem 2.1, the comparison of posterior variances for regression coefficients between M_i and M_c is as below.

Theorem 2.2. *For* M_i *in* (1.1) *with priors* (2.1) *and* (2.8), *and* M_c *in* (1.2) *with priors* (2.4) *and* (2.8), *let*

$$a_i = \frac{\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_{\tilde{\boldsymbol{X}}_i})\boldsymbol{y}_i}{(n_i-2)(g_i^{-1}+1)}, \text{ and } a = \frac{\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \boldsymbol{P}_{\tilde{\boldsymbol{X}}})\boldsymbol{y}}{(n_T-2)(g^{-1}+1)}.$$

1. *For the shared regression coefficients* $\boldsymbol{\beta}_0$,

   (a) *If* $1 - \frac{a_i}{a} < 0$,

   $$VAR(\boldsymbol{\beta}_0|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|g_i, \boldsymbol{y}_i, M_i) \leq 0$$

   holds all the time.

   (b) *If* $1 - \frac{a_i}{a} \geq 0$,

   $$VAR(\boldsymbol{\beta}_0|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|g_i, \boldsymbol{y}_i, M_i) \leq 0$$

   if and only if

   $$1 - \frac{a_i}{a} \leq \lambda_{min}\{\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}\} \in (0,1).$$

2. For the specific regression coefficients $\boldsymbol{\beta}_i$*,*

(a) If $1 - \frac{a_i}{a} < 0$*,*

$$VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) \leq 0$$

holds all the time.

(b) If $1 - \frac{a_i}{a} \geq 0, \boldsymbol{X}_i'\boldsymbol{X}_{0i} \neq \boldsymbol{0}$ *and* $p_i < p_0$*,*

$$VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) \leq 0$$

if and only if

$$1 - \frac{a_i}{a} \leq \lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}}\boldsymbol{N}\boldsymbol{M}_i^{-\frac{1}{2}}) \in (0, 1).$$

Theorem 2.2 reveals that the relative magnitude of posterior variances for $\boldsymbol{\beta}_0$ or $\boldsymbol{\beta}_i$ in M_i and M_c depends highly on the ratio of a_i and a together with the design matrix. The ratio of a_i and a is related to the SSE, sample size, and the value of g_i or g in M_i and M_c. Although it is evident that $\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \tilde{\boldsymbol{P}})\boldsymbol{y} \geq \boldsymbol{y}_1'(\boldsymbol{I}_{n_1} - \tilde{\boldsymbol{P}}_1)\boldsymbol{y}_1 + \boldsymbol{y}_2'(\boldsymbol{I}_{n_2} - \tilde{\boldsymbol{P}}_2)\boldsymbol{y}_2$ or $\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \tilde{\boldsymbol{P}})\boldsymbol{y} \geq \boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i$ (equivalently, SSE is larger with combined data), the relationship between $\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \tilde{\boldsymbol{P}})\boldsymbol{y}/(n_T - 2)$ and $\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i/(n_i - 2)$ remains less clear. A deeper discussion of their connection is one of our future directions. In addition, compared with Case 2.2.1, we can see that the Bayesian estimators in both models have the same form and therefore their frequentist distributions are the same. Consequently, results in Remark 2.1 can be extended directly to this case. At last, our theorem indicates that it is easy to verify these conditions and therefore guide the choice of g_i or g in terms of posterior variances, given the observations and design matrices. However, the justification of such choices still requires further investigation. Alternatively, a prior on g could be specified, which enables a formal Bayesian procedure, and such specification has been proved to receive many benefits in the context of model selection. A detailed discussion is offered about this option in the next subsection.

2.3 Zellner-Siow Prior

For the conventional g-prior, where g is a fixed value, the data-driven calibration of g has been discussed to improve its performance in model selection including Clyde and George (2000); George (2000); Kass and Wasserman (1995) and many others. As an alternative, Zellner and Siow (1980) proposed to introduce a prior on g to enable a fully Bayesian analysis, which has been referred to as Zellner-Siow prior. It was limited at the time due to computational challenges in integrating g. As the development of computational tools, the benefits of Zellner-Siow prior have been recognized and many variants have been developed. For example, in linear regressions, Liang et al. (2008) demonstrated that a fixed choice of g subjects to the information paradox and proposed hyper-g prior as a solution, where a distribution for g is used. Maruyama and George (2011) proposed a generalization that allows coefficient dimensions to be greater than the number of observations. Moreover, Li and Clyde (2018); Wu et al. (2016) extended g-prior to the generalized linear mixed model.

In this section, we focus on the classical Zeller-Siow prior and study its relative performance in M_i and M_c. We first present the priors and posteriors distributions. Then, we aim at the Bayesian estimators and posterior variances analyses from two perspectives. One perspective is the Laplace approximation, and the other is the behavior of posterior variance in a special case.

2.3.1 Posterior Distribution and Computation

For M_i, the priors for regression coefficients $\tilde{\boldsymbol{\beta}}_i$ and σ^2 are specified as

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_i|\sigma^2 &\propto \left(1+\tilde{\boldsymbol{\beta}}_i'\frac{\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i}{n\sigma^2}\tilde{\boldsymbol{\beta}}_i\right)^{-\frac{p_I+1}{2}}, \\ \pi(\sigma^2) &\propto \frac{1}{\sigma^2},\end{aligned}$$

which is a multivariate Cauchy distribution with the precision being unit Fisher information matrix. One benefit of this specification is that it is equivalent to the following hierarchical structure:

$$\tilde{\boldsymbol{\beta}}_i|g_i,\sigma^2 \sim N_{p_I}(0, g_i\sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}), g_i \sim IG(1/2, n_i/2), \pi(\sigma^2) \propto 1/\sigma^2, \tag{2.11}$$

which enables a faster computation.

Similarly, for M_c, priors are specified as:

$$\tilde{\boldsymbol{\beta}}|g,\sigma^2 \sim N_{p_T}(0, g\sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}), g \sim IG(1/2, n_T/2), \pi(\sigma^2) \sim 1/\sigma^2. \tag{2.12}$$

Then, priors in (2.11) and (2.12) are referred to as Zellner-Siow (ZS) prior.

Fact 2.5. *Given the ZS prior and model, we could obtain the posterior mean and variance through the law of total expectation and law of total variance.*

(a) For M_i, with priors in (2.11), the posterior mean and variance for $\tilde{\boldsymbol{\beta}}_i$ is

$$E(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, M_i) = E\left(\frac{g_i}{1+g_i}|\boldsymbol{y}_i\right)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i,$$

$$VAR\left(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, M_i\right) = E\left(\frac{g_i}{1+g_i}\frac{\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \frac{g_i}{1+g_i}\tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i}{n_i - 2}|\boldsymbol{y}_i\right)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}$$
$$+ VAR\left(\frac{g_i}{1+g_i}|\boldsymbol{y}_i\right)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i\boldsymbol{y}_i'\tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}.$$

(b) Similarly, for M_c with priors (2.12), the posterior mean and variance for $\tilde{\boldsymbol{\beta}}$ is

$$E(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}, M_c) = E\left(\frac{g}{1+g}|\boldsymbol{y}\right)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y},$$

$$VAR(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}, M_c) = E\left(\frac{g}{1+g}\frac{\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \frac{g}{1+g}\tilde{\boldsymbol{P}})\boldsymbol{y}}{n_T - 2}|\boldsymbol{y}\right)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}$$
$$+ VAR\left(\frac{g}{1+g}|\boldsymbol{y}\right)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}\boldsymbol{y}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}.$$

Compared with g-prior in Sections 2.2.1 and 2.2.2, marginalizing over g in the Bayesian estimator in Fact 2.5 allows a data-adaptive shrinkage of the least squares estimator. Since a tractable form of the marginal distribution for $\tilde{\boldsymbol{\beta}}_i$ is not available, the following posterior distributions can be utilized for computation in M_i.

$$\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_i, \boldsymbol{y}_i \sim N_{p_I}\left(\frac{g_i}{1+g_i}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i, \frac{g_i\sigma^2}{1+g_i}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\right); \tag{2.13}$$

$$\sigma^2|g_i, \boldsymbol{y}_i \sim IG\left(\frac{n_i}{2}, \frac{1}{2}\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \frac{g_i}{1+g_i}\tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i\right); \tag{2.14}$$

$$\pi(g_i|\boldsymbol{y}_i) \propto (1+g_i)^{-\frac{p_I}{2}} g_i^{-\frac{3}{2}} exp(-\frac{n_i}{2g_i})[\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \frac{g_i}{1+g_i}\tilde{\boldsymbol{P}}_i)\boldsymbol{y}_i]^{-\frac{n_i}{2}}, \tag{2.15}$$

Similarly, for M_c, the corresponding posterior distributions are

$$\tilde{\boldsymbol{\beta}}|\sigma^2, g, \boldsymbol{y} \sim N_{p_T}\left(\frac{g}{1+g}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}, \frac{g\sigma^2}{1+g}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\right); \tag{2.16}$$

$$\sigma^2|g, \boldsymbol{y} \sim IG\left(\frac{n_T}{2}, \frac{1}{2}\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \frac{g}{1+g}\tilde{\boldsymbol{P}})\boldsymbol{y}\right); \tag{2.17}$$

$$\pi(g|\boldsymbol{y}) \propto (1+g)^{-\frac{p_T}{2}} g^{-\frac{3}{2}} exp(-\frac{n_T}{2g})[\boldsymbol{y}'(\boldsymbol{I}_{n_T} - \frac{g}{1+g}\tilde{\boldsymbol{P}})\boldsymbol{y}]^{-\frac{n_T}{2}}. \tag{2.18}$$

The introduction of a hyper parameter g facilitates the computation because the integration of the marginal distribution of g in (2.15) or (2.18) is only one-dimensional, which can be performed through standard integration techniques or approximation with reasonable accuracy.

2.3.2 Posterior Variance Analysis

This section takes a focused investigation on the posterior variances in Fact 2.5. To evaluate these quantities, we need to deal with the marginal distributions for g_i or g in (2.15) or (2.18). As these distributions are not standard, we take an approximation approach and consider the Laplace approximation. It is a popular method for approximating integrals and is a candidate for analyzing the posterior mean and variance of $g/(1+g)$. There are many formulations for Laplace approximation and two of which are discussed here. The first is the fully exponential Laplace approximation (Tierney and B.Kadane (1986)). The second is the regular Laplace approximation, which provides more insights in our situation.

Since the posterior means and variances for M_i and M_c have similar structures, out of simplicity, only in this part, unless otherwise mentioned, we omit subscripts in

Fact 2.5 and study following quantities:

$$\pi(g|\boldsymbol{y}) \propto (1+g)^{-\frac{p}{2}} g^{-\frac{3}{2}} exp(-\frac{n}{2g})[\boldsymbol{y}'(\boldsymbol{I} - \frac{g}{1+g}\tilde{\boldsymbol{P}})\boldsymbol{y}]^{-\frac{n}{2}}, \tag{2.19}$$

$$E(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}) = E\left(\frac{g}{1+g}|\boldsymbol{y}\right)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}, \tag{2.20}$$

$$VAR(\tilde{\boldsymbol{\beta}}|\boldsymbol{y}) = E\left(\frac{g}{1+g}\frac{\boldsymbol{y}'(\boldsymbol{I}_n - \frac{g}{1+g}\tilde{\boldsymbol{P}})\boldsymbol{y}}{n-2}|\boldsymbol{y}\right)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1} + VAR(\frac{g}{1+g}|\boldsymbol{y})(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}\boldsymbol{y}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}. \tag{2.21}$$

Here, $\tilde{\boldsymbol{\beta}} \in \mathbb{R}^p$ is a vector of regression coefficients, $\tilde{\boldsymbol{X}} \in \mathbb{R}^{n\times p}$ is the design matrix, $\tilde{\boldsymbol{P}}$ is its projection matrix, n is the sample size.

Option 1: Fully exponential Laplace approximation

Since the principal regularity condition for this method is the target function to be unimodal, we will verify that this condition holds in our case first. To start with, we investigate the mode of our target functions. Let $h(g|\boldsymbol{y})$ denote the kernel of $\pi(g|\boldsymbol{y})$ and $\tilde{R}^2 = \boldsymbol{y}'\tilde{\boldsymbol{P}}\boldsymbol{y}/\boldsymbol{y}'\boldsymbol{y} \in [0,1]$, and we have the following:

$$\pi(g|\boldsymbol{y}) \propto h(g|\boldsymbol{y}) = (1+g)^{\frac{n-p}{2}} g^{-\frac{3}{2}} exp(-\frac{n}{2g})\left[1+g(1-\tilde{R}^2)\right]^{-\frac{n}{2}} \tag{2.22}$$

The posterior mean and variance can be calculated by the following quantity:

$$E\left(\frac{g^a}{(1+g)^a}|\boldsymbol{y}\right) = \int_0^{+\infty}\frac{g^a}{(1+g)^a}\pi(g|\boldsymbol{y})dg = \frac{\int_0^{+\infty}\frac{g^a}{(1+g)^a}h(g|\boldsymbol{y})dg}{\int_0^{+\infty}h(g|\boldsymbol{y})dg}. \tag{2.23}$$

If $a = 1$, (2.23) is the posterior mean for $g/(1+g)$. If $a = 2$, (2.23) is the second posterior moment for $g/(1+g)$. Furthermore, let $H_a(g|\boldsymbol{y}) = g^a h(g|\boldsymbol{y})/(1+g)^a, a =$

$0, 1, 2$, $L_a(g|\boldsymbol{y}) = Log(H_a(g|\boldsymbol{y}))$, and then:

$$L_a(g|\boldsymbol{y}) = (a - \frac{3}{2})log(g) + \frac{n-p-2a}{2}log(1+g) - \frac{n}{2g} - \frac{n}{2}log(1+g(1-\tilde{R}^2))$$

with its first derivative:

$$\frac{\partial L_a(g|\boldsymbol{y})}{\partial g} = (a - \frac{3}{2})\frac{1}{g} + \frac{n-p-2a}{2}\frac{1}{1+g} + \frac{n}{2g^2} - \frac{n}{2}\frac{1-\tilde{R}^2}{1+g(1-\tilde{R}^2)},$$

and the second derivative:

$$\frac{\partial^2 L_a(g|\boldsymbol{y})}{\partial^2 g} = \frac{1}{2}\left[\frac{n(1-\tilde{R}^2)^2}{1+g(1-\tilde{R}^2)^2} - \frac{n-p-2a}{(1+g)^2} + \frac{3-2a}{g^2} - \frac{2n}{g^3}\right].$$

To find the mode of $L_a(g|\boldsymbol{y})$, let $\partial L_a(g|\boldsymbol{y})/\partial g = 0$, which is equivalent to find roots of the cubic equation:

$$-(p+3)(1-\tilde{R}^2)g^3 + [(2a-3)(2-\tilde{R}^2)+(n-p-2a)]g^2 + [(2a-3)+(2-\tilde{R}^2)n]g + n = 0.$$

Assume that g_1, g_2, g_3 are three roots of this cubic equation, generally, it has one real root and a pair of complex conjugate roots with

$$g_1g_2g_3 = \frac{n}{(p+3)(1-\tilde{R}^2)}, \ g_1g_2 + g_1g_3 + g_2g_3 = -\frac{(2a-3)+(2-\tilde{R}^2)n}{(p+3)(1-\tilde{R}^2)}.$$

If $a = 0$,

$$g_1g_2g_3 = \frac{n}{(p+3)(1-\tilde{R}^2)} > 0, g_1g_2 + g_1g_3 + g_2g_3 = -\frac{n(2-\tilde{R}^2)-3}{(p+3)(1-\tilde{R}^2)} < 0,$$

which suggests one positive root and two negative roots exist for $n \geq 3$, and hence $L_0(g|\boldsymbol{y})$ has unique positive modal in the domain of g. It is easy to verify that $g_1g_2g_3 > 0$ and $g_1g_2 + g_1g_3 + g_2g_3 < 0$ hold for $a = 1$ and $a = 2$. Then, the Laplace approximation for $E\left(g^a/(1+g)^a|\boldsymbol{y}\right)$ is:

$$E\left(\frac{g^a}{(1+g)^a}|\boldsymbol{y}\right) = \frac{\int_0^\infty exp(log(H_a(g|\boldsymbol{y})))dg}{\int_0^\infty exp(log(H_0(g|\boldsymbol{y})))dg} \approx \frac{\hat{\sigma}_{H_a}}{\hat{\sigma}_{H_0}} \frac{exp(H_a(\hat{g}_{H_a})|\boldsymbol{y})}{exp(H_0(\hat{g}_{H_0}|\boldsymbol{y}))}, \tag{2.24}$$

where $\hat{g}_{H_a}$ denotes the mode for $H_a(g|\boldsymbol{y})$ and $\hat{\sigma}_{H_a} = \left[\partial^2 L_a(g|\boldsymbol{y})/\partial^2 g|_{g=\hat{g}_{H_a}}\right]$. The benefit of fully exponential Laplace approximation is its improved accuracy of $\mathcal{O}(n^{-2})$ for both posterior mean and variance of $g^a/(1+g)^a$. The main idea is that, when $\hat{g}_{H_a}$ is large enough, $\hat{g}_{H_a}/(1+\hat{g}_{H_a})$ approaches to 1 despite of data. To find the mode of the target function, Monte Carlo method and finding roots for the related cubic equation are used and they yield consistent results, which indicate that $\hat{g}_{H_a}$ is far from 0 despite of data. One drawback is that the mode for denominator and numerator are different, although both of them are very large, which makes it hard to justify the overall performance of $E\left(g^a/(1+g)^a|\boldsymbol{y}\right)$ based on an explicit expression.

Option 2: Conventional Laplace approximation

Another way to explore the quantity $E\left(g^a/(1+g)^a|\boldsymbol{y}\right)$ is the conventional Laplace approximation, which has an accuracy of $\mathcal{O}(n^{-1})$ and is formulated as:

$$\int_a^b w(x)e^{Mq(x)} \approx \sqrt{\frac{2\pi}{M|q''(x_0)|}}w(x_0)e^{Mq(x_0)}, \text{ as } M \to \infty, \tag{2.25}$$

where $w(x)$ is positive and continuous, $q(x)$ is continuous, unimodal and twice differentiable, x_0 is a unique global maximum at x_0, and M is a large number. To accomodate (2.25), we first rewrite $h(g|\boldsymbol{y})$ in (2.22) as $h(g|\boldsymbol{y}) = h(\boldsymbol{y}|g)h_0(g)$, where $h(\boldsymbol{y}|g) = (1+g)^{\frac{n-p}{2}}[1+g(1-\tilde{R}^2)]^{-\frac{n}{2}}, h_0(g) = g^{-\frac{3}{2}}exp(-n/2g)$. Then, suppose $w(g) = h_a(g) = g^a h_0(g)/(1+g)^a, M = n, q(g) = log(h(\boldsymbol{y}|g))/n$, the approximation for the ath posterior moment is:

$$E\left(\frac{g^a}{(1+g)^a}|\boldsymbol{y}\right) = \frac{\int_0^\infty h_a(g)exp[log(h(\boldsymbol{y}|g))]dg}{\int_0^\infty h_0(g)exp[log(h(\boldsymbol{y}|g))]dg} \approx \frac{h_a(\hat{g})}{h_0(\hat{g})} = \frac{\hat{g}^a}{(1+\hat{g})^a}, \tag{2.26}$$

where $\hat{g}$ is the mode of $log(h(\boldsymbol{y}|g))/n$ and it is calculated by setting $\partial L(\boldsymbol{y}|g)/\partial g = 0$. Also, since g is restricted in a positive parameter space, $\hat{g}$ is:

$$\hat{g} = max\left\{\frac{\tilde{R}^2/p}{(1-\tilde{R}^2)/(n-p)} - 1, 0\right\}, \tag{2.27}$$

where $\hat{g}$ has the same form with the local empirical Bayes (Hansen and Yu, 2001).

With Laplace approximation in (2.26) and (2.27), the posterior mean and variance can be approximated by:

$$E\left(\frac{g}{1+g}|\boldsymbol{y}\right) \approx \frac{\hat{g}}{\hat{g}+1}, \; E\left(\frac{g^2}{(1+g)^2}|\boldsymbol{y}\right) \approx \left(\frac{\hat{g}}{\hat{g}+1}\right)^2, \tag{2.28}$$

$$VAR\left(\frac{g}{1+g}|\boldsymbol{y}\right) = E\left(\frac{g^2}{(1+g)^2}|\boldsymbol{y}\right) - E^2\left(\frac{g}{1+g}|\boldsymbol{y}\right) \approx 0. \tag{2.29}$$

Equations in (2.28) indicate that, if $\hat{g}$ is bounded considerably far away from 0, $E(g/(1+g)|\boldsymbol{y})$ and $E(g^2/(1+g)^2|\boldsymbol{y})$ is close to 1, and hence $VAR(g/(1+g)|\boldsymbol{y})$ is small, with 0 as an extreme. Meanwhile, the expression in (2.27) implies that one way to achieve $\hat{g} \to \infty$ is $\tilde{R}^2 \to 1$.

To formally access the relationship between $E(g^a/(1+g)^a|\boldsymbol{y})$ and $\tilde{R}^2$, we express the expectation as a function of $\tilde{R}^2$:

$$E\left(\frac{g}{1+g}|\boldsymbol{y}\right) = s_1(\tilde{R}^2), E\left(\frac{g^2}{(1+g)^2}|\boldsymbol{y}\right) = s_2(\tilde{R}^2). \tag{2.30}$$

While it is evident that $s_1(\tilde{R}^2), s_2(\tilde{R}^2) \leq 1$ and no closed form is available for a direct analysis, we can prove that $s_1(\tilde{R}^2)$ and $s_2(\tilde{R}^2)$ are bounded by more tractable functions $m_1(\tilde{R}^2)$, $m_2(\tilde{R}^2)$ as below:

$$s_1(\tilde{R}^2) \geq \frac{2}{4+p}\frac{{}_2F_1(\frac{n}{2}, 2; \frac{p}{2}+3; \tilde{R}^2)}{{}_2F_1(\frac{n}{2}, 1; \frac{p}{2}+2; \tilde{R}^2)} = m_1(\tilde{R}^2), \tag{2.31}$$

$$s_2(\tilde{R}^2) \geq \frac{8}{(4+p)(6+p)}\frac{{}_2F_1(\frac{n}{2}, 3; \frac{p}{2}+4; \tilde{R}^2)}{{}_2F_1(\frac{n}{2}, 1; \frac{p}{2}+2; \tilde{R}^2)} = m_2(\tilde{R}^2), \tag{2.32}$$

where ${}_2F_1(a; b, c; x)$ is the Gaussian hypergeometric function (See Appendix A.1.5 for detailed derivations) with:

$${}_2F_1(a, b; c; z) = \frac{1}{Beta(b, c-b)}\int_0^1 x^{b-1}(1-x)^{c-b-1}(1-xz)^{-a}dx, c > b > 0.$$

Although ${}_2F_1(a; b, c; x) \to \infty$ as $x \to 1$, it is less clear about the ratio of hypergeometric functions in (2.31) and (2.32). In Figure (2.1), we present some examples to visualize (2.31) and (2.32). These graphs indicate that $s_i(\tilde{R}^2)$ and $m_i(\tilde{R}^2)$ approach 1 as $\tilde{R}^2 \to 1$.

Theorem 2.3. *With* (2.19) - (2.21), *as* $\tilde{R}^2 \to 1$, $s_1(\tilde{R}^2) \to 1$ *and* $s_2(\tilde{R}^2) \to 1$.

Proof. See Appendix A.1.5. □

Then, $VAR(g/(1+g)|\boldsymbol{y}) = s_2(\tilde{R}^2) - s_1(\tilde{R}^2)^2 \to 0$ as $\tilde{R}^2 \to 1$. Theorem 2.3 also implies that, as $\tilde{R}^2 \to 1$, the Bayesian estimator is similar to the least squares

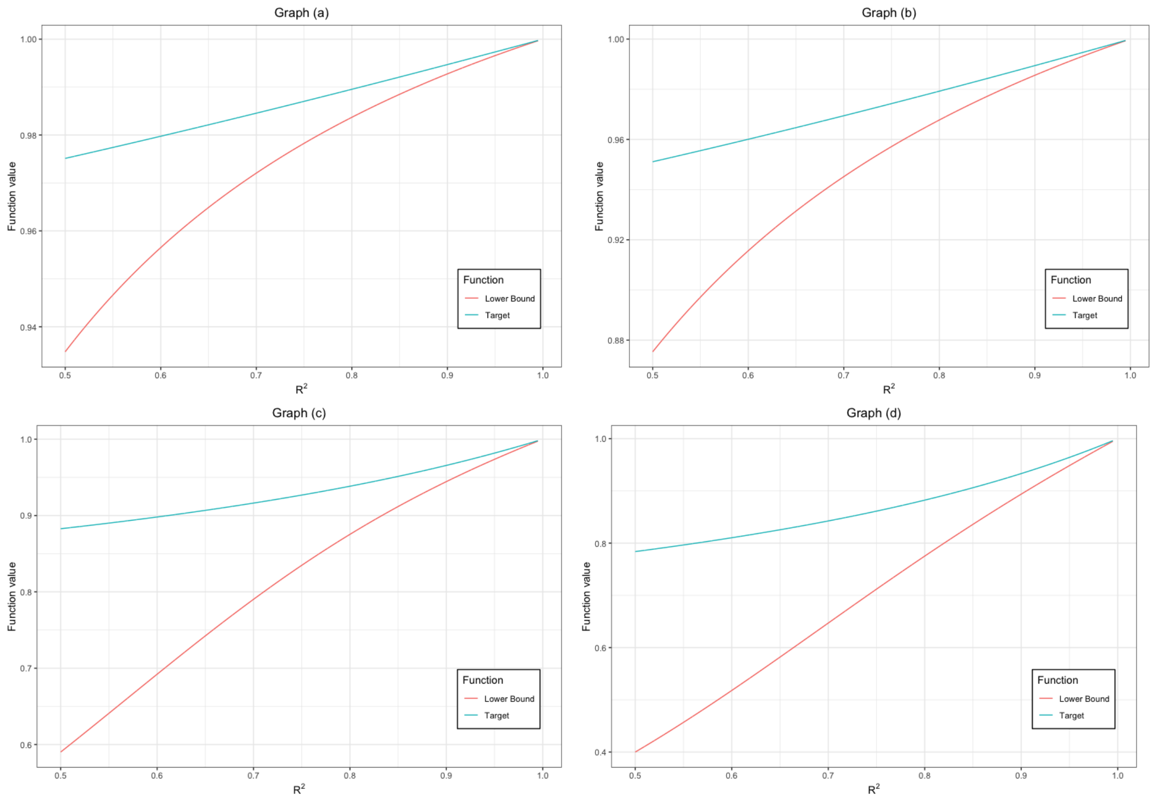

Figure 2.1: Relationship between $s_i(\tilde{R}^2)$, $m_i(\tilde{R}^2)$, $i = 1, 2$, and $\tilde{R}^2$. Graphs (a) and (b) are posterior mean and second moments for $n = 100, p = 4$, respectively. Graphs (c) and (d) are posterior mean and second moments for $n = 25, p = 4$, respectively.

estimate. As a result, the frequentist variances are smaller for M_c and the magnitude of such benefit mainly depends on the design matrices from these two data sources. For the marginal posterior variance of $\boldsymbol{\beta}$ in (2.21), with Theorem 2.3, we can roughly approximate $VAR(\boldsymbol{\beta}|\boldsymbol{y})$ by $\boldsymbol{L} = \boldsymbol{y}'(\boldsymbol{I}_n - \boldsymbol{P})\boldsymbol{y}/(n-2)(\boldsymbol{X}'\boldsymbol{X})^{-1}$, which is close to $(n-p)/(n-2)(\boldsymbol{X}'\boldsymbol{X})^{-1}$ on average. If we connect this idea with M_c and M_i, for one thing, the quantity of posterior variance primarily depends on its first term. For another, the average of $\boldsymbol{L}$ is $(n_T - p_T)/(n_T - 2)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}$ for M_c and $(n_i - p_I)/(n_i - 2)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}$ for M_i. However, the relative size of $(n_T - p_T)/(n_T - 2)$ and $(n_i - p_I)/(n_i - 2)$ remains inconclusive. As an addition, we conduct simulation studies in Section 2.4.2 to offer a view of the relative performance of posterior variances for M_i and M_c (See Tables 2.5 and 2.6). We found that, among 500 datasets, at least 96.7% has a smaller posterior variance in M_c for $\boldsymbol{\beta}_0$ and at least 55.7% has a smaller posterior variance in M_c for $\boldsymbol{\beta}_i$.

Here, Laplace approximation and Theorem 2.3 are adopted to evaluate $E(g^a/(1+g)^a|\boldsymbol{y})$ for a=1,2 from two perspectives so that an explanation can be offered for the relative performance of Bayesian estimators or posterior variances in M_i and M_c. We found that, when $\tilde{R}^2 \to 1$, the frequentist variance for the Bayesian estimator is more likely to be small in M_c while no clear pattern exists for posterior variance.

2.3.3 Extension

It is established that both Zellner's g-prior and ZS prior yield shrinkage estimation in terms of the Bayesian estimator. For a general linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{y}, \boldsymbol{\epsilon} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p, \boldsymbol{X} \in \mathbb{R}^{n\times p}$, we further consider a prior for $\boldsymbol{\beta}$ with the following hierarchical representation:

$$(\boldsymbol{\beta}|\lambda, \sigma^2) \sim N_p(\mathbf{0}, \lambda n \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}), \pi(\lambda) \propto \lambda^{-\frac{1}{2}} \exp\left(-\frac{1}{2\lambda}\right), \tag{2.33}$$

which is improper and referred to as the shrinkage prior (Berger et al., 2005). Its density function has a high peak around zero, which imposes shrinkage on the coefficients toward zero but not strictly exclude predictors. It is equivalent to:

$$\pi(\boldsymbol{\beta}) \propto \left(1 + \boldsymbol{\beta}' \frac{(\boldsymbol{X}'\boldsymbol{X})}{n\sigma^2} \boldsymbol{\beta}\right)^{-(p-1)/2}.$$

The marginal distribution for λ is:

$$\pi(\lambda|\boldsymbol{y}) \propto (1+\lambda)^{-\frac{p}{2}}\lambda^{-\frac{1}{2}}exp(-\frac{n}{2\lambda})[\boldsymbol{y}'(\boldsymbol{I}-\frac{\lambda}{1+\lambda}\tilde{\boldsymbol{P}})\boldsymbol{y}]^{-\frac{n}{2}}, \tag{2.34}$$

Compared with ZS prior, the only difference lies in the marginal distribution of λ. Therefore, all the analyses in Section 2.3.2 can be directly applied for the shrinkage prior. Similarly, we express $E(\lambda^a/(1+\lambda)^a|\boldsymbol{y})$ as functions of $\tilde{R}^2$:

$$E\left(\frac{\lambda}{1+\lambda}|\boldsymbol{y}\right) = k_1(\tilde{R}^2), E\left(\frac{\lambda^2}{(1+\lambda)^2}|\boldsymbol{y}\right) = k_2(\tilde{R}^2). \tag{2.35}$$

Theorem 2.4. *If $n \geq p+3$, when $\tilde{R}^2 \to 1$, we have $k_1(\tilde{R}^2) \to 1, k_2(\tilde{R}^2) \to 1$,*

We can see that Theorem 2.4 reaches the same conclusion as Theorem 2.3 in terms of the posterior distributions of g or λ. It is reasonable to expect ZS and shrinkage priors to behave similarly. This phenomenon can be observed in Figure 2.2. This figure shows that the shrinkage factors from the shrinkage prior and ZS prior follow the same trend and the difference between them narrows as the sample size increases. We will not go into detail about how Theorem 2.4 resonates with M_i and M_c for the sake of simplicity, because the logic of posterior analyses for ZS prior can be applied directly. More comparisons results for these two priors in terms of M_i and M_c can be found in the simulation studies in Section 2.4.2.

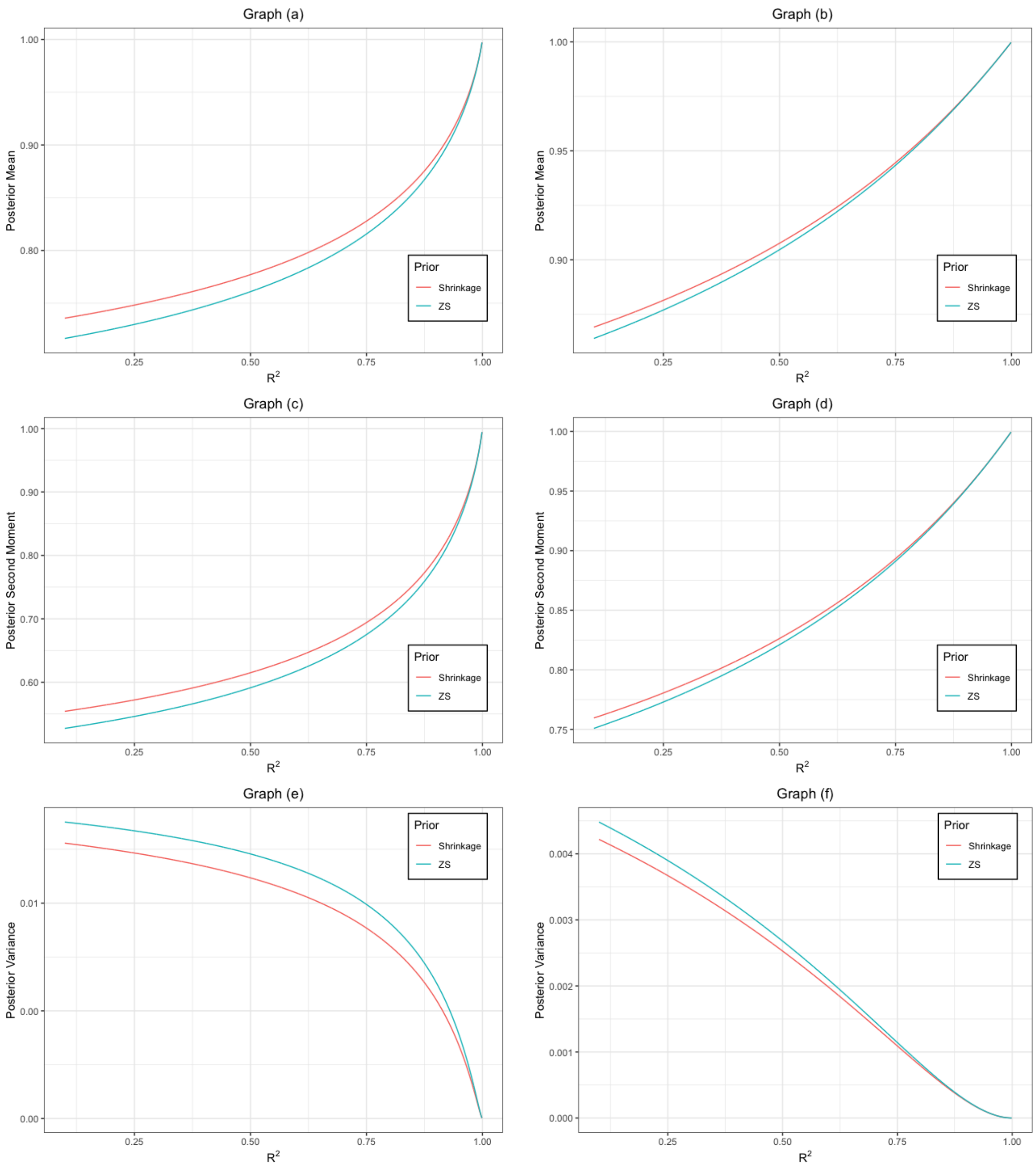

Figure 2.2: Illustrations of posterior distribution of g and λ for shrinkage and ZS prior. Graphs (a), (c) and (e) are posterior mean, second moments and variance with $n = 10, p = 5$, respectively. Graphs (b), (d) and (f) are those with $n = 30, p = 5$, respectively.

2.4 Numerical Analyses

2.4.1 Sampling Distribution

Note that ZS prior and shrinkage prior share similar form and therefore we unify their sampling distributions in one framework. Specifically, for M_i, the following distributions are applied to do the computation:

1. Sample $\sigma^2 | g_i, \boldsymbol{y}_i \sim IG\left(n_i/2,\ \boldsymbol{y}_i'\left[\boldsymbol{I}_{n_i} - g_i/(1+g_i)\tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}'\right]\boldsymbol{y}_i/2\right)$,
2. Sample $g_i | \sigma^2, \tilde{\boldsymbol{\beta}}_i, \boldsymbol{y}_i \sim IG\left(p_I/2 + l,\ \tilde{\boldsymbol{\beta}}_i{}'\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i/(2\sigma^2) + n_i/2\right)$,
3. Sample $\tilde{\boldsymbol{\beta}}_i | \sigma^2, g_i, \boldsymbol{y}_i \sim N_{p_I}\left(g_i/(1+g_i)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i,\ g_i\sigma^2/(1+g_i)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}\right)$.

Similarly, for M_c, the sampling distributions are described as below:

1. Sample $\sigma^2 | g, \boldsymbol{y} \sim IG\left(n_T/2,\ \boldsymbol{y}'\left[\boldsymbol{I}_{n_T} - g/(1+g)\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\right]\boldsymbol{y}/2\right)$,
2. Sample $g | \sigma^2, \tilde{\boldsymbol{\beta}}, \boldsymbol{y} \sim IG\left(p_T/2 + l,\ \tilde{\boldsymbol{\beta}}'\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}}/(2\sigma^2) + n_T/2\right)$,
3. Sample $\tilde{\boldsymbol{\beta}} | \sigma^2, g, \boldsymbol{y} \sim N_{p_T}\left(g/(1+g)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y},\ g\sigma^2/(1+g)(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\right)$.

Then, $l = 1/2$ and $l = 0$ correspond to ZS prior and shrinkage prior, respectively.

2.4.2 Model Comparison of M_i and M_c

We consider four sets of parameters for the regression coefficients with $p_0 = p_1 = p_2 = 3$ with respect to different sizes of coefficients $\tilde{\boldsymbol{\beta}}_i$ and error terms σ^2, where Sets 1-2 represent moderate to large coefficients while Sets 3-4 represent small coefficients.

- Set 1: $\boldsymbol{\beta}_0 = (1.1, 1.2, 1.8)'$, $\boldsymbol{\beta}_1 = (1.6, 1.2, 1.2)'$, $\boldsymbol{\beta}_2 = (1.3, 1.5, 1.7)'$, where all coefficients are large with $\sigma = 0.5$;

- Set 2: $\boldsymbol{\beta}_0 = (1.1, 1.2, 1.8)'$, $\boldsymbol{\beta}_1 = (1.6, 1.2, 1.2)'$, $\boldsymbol{\beta}_2 = (1.3, 1.5, 1.7)'$, where all coefficients are large with $\sigma = 0.1$;

- Set 3: $\boldsymbol{\beta}_0 = (0.5, 0.8, 0.4)'$, $\boldsymbol{\beta}_1 = (0.3, 0.6, 0.7)'$, $\boldsymbol{\beta}_2 = (0.5, 0.5, 0.9)'$, where all coefficients are small with with $\sigma = 0.5$;

- Set 4: $\boldsymbol{\beta}_0 = (0.5, 0.8, 0.4)'$, $\boldsymbol{\beta}_1 = (0.3, 0.6, 0.7)'$, $\boldsymbol{\beta}_2 = (0.5, 0.5, 0.9)'$, where all coefficients are small with with $\sigma = 0.1$.

For each set, we consider a small sample size $n_i = 10$ and moderate to large sample size $n_i = 20$. All design matrices are generated from the normal distribution $N(0, 1)$. To evaluate the relative performance of the Bayesian estimator, we collect its frequentist properties including its sampling variance, bias and MSE. For brevity, we report these quantities in group $\boldsymbol{\beta}_j$ rather than individual element β_{ij}. Each Bayesian estimator is computed through 20,000 samples with 10,000 burn-ins. The frequentist properties are calculated with 500 data sets. For priors, ZS prior and the shrinkage prior are considered.

Tables 2.1 - 2.4 present the $Bias$, VAR_F, and MSE for each grouped parameter. $Bias$ is the summation of absolute value of bias for each element in $\boldsymbol{\beta}_j, j = 0, 1, 2$ and describes the overall absolute difference between the expected value of the Bayesian estimator and its true value. VAR_F shows the overall sample variance of the Bayesian estimator for $\boldsymbol{\beta}_j$, which is the summation of diagonal elements of its sampling covariance matrix. Similarly, MSE is reported in groups. The bold number indicates M_c has a smaller value.

We have several main findings. First, despite the choices of sample size, random error, size of coefficients and prior options, M_c has equivalent or better performances in terms of VAR_F and MSE. Second, compared with the specific $\boldsymbol{\beta}_i$, M_c shows more reductions in VAR_F and MSE for common coefficients $\boldsymbol{\beta}_0$, which is reasonable since more information is available for $\boldsymbol{\beta}_0$ in M_c. Third, across all combinations, data combining is more advantageous with moderate to large coefficients, large σ and small sample size n_i. For example, with ZS prior, the reduction $VAR_F(\boldsymbol{\beta}_0)$ for $\sigma = 0.5, n_i = 10$ from M_1 to M_c is 0.1295 while for $\sigma = 0.1, n_i = 10$ is 0.0056. This is within our expectation since a larger sample size is needed to obtain more precise estimates when the variability in the data is large. Fourth, across all cells in terms of Models (M_1, M_2, M_C), parameters ($\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$), sample size and coefficients sets, ZS and shrinkage priors offer similar behaviors in terms of frequentist properties while one might outperform the other from different perspectives. For example, for $Bias$, shrinkage prior offers an equivalent or smaller bias in 46 out of 56 comparisons. In contrast, for VAR_F, ZS prior shows an equivalent or smaller frequentist variance in 35 out of 56 combinations. When it comes to MSE, ZS prior outperforms in 27 out of 56 combinations, which is close to half of the combinations.

Tables 2.5 and 2.6 present the percentages of cases where M_c outperforms M_1 and M_2 in terms of smaller posterior variances in 500 data sets, respectively. The key findings are summarized as below. To start with, the percentage is uniformly higher regarding the common $\boldsymbol{\beta}_0$ than the specific $\boldsymbol{\beta}_i$ across all settings. For example, it could be as high as 99.80% while the highest percentage for $\boldsymbol{\beta}_i$ is only 90.6%. This implies that M_c is more likely to offer a smaller posterior variance for $\boldsymbol{\beta}_0$ rather than $\boldsymbol{\beta}_i$. Second, shrinkage prior shows a uniformly lower percentage than ZS prior. This indicates that, if the research interest lies in the posterior variance, ZS prior has a better chance to offer a smaller posterior variance in M_c.

Table 2.1: Comparisons of M_1 and M_c in Sets 1 - 2

Prior	Design	Statistics	$n_1 = n_2$=10				$n_1 = n_2$=20			
			Set 1		Set 2		Set 1		Set 2	
			M_1	M_c	M_1	M_c	M_1	M_c	M_1	M_c
ZS	$\boldsymbol{\beta}_0$	$Bias$	0.1846	**0.1103**	0.0187	**0.0087**	0.0834	**0.0400**	0.0061	**0.0028**
		VAR_F	0.2142	**0.0847**	0.0091	**0.0035**	0.0580	**0.0228**	0.0024	**0.0009**
		MSE	0.2263	**0.0888**	0.0092	**0.0035**	0.0604	**0.0233**	0.0024	**0.0009**
	$\boldsymbol{\beta}_1$	$Bias$	0.1272	**0.0645**	0.0074	**0.0024**	0.0652	**0.0275**	0.0027	**0.0012**
		VAR_F	0.1413	**0.1371**	0.0059	**0.0057**	0.0714	**0.0644**	0.0029	**0.0026**
		MSE	0.1467	**0.1385**	0.0059	**0.0057**	0.0728	**0.0647**	0.0029	**0.0026**
Shrinkage	$\boldsymbol{\beta}_0$	$Bias$	0.0972	**0.0890**	0.0055	**0.0068**	0.0451	**0.0250**	0.0041	**0.0015**
		VAR_F	0.2096	**0.0831**	0.0087	**0.0034**	0.0577	**0.0242**	0.0023	**0.0010**
		MSE	0.2130	**0.0858**	0.0087	**0.0034**	0.0586	**0.0245**	0.0024	**0.0010**
	$\boldsymbol{\beta}_1$	$Bias$	0.0886	**0.0705**	0.0044	**0.0043**	0.0661	**0.0420**	0.0047	**0.0047**
		VAR_F	0.1529	**0.1479**	0.0063	**0.0061**	0.0691	**0.0635**	0.0028	**0.0026**
		MSE	0.1556	**0.1497**	0.0063	**0.0061**	0.0706	**0.0642**	0.0028	**0.0026**

Table 2.2: Comparisons of M_1 and M_c in Sets 3 - 4

Prior	Design	Statistics	$n_1 = n_2$=10				$n_1 = n_2$=20			
			Set 3		Set 4		Set 3		Set 4	
			M_1	M_c	M_1	M_c	M_1	M_c	M_1	M_c
ZS	$\boldsymbol{\beta}_0$	$Bias$	0.2421	**0.1837**	0.0373	**0.0175**	0.1200	**0.0702**	0.0096	**0.0047**
		VAR_F	0.1812	**0.0745**	0.0089	**0.0035**	0.0539	**0.0218**	0.0023	**0.0009**
		MSE	0.2028	**0.0866**	0.0095	**0.0036**	0.0590	**0.0235**	0.0024	**0.0009**
	$\boldsymbol{\beta}_1$	$Bias$	0.1804	**0.1331**	0.0247	**0.0081**	0.0980	**0.0550**	0.0059	**0.0022**
		VAR_F	0.1207	**0.1198**	0.0059	**0.0056**	0.0660	**0.0610**	0.0029	**0.0026**
		MSE	0.1331	**0.1267**	0.0061	**0.0056**	0.0696	**0.0621**	0.0029	**0.0026**
Shrinkage	$\boldsymbol{\beta}_0$	$Bias$	0.1623	**0.1561**	0.0180	**0.0140**	0.0777	**0.0520**	0.0034	**0.0021**
		VAR_F	0.1837	**0.0745**	0.0086	**0.0034**	0.0544	**0.0234**	0.0023	**0.0010**
		MSE	0.1930	**0.0832**	0.0087	**0.0035**	0.0564	**0.0244**	0.0023	**0.0010**
	$\boldsymbol{\beta}_1$	$Bias$	0.1465	**0.1317**	0.0156	**0.0100**	0.0939	**0.0660**	0.0073	**0.0061**
		VAR_F	0.1349	**0.1314**	0.0063	**0.0060**	0.0648	**0.0606**	0.0028	**0.0026**
		MSE	0.1435	**0.1383**	0.0064	**0.0061**	0.0683	**0.0625**	0.0028	**0.0026**

2.4.3 Sensitivity Analyses of M_c

While Section 2.4.2 gives us a big picture of relative performances of M_i and M_c, one natural question arises regarding the validity of data combining in M_c, and we may wonder how far M_c is from the golden standard. Here, the golden standard model,

Table 2.3: Comparisons of M_2 and M_c in Sets 1 - 2

Prior	Design	Statistics	$n_1 = n_2$=10				$n_1 = n_2$=20			
			Set 1		Set 2		Set 1		Set 2	
			M_2	M_c	M_2	M_c	M_2	M_c	M_2	M_c
ZS	$\boldsymbol{\beta}_0$	$Bias$	0.2219	**0.1103**	0.0179	**0.0087**	0.0424	**0.0400**	0.0021	0.0028
		VAR_F	0.3236	**0.0847**	0.0141	**0.0035**	0.0416	**0.0228**	0.0017	**0.0009**
		MSE	0.3414	**0.0888**	0.0142	**0.0035**	0.0423	**0.0233**	0.0017	**0.0009**
	$\boldsymbol{\beta}_2$	$Bias$	0.2500	**0.0999**	0.0211	**0.0053**	0.0536	**0.0464**	0.0057	0.0058
		VAR_F	0.1423	**0.1312**	0.0060	**0.0054**	0.0432	**0.0411**	0.0017	**0.0017**
		MSE	0.1635	**0.1346**	0.0062	**0.0054**	0.0444	**0.0420**	0.0018	**0.0017**
Shrinkage	$\boldsymbol{\beta}_0$	$Bias$	0.2066	**0.0890**	0.0206	**0.0068**	0.0332	**0.0250**	0.0015	**0.0015**
		VAR_F	0.3142	**0.0831**	0.0133	**0.0034**	0.0408	**0.0242**	0.0016	**0.0010**
		MSE	0.3291	**0.0858**	0.0134	**0.0034**	0.0411	**0.0245**	0.0016	**0.0010**
	$\boldsymbol{\beta}_2$	$Bias$	0.1749	**0.0762**	0.0118	**0.0050**	0.0484	**0.0420**	0.0041	**0.0034**
		VAR_F	0.1439	**0.1327**	0.0060	**0.0055**	0.0471	**0.0454**	0.0019	**0.0018**
		MSE	0.1545	**0.1349**	0.0060	**0.0055**	0.0479	**0.0460**	0.0019	**0.0018**

Table 2.4: Comparisons of M_2 and M_c in Sets 3 - 4

Prior	Design	Statistics	$n_1 = n_2$=10				$n_1 = n_2$=20			
			Set 3		Set 4		Set 3		Set 4	
			M_2	M_c	M_2	M_c	M_2	M_c	M_2	M_c
ZS	$\boldsymbol{\beta}_0$	$Bias$	0.2226	**0.1837**	0.0355	**0.0175**	0.0816	**0.0702**	0.0047	**0.0047**
		VAR_F	0.2738	**0.0745**	0.0137	**0.0035**	0.0394	**0.0218**	0.0017	**0.0009**
		MSE	0.2910	**0.0866**	0.0142	**0.0036**	0.0418	**0.0235**	0.0017	**0.0009**
	$\boldsymbol{\beta}_2$	$Bias$	0.2543	**0.1858**	0.0411	**0.0152**	0.0978	**0.0806**	0.0071	**0.0069**
		VAR_F	0.1220	**0.1149**	0.0060	**0.0054**	0.0412	**0.0393**	0.0017	**0.0017**
		MSE	0.1456	**0.1272**	0.0066	**0.0055**	0.0451	**0.0420**	0.0018	**0.0017**
Shrinkage	$\boldsymbol{\beta}_0$	$Bias$	0.2284	**0.1561**	0.0326	**0.0140**	0.0658	**0.0520**	0.0036	**0.0021**
		VAR_F	0.2746	**0.0745**	0.0131	**0.0034**	0.0391	**0.0234**	0.0016	**0.0010**
		MSE	0.2944	**0.0832**	0.0135	**0.0035**	0.0407	**0.0244**	0.0016	**0.0010**
	$\boldsymbol{\beta}_2$	$Bias$	0.2047	**0.1541**	0.0257	**0.0112**	0.0850	**0.0723**	0.0064	**0.0053**
		VAR_F	0.1285	**0.1187**	0.0060	**0.0054**	0.0451	**0.0436**	0.0019	**0.0018**
		MSE	0.1441	**0.1280**	0.0062	**0.0055**	0.0477	**0.0455**	0.0019	**0.0018**

denoted by "M_{gs}", is referred to as the case where covariates are fully observed as

Table 2.5: Comparisons of M_1 and M_c regarding posterior variance

Prior	Design	$n_1 = n_2$=10				$n_1 = n_2$=20			
	Parameter	Set 1	Set 2	Set 3	Set 4	Set 1	Set 2	Set 3	Set 4
ZS	$\boldsymbol{\beta}_0$	98.00%	98.60%	97.60%	98.40%	99.80%	99.80%	99.80%	99.60%
	$\boldsymbol{\beta}_1$	86.40%	90.60%	69.20%	88.80%	73.60%	75.40%	66.40%	74.40%
Shrinkage	$\boldsymbol{\beta}_0$	95.40%	95.80%	95.00%	95.60%	99.40%	99.40%	99.40%	99.40%
	$\boldsymbol{\beta}_1$	75.80%	79.00%	63.80%	77.40%	72.80%	74.00%	68.80%	73.80%

Table 2.6: Comparisons of M_2 and M_c regarding posterior variance

Prior	Design	$n_1 = n_2$=10				$n_1 = n_2$=20			
	Parameter	Set 1	Set 2	Set 3	Set 4	Set 1	Set 2	Set 3	Set 4
ZS	$\boldsymbol{\beta}_0$	99.00%	99.20%	98.80%	99.20%	97.60%	97.60%	97.80%	97.60%
	$\boldsymbol{\beta}_2$	77.80%	88.60%	61.80%	85.00%	71.20%	71.00%	69.00%	71.20%
Shrinkage	$\boldsymbol{\beta}_0$	97.20%	97.40%	97.00%	97.40%	96.20%	96.20%	96.40%	96.20%
	$\boldsymbol{\beta}_2$	69.00%	75.60%	55.60%	75.00%	64.80%	65.20%	63.60%	65.00%

below:

$$\begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{01} \\ \boldsymbol{X}_{02} \end{pmatrix} \boldsymbol{\beta}_0 + \begin{pmatrix} \boldsymbol{X}_1 & \boldsymbol{X}_{12} \\ \boldsymbol{X}_{21} & \boldsymbol{X}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{pmatrix}. \tag{2.36}$$

Therefore, as a complement to Section 2.4.2, we perform a sensitivity analysis to evaluate the behaviors of M_i, M_c and M_{gs}. Since Section 2.4.2 focused on a balanced design in terms of sample size ($n_1 = n_2$) and dimension of coefficients ($p_0 = p_1 = p_2$), we additionally consider the imbalanced design to offer a more complete view of candidate models.

Our analysis is conducted under two designs. In Design 1, fixing the dimension of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ at $p_0 = 4, p_1 = p_2 = 2$, we let the ratio of sample size vary from 0.5 to 1.0 for two sources with the following setups $n_1 = 30, n_2 = 15;\ n_1 = 30, n_2 = 30;\ n_1 = 30, n_2 = 45;\ n_1 = 30, n_2 = 60$. In Design 2, fixing the sample size at $n_1 = n_2 = 15$, we let the ratio of dimension for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ vary from 0.25 to 1.0 with the following setups $p_0 = 4, p_1 = p_2 = 1$; $p_0 = 4, p_1 = p_2 = 2$; $p_0 = 4, p_1 = p_2 = 3$ and $p_0 = p_1 = p_2 = 4$. At last, we examine these two scenarios under large and small size of coefficients. The largest model for large and small coefficients are listed in Sets 1 and 2 as below:

- Set 5: $\boldsymbol{\beta}_0 = (1.6, 1.5, 1.7, 1.6)'$, $\boldsymbol{\beta}_1 = (1.5, 1.7, 1.4, 1.2)'$, $\boldsymbol{\beta}_2 = (1.4, 1.8, 1.3, 1.3)'$, where all coefficients are large with $\sigma = 0.5$;

- Set 6: $\boldsymbol{\beta}_0 = (1.2, 0.8, 1.1, 0.9)'$, $\boldsymbol{\beta}_1 = (0.7, 0.4, 0.5, 0.3)'$, $\boldsymbol{\beta}_2 = (0.5, 0.6, 0.6, 0.2)'$, where all coefficients are small with $\sigma = 0.5$.

When $p_1,\ p_2 \leq 4$, the coefficients correspond to the first k elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in Sets 1 and 2. For each combination of two scenarios and two sets of coefficients, we collect quantities including relative bias (RBias), standard deviation (SD), relative MSE (RMSE), for $\boldsymbol{\beta}_i$, $i = 0, 1, 2$,

$$RBias = \frac{\sum_{r=1}^{500} \sum_{j=0}^{p_i} |\hat{\beta}_{ij}^{(r)} - \beta_{ij}|}{500 p_i \sum_{j=0}^{p_i} \beta_{ij}}, \tag{2.37}$$

$$RMSE = \frac{\sqrt{\sum_{r=1}^{500} \sum_{j=0}^{p_i} (\hat{\beta}_{ij}^{(r)} - \beta_{ij})^2}}{500 p_i \sum_{j=0}^{p_i} \beta_{ij}}. \tag{2.38}$$

Tables 2.7 and 2.8 present the frequentist properties of Bayesian estimators for large coefficients while Tables 2.9 and 2.10 show the frequentist properties for small coefficients. We have several main findings. First, consider M_{gs} as the reference model, M_c shows considerable advantage of M_i in terms of smaller deviation from M_{gs} for SD_F and $RMSE$ but not necessarily for bias. Second, the magnitude of advantage in M_c depends on the sample size and the dimension of coefficients. Specifically, when the dimension of coefficients is fixed, as shown in Tables 2.7 and 2.9, the deviation between M_c and M_{gs} decreases as the increase of sample size in terms of all frequentist properties studied in this context. In contrast, when the sample size is fixed, the deviation of M_c and M_{gs} increases as the dimension of specific coefficients $\boldsymbol{\beta}_i$ increases. Third, M_c is pretty robust to the misspecification compared with M_i in terms of frequentist variances especially for common coefficients $\boldsymbol{\beta}_0$. Occasionally, M_c has a smaller frequentist variance than M_{gs}. For example, in Table 2.7, the frequentist variances for $\boldsymbol{\beta}_0$ in M_c are all smaller than those in M_{gs} across all cases. Fourth, the Bayesian estimates in M_c differ from M_{gs} mainly in bias and the magnitude in MSE mainly depends on the bias rather than the frequentist variances. A deeper relevant discussion can be found in Section 2.6.

2.5 One Real Data Example

This section presents a student performance dataset to serve as a paradigm for our data combining method. This dataset was collected from two Portuguese secondary schools through mark reports and questionnaires (Cortez and Silva, 2008). It has been frequently used to study the association between student performance, which is reflected by their scores, and covariates including demographic, social and family features from multiple perspectives. We first tailor this dataset to our context by applying group lasso to select covariates included in our model and then we divide the dataset into two according to schools so that M_1, M_2, and M_c could be simulated. Here, we also use M_i or M_c to indicate the related datasets. As a result, M_1 corresponds to Gabriel Pereira school with 349 observations and M_2 corresponds to Mousinho da Silveira school with 46 observations. The common covariates include the

Table 2.7: Sensitivity analysis for Design 1 with Set 5

Pamameters	Design	$n_1 = 30$, $n_2 = 15$				$n_1 = 30$, $n_2 = 30$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0379	0.0480	0.0364	0.0012	0.0305	0.0571	0.0457	0.0004
	SD_F	0.0504	0.0846	0.0421	0.0424	0.0524	0.0439	0.0334	0.0336
	$RMSE$	0.0244	0.0314	0.0232	0.0066	0.0377	0.0377	0.0337	0.0053
$\boldsymbol{\beta}_1$	$RBias$	0.1824	-	0.1785	0.0045	0.0515	-	0.0449	0.0010
	SD_F	0.0837	-	0.0822	0.0676	0.0753	-	0.0698	0.0492
	$RMSE$	0.1317	-	0.1288	0.0212	0.0443	-	0.0385	0.0154
$\boldsymbol{\beta}_2$	$RBias$	-	0.1675	0.1130	0.0003	-	0.0203	0.0205	0.0012
	SD_F	-	0.1255	0.1056	0.0557	-	0.0747	0.0738	0.0456
	$RMSE$	-	0.1461	0.1041	0.0174	-	0.0278	0.0273	0.0143
	Design	$n_1 = 30$, $n_2 = 45$				$n_1 = 30$, $n_2 = 60$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0265	0.0239	0.0212	0.0006	0.0305	0.0343	0.0245	0.0006
	SD_F	0.0496	0.0420	0.0298	0.0301	0.0471	0.0338	0.0276	0.0281
	$RMSE$	0.1004	0.0926	0.0769	0.0302	0.1226	0.1176	0.0906	0.0282
$\boldsymbol{\beta}_1$	$RBias$	0.0950	-	0.0900	0.0010	0.1241	-	0.0795	0.0011
	SD_F	0.0696	-	0.0665	0.0384	0.0763	-	0.0725	0.0361
	$RMSE$	0.0727	-	0.0701	0.0120	0.1019	-	0.0675	0.0113
$\boldsymbol{\beta}_2$	$RBias$	-	0.0367	0.0365	0.0009	-	0.0420	0.0399	0.0008
	SD_F	-	0.0608	0.0581	0.0433	-	0.0444	0.0442	0.0367
	$RMSE$	-	0.0398	0.0363	0.0135	-	0.0415	0.0397	0.0115

age of a student β_{01} and the number of past class failures β_{02}. The specific covariate is the mother's education β_{11} for M_1 and the number of school absences β_{21} for M_2. The response variable is the average score of three exams (first period grade, second period grade and final grade) for a student. After 20,000 samples with 10,000 burn-in in MCMC, we collect the posterior mean (M_p), posterior variance (VAR_p), 95% credible intervals (CI) and its corresponding width. Table 2.11 summarizes the key results for the analysis. There are several main findings. First, for each parameter, the posterior variances for the combined data are smaller compared using individual data alone. Second, for each parameter, the width of 95% credible intervals for smaller for

Table 2.8: Sensitivity analysis for Design 2 with Set 5

Parameters	Design	$p_0 = 4, p_1 = p_2 = 1$				$p_0 = 4, p_1 = p_2 = 2$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0346	0.0352	0.0213	0.0013	0.0608	0.0360	0.0144	0.0017
	SD_F	0.0695	0.0911	0.0501	0.0499	0.0832	0.0712	0.0532	0.0528
	$RMSE$	0.0244	0.0277	0.0152	0.0078	0.0379	0.0259	0.0122	0.0083
$\boldsymbol{\beta}_1$	$RBias$	0.1481	-	0.0148	0.0001	0.1166	-	0.0565	0.0025
	SD_F	0.1435	-	0.1356	0.0894	0.1140	-	0.1089	0.0886
	$RMSE$	0.1764	-	0.0917	0.0596	0.0898	-	0.0534	0.0278
$\boldsymbol{\beta}_2$	$RBias$	-	0.6261	0.2123	0.0070	-	0.1276	0.0498	0.0025
	SD_F	-	0.2963	0.2193	0.1257	-	0.1313	0.1183	0.0915
	$RMSE$	-	0.6610	0.2638	0.0901	-	0.1068	0.0571	0.0287
	Design	$p_0 = 4, p_1 = p_2 = 3$				$p_0 = 4, p_1 = p_2 = 4$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0914	0.0634	0.0651	0.0006	0.0960	0.0752	0.0405	0.0011
	SD_F	0.0775	0.1085	0.0578	0.0612	0.1014	0.0860	0.0492	0.0552
	$RMSE$	0.0532	0.0382	0.0376	0.0096	0.0747	0.0599	0.0262	0.0086
$\boldsymbol{\beta}_1$	$RBias$	0.1207	-	0.0957	0.0023	0.1570	-	0.0758	0.0018
	SD_F	0.0927	-	0.0866	0.0665	0.0875	-	0.0747	0.0652
	$RMSE$	0.0763	-	0.0732	0.0145	0.0889	-	0.0519	0.0113
$\boldsymbol{\beta}_2$	$RBias$	-	0.3015	0.2874	0.0018	-	0.0619	0.0564	0.0009
	SD_F	-	0.0927	0.0896	0.0611	-	0.1027	0.0865	0.0598
	$RMSE$	-	0.1899	0.1852	0.0136	-	0.0471	0.0386	0.0103

the combined data. Third, the benefit of data combining reach its best for common coefficients rather than the specific coefficients in terms of both posterior variances and width of 95% credible intervals. Fourth, individual model and combined model yield the same conclusion regarding whether the 95% cover 0. Specifically, the CIs for age, the number of failures, and mother's education exclude 0 while the CIs for school absences include 0 across all models.

Table 2.9: Sensitivity analysis for Design 1 with Set 6

Parameters	Design	$n_1 = 30$, $n_2 = 15$				$n_1 = 30$, $n_2 = 30$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0329	0.0454	0.0238	0.0024	0.0381	0.0225	0.0301	0.0023
	SD_F	0.0500	0.1001	0.0400	0.0393	0.0500	0.0555	0.0342	0.0349
	$RMSE$	0.0227	0.0379	0.0180	0.0099	0.0267	0.0184	0.0205	0.0088
$\boldsymbol{\beta}_1$	$RBias$	0.2370	-	0.1867	0.0034	0.0220	-	0.0448	0.0042
	SD_F	0.0792	-	0.0730	0.0610	0.0718	-	0.0650	0.0439
	$RMSE$	0.1835	-	0.1503	0.0557	0.0674	-	0.0671	0.0401
$\boldsymbol{\beta}_2$	$RBias$	-	0.1678	0.0503	0.0062	-	0.1700	0.1297	0.0030
	SD_F	-	0.1855	0.1070	0.0510	-	0.0598	0.0581	0.0427
	$RMSE$	-	0.2181	0.1045	0.0466	-	0.1322	0.1064	0.0388
	Design	$n_1 = 30$, $n_2 = 45$				$n_1 = 30$, $n_2 = 60$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0348	0.0359	0.0334	0.0020	0.0262	0.0151	0.0104	0.0008
	SD_F	0.0441	0.0437	0.0302	0.0307	0.0562	0.0348	0.0284	0.0277
	$RMSE$	0.0231	0.0212	0.0189	0.0078	0.0207	0.0124	0.0098	0.0069
$\boldsymbol{\beta}_1$	$RBias$	0.3941	-	0.3740	0.0049	0.1034	-	0.0861	0.0025
	SD_F	0.0709	-	0.0704	0.0520	0.0908	-	0.0792	0.0415
	$RMSE$	0.2942	-	0.2792	0.0475	0.1133	-	0.0956	0.0378
$\boldsymbol{\beta}_2$	$RBias$	-	0.0387	0.0330	0.0035	-	0.1004	0.0596	0.0008
	SD_F	-	0.0550	0.0539	0.0406	-	0.0512	0.0490	0.0400
	$RMSE$	-	0.0571	0.0542	0.0369	-	0.0853	0.0618	0.0364

2.6 Discussion

In this chapter, we evaluated the use of M_i and M_c in terms of posterior variances and frequentist properties under Zellner's g-prior, ZS prior, and shrinkage prior. When g is known, Theorems 2.1 and 2.2 establish the sufficient and necessary conditions to achieve a smaller posterior variance in M_c. These conditions depend on the value of g and the minimum eigenvalue of a function of design matrices while additionally rely on the observations when σ^2 is unknown. Furthermore, we considered a more general case, where we allow g to follow an inverse-gamma distribution. To handle the

Table 2.10: Sensitivity analysis for Design 2 with Set 6

Parameter	Design	$p_0 = 4, p_1 = p_2 = 1$				$p_0 = 4, p_1 = p_2 = 2$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
	$RBias$	0.0324	0.0290	0.0203	0.0036	0.0102	0.0648	0.0088	0.0040
$\boldsymbol{\beta}_0$	SD_F	0.1054	0.0680	0.0530	0.0552	0.0995	0.0948	0.0560	0.0554
	$RMSE$	0.0326	0.0261	0.0203	0.0139	0.0257	0.0414	0.0147	0.0140
	$RBias$	0.1019	-	0.0764	0.0144	0.2486	-	0.1857	0.0115
$\boldsymbol{\beta}_1$	SD_F	0.1330	-	0.1237	0.1044	0.0966	-	0.0795	0.0716
	$RMSE$	0.2157	-	0.1927	0.1498	0.1965	-	0.1501	0.0657
	$RBias$	-	0.2806	0.2834	0.0022	-	0.0710	0.1857	0.0065
$\boldsymbol{\beta}_2$	SD_F	-	0.1587	0.1568	0.1404	-	0.1067	0.0795	0.0728
	$RMSE$	-	0.4233	0.4224	0.2807	-	0.1165	0.1501	0.0663
	Design	$p_0 = 4, p_1 = p_2 = 3$				$p_0 = 4, p_1 = p_2 = 4$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
	$RBias$	0.0562	0.0374	0.0386	0.0050	0.0425	0.0514	0.0404	0.0078
$\boldsymbol{\beta}_0$	SD_F	0.0854	0.0996	0.0566	0.0606	0.0865	0.0875	0.0590	0.0643
	$RMSE$	0.0369	0.0314	0.0241	0.0154	0.0330	0.0351	0.0263	0.0166
	$RBias$	0.0768	-	0.0699	0.0064	0.0439	-	0.0513	0.0079
$\boldsymbol{\beta}_1$	SD_F	0.1015	-	0.0975	0.0656	0.0936	-	0.0891	0.0645
	$RMSE$	0.0783	-	0.0791	0.0412	0.0545	-	0.0537	0.0343
	$RBias$	-	0.0659	0.0555	0.0055	-	0.1134	0.1104	0.0098
$\boldsymbol{\beta}_2$	SD_F	-	0.1001	0.0913	0.0645	-	0.0865	0.0791	0.0652
	$RMSE$	-	0.0752	0.0626	0.0382	-	0.0778	0.0754	0.0347

non-standard marginal distribution for g given data, we utilized two popular Laplace approximation methods to evaluate the Bayesian estimator, posterior variance and frequentist variances. We found that, in general, M_c is equivalent or better than M_i in terms of a smaller posterior variance and frequentist variance, especially for common coefficients. Our simulation studies also show that this conclusion holds for most scenarios we considered. The advantage of M_c over M_i is more evident when the sample size is small and/or the dimension of specific coefficients is small compared with the dimension of common coefficients. We also performed a sensitivity

Table 2.11: Results for the student performance dataset

Parameter		M_1	M_c	M_2
β_{01}	M_p (VAR_p)	0.5740 (0.0012)	0.5984 (0.0007)	0.6291 (0.0017)
	95 % CI	(0.5068, 0.6411)	(0.5474, 0.6497)	(0.5467, 0.7104)
	Width	0.1343	0.1023	0.1637
β_{02}	M_p (VAR_p)	-1.8936 (0.0741)	-1.9282 (0.0629)	-1.8607 (0.4573)
	95 % CI	(-2.4172, -1.3518)	(-2.4186, -1.4382)	(-3.1843, -0.5230)
	Width	1.0655	0.9804	2.6613
β_{11}	M_p (VAR_p)	0.6283 (0.0330)	0.5052 (0.0210)	-
	95 % CI	(0.2722, 0.9782)	(0.2208, 0.7875)	-
	Width	0.7060	0.5666	-
β_{21}	M_p (VAR_p)	-	-0.0274 (0.0113)	-0.1041 (0.0158)
	95 % CI	-	(-0.2392, 0.1790)	(-0.3504, 0.1393)
	Width	-	0.4182	0.4897

analysis in terms of M_c compared with the golden standard, where the covariates are fully observed, to evaluate whether M_c is applicable. We found that the discrepancy between M_c and M_{gs} increases as dimension of coefficients increases. Thus, it is probably best not to use M_c when p_i/p_0 is large.

Finally, we discuss some issues and future directions based on the work in this chapter. To begin with, we assume that the elements in $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ are of similar sizes. However, it is common for different parameters to have different sizes. For example, we could have $\boldsymbol{\beta}_0$ with large size and $\boldsymbol{\beta}_i$ with small size. In this case, it no longer makes sense to use an overall g to govern all parameters. In fact, associated problems of using the same scale parameter for all coefficients have been noticed. For example, Agliari and Parisetti (1988) derived A-g prior and found that the limiting behavior of a single parameter is affected by different scales in the prior. Som et al. (2015) further revealed that using a single g to regulate all parameters may result in unsatisfactory fixed p - fixed n conditional information paradox. Second, we strictly explore the posterior variance and frequentist variance when $R^2 \to 1$ to demonstrate the limiting behavior to offer a reasonable explanation regarding the properties of the Bayesian estimator. In the next chapter, we study the independent g-prior and its relative performances in M_i and M_c under the conditional asymptotic defined by Som et al. (2016).

Chapter 3

Independent Mixtures of G-Priors

3.1 Introduction

In the previous chapter, we employed the standard g-prior, ZS prior, and shrinkage prior on the coefficient with a single g controlling the shrinkage. However, several problems for such specification still exist. For one thing, it requires the design matrix for all coefficients to be of full column rank, which put more constraints on the number of observations. For another thing, it has an undesirable theoretical property in the context of model selection Som et al. (2015). Specifically, a new form of conditional asymptotic limit driven by a situation arising in many practical problems when one or more groups of regression coefficients are much larger than the rest. Under this asymptotic, many prominent "g-type" priors, such as hyper-g prior (Liang et al., 2008) and robust g prior (Bayarri et al., 2012), are shown to suffer from the Conditional Lindley's Paradox (CLP), which is interpreted as, "*if at least one of the regression co-*

efficients common to both models is quite large compared to the additional coefficients in the bigger model, then the Bayes factor due to the hyper-g shows unwarranted bias toward choosing the smaller model." The rationale behind this undesired behavior is that the common mixing parameter g in these priors introduces a mono-shrinkage.

One way to alleviate CLP is to employ the block g-prior proposed by Som et al. (2015), which has also been explored as independent g-prior in Min and Sun (2016). With the new form of the asymptotic limit, they focused on the demonstration of these undesirable issues in the traditional g-prior while we aim at the estimation of the coefficients such as posterior variances and the frequentist variances of the Bayesian estimator with the Zellner-Siow prior. More importantly, we need to examine such quantities through the comparison of M_c and M_i. For instance, we are interested in how the coefficient estimators differ in the M_c and M_i when the dominated one is shared, and what is the tendency of the bias and covariance of the Bayesian estimators. Therefore, in this chapter, we focus on the independent g-priors for $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_i$. The independent version of g priors not only allows us to specify different shrinkage effects for $\boldsymbol{\beta}_{0i}$ and $\boldsymbol{\beta}_i$ but also offer a more flexible requirement for the rank of the design matrix. For example, independent g-priors only requires $\boldsymbol{X}_{i0}$ and $\boldsymbol{X}_i$ to be of full column rank, respectively, while dependent g-prior requires $(\boldsymbol{X}_{0i}, \boldsymbol{X}_i)$ to be of full column rank. Notice that when the whole design matrix $\tilde{\boldsymbol{X}}_i$ or $\tilde{\boldsymbol{X}}$ are block diagonal (equivalently, $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$) and $g_{0i} = g_i$ for M_i, $g_0 = g_1 = g_2 = g$ for M_c, the independent g-priors reduce to dependent g-priors.

Notice that the sample size in the ZS prior remains the same for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ across M_i and M_c. This does not agree with our intuitions since at most n_i observations contribute for the estimation of $\boldsymbol{\beta}_i$ even in M_c. As a matter of fact, the potential misuse of the number of observations as sample size has been identified and discussed in the context of model selection. For example, the well-known Schwarz criterion or Bayesian information criterion (BIC) (Schwarz, 1978) is formulated as BIC $= 2l(\hat{\boldsymbol{\beta}}) - p\log(n)$, where $l(\hat{\boldsymbol{\beta}})$ is the estimated log-likelihood for the model, p is the dimension of the model parameter $\boldsymbol{\beta}$, and n is the sample size. It has become a standard procedure in model selection since it serves as an approximation to the logrithm of the Bayes factor with large samples. However, the sample size n has been suggested to be determined carefully. For one thing, its derivation reveals that n should reflect the number of data values contributing to the summation that appears in the formula for the Hessian and the approximation only works well for limited settings (Kass and Raftery (1995), Stone (1979), Weakliem (1999)). Many efforts have been made to improve its performance in more general situations rather than iid observations including Kass and Wasserman (1995), Berger et al. (2014), Bayarri et al. (2019) and Berger et al. (2019). Similar use of n exists in Zellner-Siow prior along with its variants (Cui and George (2008), Liang et al. (2008), Wang (2017)), where n is used to adjust the prior scale. It's natural to investigate whether n is well-defined in such cases. In fact, Berger et al. (2014) has addressed this issue and proposed the effective sample size (TESS) to obtain a reasonable sample size for individual parameter by removing the corresponding scale. One example of demonstrating the benefits of TESS is Findley's (Findley, 1991) counterexample, which shows the inconsistency of BIC in hypothesis testing. This issue can be resolved with application of TESS. As pointed out by Berger et al. (2014), the utilization of TESS should not be limited to the model selection and therefore we probe the impacts of TESS from the estimation perspective.

In Section 1.3.2, we review the model and present notations. In Section 3.2, we study the independent g-prior where (g, σ^2) is considered known and presents the comparative results for M_i and M_c in terms of posterior variances. Then, in Section 3.3, we have a focused investigation on the independent ZS prior and perform an asymptotic analysis for the frequentist property regarding the Bayesian estimators obtained from M_i and M_c. Additionally, the corresponding results could be easily applied to TESS. In Section 3.4, we perform numerical analyses to illustrate the theorems in Section 3.3. At last, a real data analysis is presented for a demonstrative purpose.

3.2 Independent G-priors with Known (σ^2, g)

3.2.1 Priors and Posterior Distributions

Fact 3.1. *For M_i, independent conventional g-prior for $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i)$ is:*

$$\boldsymbol{\beta}_0|\sigma^2, g_0 \sim N_{p_0}(\mathbf{0}, g_0\sigma^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}), \ \boldsymbol{\beta}_i|\sigma^2, g_i \sim N_{p_i}(\mathbf{0}, g_i\sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}) \tag{3.1}$$

(a) Define $\boldsymbol{S}_i = diag\left(g_0(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}, g_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\right)$ and $\boldsymbol{g}_i = (g_0, g_i)'$, the posterior distribution for $(\tilde{\boldsymbol{\beta}}_i|\sigma^2, \boldsymbol{g}_i, \boldsymbol{y}_i)$ is $N_{p_I}(\tilde{\boldsymbol{\beta}}_i^B, \boldsymbol{\Sigma}_i^B)$, where

$$\tilde{\boldsymbol{\beta}}_i^B = (\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{S}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i \text{ and } \boldsymbol{\Sigma}_i^B = \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{S}_i^{-1})^{-1}.$$

(b) The marginal posterior variances for $\boldsymbol{\beta}_0$ *and* $\boldsymbol{\beta}_i$ *are:*

$$VAR(\boldsymbol{\beta}_0|\sigma^2,\boldsymbol{g}_i,\boldsymbol{y}_i,M_i) = \sigma^2\big\{\boldsymbol{X}_{0i}^{'}[(1+g_0^{-1})\boldsymbol{I}_{n_i} - (1+g_i^{-1})^{-1}\boldsymbol{P}_i]\boldsymbol{X}_{0i}\big\}^{-1},$$

$$VAR(\boldsymbol{\beta}_i|\sigma^2,\boldsymbol{g}_i,\boldsymbol{y}_i,M_i) = \sigma^2\Big\{(g_i^{-1}+1)^{-1}(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1} + (g_i^{-1}+1)^{-2}(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1}$$
$$\boldsymbol{X}_i^{'}\boldsymbol{X}_{0i}\{\boldsymbol{X}_{0i}^{'}[(1+g_0^{-1})\boldsymbol{I}_{n_i} - (1+g_i^{-1})^{-1}\boldsymbol{P}_i]\boldsymbol{X}_{0i}\}^{-1}\boldsymbol{X}_{0i}^{'}\boldsymbol{X}_i(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1}\Big\}.$$

Fact 3.2. *For* M_c*, the independent conventional g-prior is:*

$$\boldsymbol{\beta}_0 \sim N_{p_0}(\boldsymbol{0}, g_0\sigma^2(\boldsymbol{X}_0^{'}\boldsymbol{X}_0)^{-1}), \boldsymbol{\beta}_i \sim N_{p_i}(\boldsymbol{0}, g_i\sigma^2(\boldsymbol{X}_i^{'}\boldsymbol{X}_i)^{-1}), i = 1,2. \tag{3.2}$$

(a) Define $\boldsymbol{S} = diag\,(g_0(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}, g_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}, g_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1})$ *and* $\boldsymbol{g} = (g_0, g_1, g_2)$*. The posterior distribution for* $(\tilde{\boldsymbol{\beta}}|\sigma^2, \boldsymbol{g}, \boldsymbol{y})$ *is normal distribution with posterior mean* $\tilde{\boldsymbol{\beta}}^B$ *and* $\boldsymbol{\Sigma}^B$*, where*

$$\tilde{\boldsymbol{\beta}}^B = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{S}^{-1})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y} \text{ and } \boldsymbol{\Sigma}^B = \sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{S}^{-1})^{-1}.$$

(b) The marginal posterior variances for $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ *are:*

$$VAR(\boldsymbol{\beta}_0|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c) = \sigma^2\big\{\boldsymbol{X}_{01}^{'}[(1+g_0^{-1})\boldsymbol{I}_{n_1} - (1+g_1^{-1})^{-1}\boldsymbol{P}_1]\boldsymbol{X}_{01}$$
$$+ \boldsymbol{X}_{02}^{'}[(1+g_0^{-1})\boldsymbol{I}_{n_2} - (1+g_2^{-1})^{-1}\boldsymbol{P}_2]\boldsymbol{X}_{02}\big\}^{-1},$$

$$VAR(\boldsymbol{\beta}_1|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c) = \sigma^2\Big\{(1+g_1^{-1})^{-1}(\boldsymbol{X}_1^{'}\boldsymbol{X}_1)^{-1} + (1+g_1^{-1})^{-2}(\boldsymbol{X}_1^{'}\boldsymbol{X}_1)^{-1}$$
$$\boldsymbol{X}_1^{'}\boldsymbol{X}_{01}\big\{\sum_{i=1}^{2}\boldsymbol{X}_{0i}^{'}[(1+g_0^{-1})\boldsymbol{I}_{n_i} - (1+g_i^{-1})^{-1}\boldsymbol{P}_i]\boldsymbol{X}_{0i}\big\}^{-1}\boldsymbol{X}_{01}^{'}\boldsymbol{X}_1(\boldsymbol{X}_1^{'}\boldsymbol{X}_1)^{-1}\Big\},$$

$$VAR(\boldsymbol{\beta}_2|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c) = \sigma^2\Big\{(1+g_2^{-1})^{-1}(\boldsymbol{X}_2^{'}\boldsymbol{X}_2)^{-1} + (1+g_2^{-1})^{-2}(\boldsymbol{X}_2^{'}\boldsymbol{X}_2)^{-1}$$

$$\boldsymbol{X}_2'\boldsymbol{X}_{02}\Big\{\sum_{i=1}^{2}\boldsymbol{X}_{0i}'[(1+g_0^{-1})\boldsymbol{I}_{n_i}-(1+g_i^{-1})^{-1}\boldsymbol{P}_i]\boldsymbol{X}_{0i}\Big\}^{-1}\boldsymbol{X}_{02}'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\Big\}.$$

Theorem 3.1. *With independent g-priors in (3.1) for M_i and in (3.2) for M_c,*

$$VAR(\boldsymbol{\beta}_0|\sigma^2,\boldsymbol{g}_i,\boldsymbol{y}_i,M_i) > VAR(\boldsymbol{\beta}_0|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c), \tag{3.3}$$

$$VAR(\boldsymbol{\beta}_i|\sigma^2,\boldsymbol{g}_i,\boldsymbol{y}_i,M_i) \geq VAR(\boldsymbol{\beta}_i|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c). \tag{3.4}$$

Proof. See Appendix A.2.1. □

In fact, results in 3.3 and 3.4 hold no matter what values g_0, g_1, g_2, n_1 and n_2 take. On the premises of Theorem 3.1, for common regression coefficients, combining data always provides more precise estimates. For specific regression coefficients, combining data provides at least equivalently precise estimates with respect to posterior variances. One special case would be $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, which has been widely adopted in hypothesis testing or variable selection using g-priors. It's easy to check that $VAR(\boldsymbol{\beta}_i|\sigma^2,\boldsymbol{g}_i,\boldsymbol{y}_i,M_i) = VAR(\boldsymbol{\beta}_i|\sigma^2,\boldsymbol{g},\boldsymbol{y},M_c)$ by the proof of Theorem 3.1. This indicates that, when the design matrix is block diagonal, there is no benefits for $\boldsymbol{\beta}_i$ with data combining from the estimation perspective. Also, notice that Theorem 3.1 assumes g_0, g_1 and g_2 are the same for M_i and M_c. One may be interested in the case of setting different g values before and after the combining. We wouldn't pursue this aspect here. For one thing, recommended fixed g-priors in model selection generally lead to overly biased posterior means and it would be better to set g as a large number for estimation. Since ZS prior has better properties and depends on the sample size, we would pursue using different g-priors for M_i and M_c with ZS prior.

The following corollary gives a special case where the intercept is only term shared by two data sources.

Corollary 3.1. *Assume that* $\boldsymbol{X}_{0i} = \mathbf{1}_{n_i}$, *it reduces to the case when two sources only share the same intercept. Define* s_i *as the summation of all elements in the projection matrix* $\boldsymbol{P}_i$. *For* M_i,

$$
\begin{aligned}
VAR(\boldsymbol{\beta}_0|\boldsymbol{g}_i, \boldsymbol{y}_i, M_i) &= \sigma^2\{(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i\}^{-1},\\
VAR(\boldsymbol{\beta}_i|\boldsymbol{g}_i, \boldsymbol{y}_i, M_i) &= \sigma^2\big\{(g_i^{-1}+1)^{-1}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} + (g_i^{-1}+1)^{-2}\\
&\qquad \{(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i\}^{-1}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{J}_{n_i}\boldsymbol{X}_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\big\}.
\end{aligned}
$$

For M_c,

$$
\begin{aligned}
VAR(\boldsymbol{\beta}_0|\boldsymbol{g}, \boldsymbol{y}, M_c) &= \sigma^2\{\sum_{i=1}^{2}(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i\}^{-1},\\
VAR(\boldsymbol{\beta}_1|\boldsymbol{g}, \boldsymbol{y}, M_c) &= \sigma^2\big\{(g_1^{-1}+1)^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} + (g_1^{-1}+1)^{-2}\\
&\qquad \{\sum_{i=1}^{2}(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i\}^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{J}_{n_1}\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\big\},\\
VAR(\boldsymbol{\beta}_2|\boldsymbol{g}, \boldsymbol{y}, M_c) &= \sigma^2\big\{(g_2^{-1}+1)^{-1}(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1} + (g_2^{-1}+1)^{-2}\\
&\qquad \{\sum_{i=1}^{2}(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i\}^{-1}(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{J}_{n_2}\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\big\}.
\end{aligned}
$$

We can see that $VAR(\boldsymbol{\beta}_i|\boldsymbol{g}, \boldsymbol{y}, M_c) \leq VAR(\boldsymbol{\beta}_i|\boldsymbol{g}_i, \boldsymbol{y}_i, M_i)$ *and* $VAR(\boldsymbol{\beta}_0|\boldsymbol{g}, \boldsymbol{y}, M_c) \leq VAR(\boldsymbol{\beta}_0|\boldsymbol{g}_i, \boldsymbol{y}_i, M_i)$ *if* $g_i \geq g$ *for* $i = 1, 2$. *This result can be obtained by the inequality* $(1+g_0^{-1})n_i - (1+g_i^{-1})^{-1}s_i \geq 0$.

3.2.2 Frequentist Properties for $\tilde{\boldsymbol{\beta}}_i^B$ and $\tilde{\boldsymbol{\beta}}^B$

Fact 3.3. *(a) For M_i, the frequentist distribution for $\tilde{\boldsymbol{\beta}}_i^B$ is $N(\boldsymbol{m}_i, \boldsymbol{V}_i)$, where*

$$\boldsymbol{m}_i = (\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{S}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i,\ \boldsymbol{V}_i = \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{S}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{S}_i^{-1})^{-1}.$$

(b) For M_c, the frequentist distribution for $\tilde{\boldsymbol{\beta}}^B$ is $N(\boldsymbol{m}, \boldsymbol{V})$, where

$$\boldsymbol{m} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{S}^{-1})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\tilde{\boldsymbol{\beta}},\ \boldsymbol{V} = \sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{S}^{-1})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{S}^{-1})^{-1}.$$

Remark 3.1. *If we let g_0, g_1, g_2 go ∞, the frequentist variances reduce to $\boldsymbol{V}_i = \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i)^{-1}$ and $\boldsymbol{V} = \sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}$. Consequently, the frequentist variances from M_c are smaller than or equal to M_i for $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, respectively.*

3.3 Independent Zellner-Siow Priors

In this subsection, we consider independent ZS prior, where (σ^2, g) is considered unknown.

3.3.1 Priors and Posterior Distributions

We first present the priors for M_i and M_c, and show their corresponding posterior distributions. For M_i, we use priors:

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2},$$
$$\boldsymbol{\beta}_0|g_0, \sigma^2 \sim N_{p_0}(0, g_0 n_i \sigma^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}),$$

$$\boldsymbol{\beta}_i|g_i,\sigma^2 \sim N_{p_i}(0, g_i n_i \sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}),$$
$$\pi(g_i) \propto g_i^{-\frac{3}{2}} \exp(-\frac{1}{2g_i}), i = 0, 1, 2. \tag{3.5}$$

For M_c, we use priors:

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2},$$
$$\boldsymbol{\beta}_0|g_{0c},\sigma^2 \sim N_{p_0}(0, g_{0c} n \sigma^2(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}),$$
$$\boldsymbol{\beta}_i|g_{ic},\sigma^2 \sim N_{p_i}(0, g_{ic} n \sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}), i = 1, 2,$$
$$\pi(g_{ic}) \propto g_{ic}^{-\frac{3}{2}} \exp(-\frac{1}{2g_{ic}}), i = 0, 1, 2. \tag{3.6}$$

Define the covariance matrix for $\tilde{\boldsymbol{\beta}}_i$ as $\boldsymbol{C}_i = n_i diag\big(g_0(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}, g_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\big)$ and the covariance matrix for $\tilde{\boldsymbol{\beta}}$ as $\boldsymbol{C} = n diag\big(g_{0c}(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}, g_{1c}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}, g_{2c}(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\big)$ in M_i and M_c, respectively.

Remark 3.2. *In Section 3.2, we use the same g_0, g_1, g_2 for M_i and M_c. However, in Section 3.3, it is equivalent to use $g_i \sim IG(1/2, n_i/2)$ for M_i but $g_{ic} \sim IG(1/2, n/2)$, where we allow g_i and g_{ic} to adjust according to the sample size for M_i and M_c.*

Since M_i and M_c share similar structures regarding the posterior distributions, we only present the results for M_i for brevity. The Bayesian estimator $\tilde{\boldsymbol{\beta}}_i^B$ for $\tilde{\boldsymbol{\beta}}_i$ is:

$$\tilde{\boldsymbol{\beta}}_i^B = E(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, M_i) = E\big(E(\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_0, g_i, \boldsymbol{y}_i)|\boldsymbol{y}_i, M_i\big)$$
$$= E\big((\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i|\boldsymbol{y}_i, M_i\big). \tag{3.7}$$

The posterior variance of $(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, M_i)$ can be computed by the total law of variation and the law of total expectation as below:

$$
\begin{aligned}
& VAR(\tilde{\boldsymbol{\beta}}_i|\boldsymbol{y}_i, M_i) \\
=& E(VAR(\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_0, g_i, \boldsymbol{y}_i)|\boldsymbol{y}_i, M_i) + VAR\big(E(\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_0, g_i, \boldsymbol{y}_i)|\boldsymbol{y}_i, M_i\big) \\
=& E\left(\sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}|\boldsymbol{y}_i, M_i\right) + VAR\big((\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i|\boldsymbol{y}_i, M_i\big) \\
=& E\big[E(\sigma^2|g_0, g_i, \boldsymbol{y}_i)(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}|\boldsymbol{y}_i, M_i\big] + VAR\big[(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i|\boldsymbol{y}_i, M_i\big] \\
=& E\left(\frac{\boldsymbol{y}_i'(\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i')\boldsymbol{y}_i}{n_i - 2}(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}|\boldsymbol{y}_i, M_i\right) \\
& + VAR\big((\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i|\boldsymbol{y}_i, M_i\big). \qquad (3.8)
\end{aligned}
$$

3.3.2 Reparameterization of g

Next, we explore the behaviors of the Bayesian estimators from M_i and M_c, which are $\tilde{\boldsymbol{\beta}}_i^B$ and $\tilde{\boldsymbol{\beta}}_c^B$, respectively. Since a direct integration for the marginal posterior distributions in terms of g_0 or g_i is mathmetically difficult, we consider a special case, where $\boldsymbol{X}_{0i}$ and $\boldsymbol{X}_i$ are orthogonal, for a theoretical guidance. Notice that this is actually the worse case where there is no information borrowing between the common and specific factors. Namely, we expect better performances for the non-orthogonal design.

When $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, the joint posterior distribution of g_0 and g_i in M_i is:

$$
f(g_0, g_i|\boldsymbol{y}_i) \propto \frac{(g_0 g_i)^{-\frac{3}{2}}(1+g_0 n_i)^{-\frac{p_0}{2}}(1+g_i n_i)^{-\frac{p_i}{2}} exp\big(-\frac{g_0+g_i}{2g_0 g_i}\big)}{\Big[\boldsymbol{y}_i'\big(\boldsymbol{I}_{n_i} - \frac{g_0 n_i}{1+g_0 n_i}\boldsymbol{P}_{X_{0i}} - \frac{g_i n_i}{1+g_i n_i}\boldsymbol{P}_{X_i}\big)\boldsymbol{y}_i\Big]^{n_i/2}}, \qquad (3.9)
$$

where $\boldsymbol{P}_{X_{0i}}$ and $\boldsymbol{P}_{X_i}$ are the projection matrices generated by $\boldsymbol{X}_{0i}$ and $\boldsymbol{X}_i$. To make

(3.9) more tracktable, we transform (g_0, g_i) through:

$$t_0 = \frac{g_0 n_i}{1 + g_0 n_i}, t_i = \frac{g_i n_i}{1 + g_i n_i}, \tag{3.10}$$

$R^2_{0i} = \boldsymbol{y}_i' \boldsymbol{P}_{X_{0i}} \boldsymbol{y}_i / \boldsymbol{y}_i' \boldsymbol{y}_i$ and $R^2_i = \boldsymbol{y}_i' \boldsymbol{P}_{X_i} \boldsymbol{y}_i / \boldsymbol{y}_i' \boldsymbol{y}_i$, the density in (3.9) is equivalent to:

$$f(t_0, t_i | \boldsymbol{y}_i) \propto \frac{(t_0 t_i)^{-\frac{3}{2}} (1 - t_0)^{\frac{p_0 - 1}{2}} (1 - t_i)^{\frac{p_i - 1}{2}} exp\big(- \frac{n_i (t_0 + t_i)}{2 t_0 t_i} \big)}{(1 - t_0 R^2_{0i} - t_i R^2_i)^{n_i/2}}. \tag{3.11}$$

Similarly, for M_c, if we transform (g_0, g_1, g_2) into:

$$t_{0c} = \frac{g_0 n}{1 + g_0 n}, t_{1c} = \frac{g_1 n}{1 + g_1 n}, t_{2c} = \frac{g_2 n}{1 + g_2 n}, \tag{3.12}$$

then we have:

$$f(t_{0c}, t_{1c}, t_{2c} | \boldsymbol{y}) \propto \frac{\prod_{j=0}^{2} t_{jc}^{-3/2} (1 - t_{jc})^{(p_j - 1)/2} exp\big(- n/(2 t_{jc}) \big)}{(1 - t_{0c} R^2_0 - t_{1c} R^2_1 - t_{2c} R^2_2)^{n/2}}, \tag{3.13}$$

where $R^2_0 = \boldsymbol{y}' \boldsymbol{P}_{X_0} \boldsymbol{y} / \boldsymbol{y}' \boldsymbol{y}$. With the simplified expressions in densities (3.11) and (3.13), the Bayesian estimators defined in equation (3.7) reduce to:

$$\tilde{\boldsymbol{\beta}}^B_i = \begin{pmatrix} \boldsymbol{\beta}^B_{i,0} \\ \boldsymbol{\beta}^B_{i,i} \end{pmatrix} = \begin{pmatrix} E(t_0 | \boldsymbol{y}_i, M_i)(\boldsymbol{X}_{0i}' \boldsymbol{X}_{0i})^{-1} \boldsymbol{X}_{0i}' \boldsymbol{y}_i \\ E(t_i | \boldsymbol{y}_i, M_i)(\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i' \boldsymbol{y}_i \end{pmatrix} = \begin{pmatrix} E(t_0 | \boldsymbol{y}_i, M_i) \hat{\boldsymbol{\beta}}^L_{i,0} \\ E(t_i | \boldsymbol{y}_i, M_i) \hat{\boldsymbol{\beta}}^L_{i,i} \end{pmatrix}, \tag{3.14}$$

where $\boldsymbol{\beta}^B_{i,0}$ and $\boldsymbol{\beta}^B_{i,i}$ indicate the Bayesian estimators for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ in M_i, respectively, $\hat{\boldsymbol{\beta}}^L_{i,0} = (\boldsymbol{X}_{0i}' \boldsymbol{X}_{0i})^{-1} \boldsymbol{X}_{0i}' \boldsymbol{y}_i$ and $\hat{\boldsymbol{\beta}}^L_{i,i} = (\boldsymbol{X}_i' \boldsymbol{X}_i)^{-1} \boldsymbol{X}_i' \boldsymbol{y}_i$ are the least squares estimators for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$.

Similarly, in M_c, the Bayesian estimator for $\tilde{\boldsymbol{\beta}}$ is denoted as $\tilde{\boldsymbol{\beta}}_c^B$ with:

$$\tilde{\boldsymbol{\beta}}_c^B = \begin{pmatrix} \boldsymbol{\beta}_{c,0}^B \\ \boldsymbol{\beta}_{c,1}^B \\ \boldsymbol{\beta}_{c,2}^B \end{pmatrix} = \begin{pmatrix} E(t_{0c}|\boldsymbol{y}, M_c)(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{y} \\ E(t_{1c}|\boldsymbol{y}, M_c)(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{y}_1 \\ E(t_{2c}|\boldsymbol{y}, M_c)(\boldsymbol{X}_2'\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\boldsymbol{y}_2 \end{pmatrix} = \begin{pmatrix} E(t_{0c}|\boldsymbol{y}, M_c)\hat{\boldsymbol{\beta}}_{c,0}^L \\ E(t_{1c}|\boldsymbol{y}, M_c)\hat{\boldsymbol{\beta}}_{c,1}^L \\ E(t_{2c}|\boldsymbol{y}, M_c)\hat{\boldsymbol{\beta}}_{c,2}^L \end{pmatrix}, \quad (3.15)$$

where $\boldsymbol{\beta}_{c,0}^B$ and $\boldsymbol{\beta}_{c,i}^B$ are the Bayesian estimators for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ in M_c, respectively, $\hat{\boldsymbol{\beta}}_{c,0}^L = (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{y}$ and $\hat{\boldsymbol{\beta}}_{c,i}^L = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i$ correspond to the least squares estimators for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$.

3.3.3 Conditional Information Asymptotic Analyses

Here, we establish asymptotic frequentist properties regarding $\tilde{\boldsymbol{\beta}}_i^B$ and $\tilde{\boldsymbol{\beta}}_c^B$ respectively. The asymptotic process under consideration is defined in Som et al. (2015) and Som et al. (2016) in the context of hyper-g prior. It describes a situation where the model is dominated by a typical group of coefficients or, equivalently, the size of one group of coefficients is much larger than the rest. Different from the regular asymptotic theories, it is referred to as the sample size n fixed and dimension p fixed asymptotic process and considered to address the conditional Lindley's paradox (CLP), which will occur if a dependent hyper-g prior is employed. Specifically, with the dependent version of hyper-g prior, in the comparison of a pair of nested models, the Bayes factor always chooses a smaller model if at least one of the common regression coefficients is relatively larger compared with additional coefficients in the bigger model. This phenomenon not only exists in hyper-g prior but also the ZS prior when it comes to the hypothesis testing. Since our primary interest lies in the independent ZS prior in this subsection, the CLP is no longer our concern, but how the defined sequence impact the Bayesian estimator in M_i and M_c in terms of estimation and whether the sequence has the same impact on M_i and M_c remain unstudied.

Recall that our models includes three groups of regression coefficients $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ and we perform two asymptotic analyses, where subsection 3.3.4 probes the case where the common $\boldsymbol{\beta}_0$ is dominant over the specific $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$ and 3.3.5 investigates the case where the specific $\boldsymbol{\beta}_1$ is dominant over the common $\boldsymbol{\beta}_0$ and the specific $\boldsymbol{\beta}_2$.

3.3.4 Dominant Common Coefficients

In this subsection, we consider the case where dominant variables are common coefficients $\boldsymbol{\beta}_0$ in the sense that the size of $\boldsymbol{\beta}_0$ is relatively large compared with $\boldsymbol{\beta}_i$. For M_c, we consider $\{L_c^{(k)}\}_{k=1}^{\infty}$, where each element $L_c^{(k)}$ represents the linear model with:

$$L_c^{(k)} = \{\boldsymbol{X}_0, \boldsymbol{\beta}_0^{(k)}, \boldsymbol{X}_1, \boldsymbol{\beta}_1, \boldsymbol{X}_2, \boldsymbol{\beta}_2, \boldsymbol{\epsilon}\}, \tag{3.16}$$

where we let $||\boldsymbol{\beta}_0^{(k)}||^2 \to \infty$ as $k \to \infty$ while $\{\boldsymbol{X}_0, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\epsilon}\}$ are held. Interested readers may refer to Som et al. (2016) for more details and discussions. Here, $\boldsymbol{\epsilon}$ is held in the sense that $\boldsymbol{\epsilon}$ remains the same for all k. Notice that this represents the situation where the likelihood is driven by one particular set of predictor variables. Naturally, the sequence for M_i, i=1,2 is $\{L_i^{(k)}\}_{k=1}^{\infty}$ with element:

$$L_i^{(k)} = \{\boldsymbol{X}_{0i}, \boldsymbol{\beta}_0^{(k)}, \boldsymbol{X}_i, \boldsymbol{\beta}_i, \boldsymbol{\epsilon}_i\}. \tag{3.17}$$

Lemma 3.1. *With $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \mathbf{0}$ and the defined sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$, as $k \to \infty$, $E(1/R_{0i}^{2(k)}|\boldsymbol{X}_{0i}, \boldsymbol{\beta}_0^{(k)}, \boldsymbol{X}_i, \boldsymbol{\beta}_i) \to 1$ with $R_{0i}^{2(k)} \in (0,1)$.*

Proof. See Appendix A.2.2. □

Remark 3.3. *By Lemma 3.1, since* $1/R_{0i}^{2(k)} > 1$*, as* $k \to \infty$*, we have* $E(|1/R_{0i}^{2(k)} - 1|) = E(1/R_{0i}^{2(k)} - 1) \to 0$*, and therefore* $1/R_{0i}^{2(k)} \xrightarrow{L_1} 1$*, which implies that* $1/R_{0i}^{2(k)}$ *converges to 1 in probability. By the continuous mapping theorem, we have* $R_{0i}^{2(k)}$ *converges to 1 in probability. A similar argument can be applied to* $R_0^{2(k)}, R_1^{2(k)}$ *and* $R_2^{2(k)}$*.*

Remark 3.4. *Several seminal papers have addressed the Lindley's paradox (Liang et al. (2008)) as* $R_{0i}^2 \to 1$*. However, the underlying sequence has been seldom explicitly addressed. To the best of our knowledge, Som et al. (2015) is the first to formally state the underlying sequence* $\{L_i^{(k)}\}$ *to explain* $R_{0i}^{(k)} \to 1$ *and utilize the sequence to demonstrate CLP. However, Som et al. (2015) utilized this type of convergence vaguely and other types of convergence are clearer such as converging in probability when it comes to model selection consistency or prediction consistency. Lemma 3.1 and Remark 3.3 first show that the convergence type with respect to the sequence* $\{L_i^{(k)}\}$ *refers to the convergence in probability.*

Lemma 3.2. *For* M_i*, with* $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$ *and the defined sequence* $\{L_i^{(k)}\}_{k=1}^{\infty}$ *in (3.17), as* $k \to \infty$ *and* $||\boldsymbol{\beta}_0^{(k)}||^2 \to \infty$*,* $\boldsymbol{\beta}_{i,0}^{B(k)} - \boldsymbol{\beta}_0^{(k)} \to (\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i$ *in probability if* $n_i - p_0 > 4$*. Similarly, for* M_c*, we have* $\boldsymbol{\beta}_{c,0}^{B(k)} - \boldsymbol{\beta}_0^{(k)} \to (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{\epsilon}$ *in probability.*

Proof. See Appendix A.2.3. □

The lemma indicates that, under our defined sequences, the difference between the Bayesian estimator for dominated variables and the true value converges in probability to a random variable, whose expectation is zero.

Now, we formally address the conditional asymptotic analyses with respect to the Bayesian estimators for $\boldsymbol{\beta}_0$ in M_i and M_c.

Theorem 3.2. *Suppose $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, and consider M_i and priors in (3.5), $n_i - p_0 > 4$, under the sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$ in (3.17), there exists a subsequence with respect to $\boldsymbol{\beta}_0^{(m_k)}$ such that if $||\boldsymbol{\beta}_0^{(m_k)}||^2 \to \infty$,*

(a) $\lim_{m_k \to \infty} E\big(\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}\big) = \boldsymbol{0}$;

(b) $\lim_{m_k \to \infty} E\left([\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}][\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}]'\right) = \sigma^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}$.

Consider M_c and priors in (3.6), under the sequence $\{L_c^{(k)}\}_{k=1}^{\infty}$ in (3.16), there exists a subsequence with respect to $\boldsymbol{\beta}_0^{(m_k)}$ such that if $||\boldsymbol{\beta}_0^{(m_k)}||^2 \to \infty$,

(c) $\lim_{m_k \to \infty} E\big(\boldsymbol{\beta}_{c,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}\big) = \boldsymbol{0}$;

(d) $\lim_{m_k \to \infty} E\left([\boldsymbol{\beta}_{c,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}][\boldsymbol{\beta}_{c,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}]'\right) = \sigma^2(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}$.

Proof. See Appendix A.2.5. □

Theorem 3.2 (a) and (c) reveal that the Bayesian estimators for both M_i and M_c are conditional asymptotically unbiased for the dominant common $\boldsymbol{\beta}_0$. Comparing Theorem 3.2 (b) and (d), the frequentist variances for $\boldsymbol{\beta}_{c,0}^{B(m_k)}$ in M_c is conditional asymptotically smaller than $\boldsymbol{\beta}_{i,0}^{B(m_k)}$ in M_i. Consequently, the MSEs for $\boldsymbol{\beta}_{c,0}^{B(m_k)}$ in M_c is also conditional asymptotically smaller than $\boldsymbol{\beta}_{i,0}^{B(m_k)}$ in M_i.

Next, we would establish the conditional asymptotic analysis with respect to the Bayesian estimators for the specific $\boldsymbol{\beta}_i$ in M_i and M_c.

Recall that in M_i, $E(t_i|\boldsymbol{y}_i)$ has the following form:

$$\frac{\int_0^1\int_0^1 t_0^{-\frac{3}{2}} t_i^{-\frac{1}{2}}(1-t_0)^{\frac{p_0-1}{2}}(1-t_i)^{\frac{p_i-1}{2}} e^{-\frac{n_i}{2t_0}-\frac{n_i}{2t_i}}(1-t_0R_{0i}^{2(k)}-t_iR_i^{2(k)})^{-\frac{n_i}{2}}dt_0dt_i}{\int_0^1\int_0^1 t_0^{-\frac{3}{2}} t_i^{-\frac{3}{2}}(1-t_0)^{\frac{p_0-1}{2}}(1-t_i)^{\frac{p_i-1}{2}} e^{-\frac{n_i}{2t_0}-\frac{n_i}{2t_i}}(1-t_0R_{0i}^{2(k)}-t_iR_i^{2(k)})^{-\frac{n_i}{2}}dt_0dt_i}. \quad (3.18)$$

In fact, due to $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, if $||\boldsymbol{\beta}_0^{(k)}|| \to \infty$, then $R_{0i}^{2(k)}, R_0^{2(k)} \to 1$ and $R_i^{2(k)} \to 0$. Consequently, $E(t_i|\boldsymbol{y}_i, M_i)$ reduces to $G(n_i; p_i)$ and $E(t_i|\boldsymbol{y}, M_c)$ reduces to $G(n; p_i)$, where

$$G(x; p_i) = \frac{\int_0^1 t_i^{-\frac{1}{2}}(1-t_i)^{\frac{p_i-1}{2}} \exp(-\frac{x}{2t_i})dt_i}{\int_0^1 t_i^{-\frac{3}{2}}(1-t_i)^{\frac{p_i-1}{2}} \exp(-\frac{x}{2t_i})dt_i}. \tag{3.19}$$

Notice that $G(n_i; p_i)$ and $G(n, p_i)$ do not depend on data $\boldsymbol{y}_i$ or $\boldsymbol{y}$ when $R_i^{2(k)} \to 0$. In the comparison of M_i and M_c, their only difference lies in the sample size. The following lemma states that $G(x; p_i)$ is nondecreasing with respect to x.

Lemma 3.3. *If $R_{0i}^{2(k)}$ or $R_0^{2(k)} \to 1$, we have $R_i^{2(k)} \to 0$ and $G(x; p_i)$ in (3.19) is increasing with respect to x. Consequently, $G(n_i; p_i) \le G(n; p_i)$ for $i = 1, 2$.*

Proof. See Appendix A.2.4. □

Theorem 3.3. *Suppose $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, and consider M_i and priors in (3.5), under the sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$, as $||\boldsymbol{\beta}_0^{(k)}||^2 \to \infty$,*

(a) $\lim_{k\to\infty} E\big(\boldsymbol{\beta}_{i,i}^{B(k)} - \boldsymbol{\beta}_i\big) = [G(n_i; p_i) - 1]\boldsymbol{\beta}_i$;

(b) $\lim_{k\to\infty} E\left([\boldsymbol{\beta}_{i,i}^{B(k)} - E(\boldsymbol{\beta}_{i,i}^{B(k)})][\boldsymbol{\beta}_{i,i}^{B(k)} - E(\boldsymbol{\beta}_{i,i}^{B(k)})]'\right) = \sigma^2 G^2(n_i; p_i)(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$.

Consider M_c and priors in (3.6), under the sequence $\{L_c^{(k)}\}_{k=1}^{\infty}$,

(c) $\lim_{k\to\infty} E\big(\boldsymbol{\beta}_{c,i}^{B(k)} - \boldsymbol{\beta}_i\big) = [G(n; p_i) - 1]\boldsymbol{\beta}_i$;

(d) $\lim_{k\to\infty} E\left([\boldsymbol{\beta}_{c,i}^{B(k)} - E(\boldsymbol{\beta}_{c,i}^{B(k)})][\boldsymbol{\beta}_{c,i}^{B(k)} - E(\boldsymbol{\beta}_{c,i}^{B(k)})]'\right) = \sigma^2 G^2(n; p_i)(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$.

Theorem 3.3 indicates that the Bayesian estimators for relatively undominant random variables or the specific $\boldsymbol{\beta}_i$ are asymptotically biased under the defined sequence and the magnitude of this biasness depends on $G(x;p_i)$. Additionally, compared with M_i, the Bayesian estimator of the undominated specific $\boldsymbol{\beta}_i$ in M_c has an asymptotically increased frequentist variance but a decreased bias. Therefore, the resulting MSEs for M_i and M_c depend on the trade-off of bias and variance.

By Theorems 3.2 and 3.3, the following result gives the comparison of MSEs in the limiting case.

Theorem 3.4. *Suppose $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$. Under the defined sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$ in M_i and $\{L_c^{(k)}\}_{k=1}^{\infty}$ in M_c, we have*

1. *For $\boldsymbol{\beta}_0$, the MSE in M_c is asymptotically smaller.*

2. *For $\boldsymbol{\beta}_i$, let $a = \boldsymbol{\beta}_i'\boldsymbol{\beta}_i, b = \sigma^2 tr((\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1})$, if $a/(a+b) > G(n;p_i) > G(n_i;p_i)$ or $a/(a+b) < G(n;p_i) < 2a/(a+b) - G(n_i;p_i)$, MSE in M_c is smaller. If $2a/(a+b) - G(n;p_i) < G(n_i;p_i) < a/(a+b) < G(n;p_i)$, MSE in M_c is larger.*

3.3.5 Dominant Specific Coefficients

In this subsection, we investigate the case where the size of specific coefficients play a dominant role in the model. Notice that the dominant coefficient is either $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$. Accordingly, the sequence for $M_i, i = 1, 2$ is $\{L_i^{(k)}\}_{k=1}^{\infty}$ with element:

$$L_i^{(k)} = \{\boldsymbol{X}_{0i}, \boldsymbol{\beta}_0, \boldsymbol{X}_i, \boldsymbol{\beta}_i^{(k)}, \boldsymbol{\epsilon}_i\}. \tag{3.20}$$

Assume $\|\boldsymbol{\beta}_i\|$ has a dominant effect, then the sequence for M_c is specified as:

$$L_c^{(k)} = \{\boldsymbol{X}_0, \boldsymbol{\beta}_0, \boldsymbol{X}_i, \boldsymbol{\beta}_i^{(k)}, \boldsymbol{X}_j, \boldsymbol{\beta}_j, \boldsymbol{\epsilon}\}, \tag{3.21}$$

where $i+j=3, i,j=1,2$, and we let $||\boldsymbol{\beta}_i^{(k)}||^2 \to \infty$ as $k \to \infty$. Note that there is no dominant effect in M_j and no sequence is associated with it. Results for M_i with $||\boldsymbol{\beta}_i||$ being dominated can be easily extended to M_j with $||\boldsymbol{\beta}_j||$ being dominant.

Again, under the orthogonality assumption $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \mathbf{0}$, recall that $R_0^{(k)} + R_1^{(k)} + R_2^{(k)} \leq 1$ and therefore, if $||\boldsymbol{\beta}_i^{(k)}|| \to \infty$ as $k \to \infty$, $R_i^{(k)} \to 1, R_0^{(k)} \to 0, R_j^{(k)} \to 0$ for M_c and $R_i^{(k)} \to 1, R_{0i}^{(k)} \to 0$ for M_i. We would directly present the corresponding results since detailed proofs have been provided for dominant common regression coefficients in Subsection 3.3.4 and they can be tailored to dominant specific regression coefficients without efforts.

Theorem 3.5. *Suppose $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \mathbf{0}$, and consider M_i and priors in (3.5), assume $n_i - p_i > 4$, under the sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$ in (3.20), there exists a subsequence with respect to $\boldsymbol{\beta}_i^{(m_k)}$ such that if $||\boldsymbol{\beta}_i^{(m_k)}|| \to \infty$, we have:*

(a) $\lim_{m_k\to\infty} E\big(\boldsymbol{\beta}_{i,i}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}\big) = \mathbf{0}$;

(b) $\lim_{m_k\to\infty} E\left([\boldsymbol{\beta}_{i,i}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}][\boldsymbol{\beta}_{i,i}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}]'\right) = \sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$.

Consider M_c and priors in (3.6), under the sequence $\{L_c^{(k)}\}_{k=1}^{\infty}$ in (3.21), there exists a subsequence with respect to $\boldsymbol{\beta}_i^{(m_k)}$ such that if $||\boldsymbol{\beta}_i^{(m_k)}|| \to \infty$,

(c) $\lim_{m_k\to\infty} E\big(\boldsymbol{\beta}_{c,i}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}\big) = \mathbf{0}$;

(d) $\lim_{m_k\to\infty} E\left([\boldsymbol{\beta}_{c,i}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}][\boldsymbol{\beta}_{c,}^{B(m_k)} - \boldsymbol{\beta}_i^{(m_k)}]'\right) = \sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$.

Notice that when the specific regression coefficients are driven predictors, M_i provides Bayesian estimates as good as M_c for $\boldsymbol{\beta}_i^{(m_k)}$ in terms of the conditional asymptotic bias, frequentist variance and MSE. For the common $\boldsymbol{\beta}_0$, we have the following results.

Theorem 3.6. *Suppose $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \boldsymbol{0}$, and consider M_i and priors in (3.5) with the sequence $\{L_i^{(k)}\}_{k=1}^{\infty}$ in (3.20), if $||\boldsymbol{\beta}_i^{(k)}|| \to \infty$, we have:*

(a) $\lim_{k\to\infty} E\big(\boldsymbol{\beta}_{i,0}^{B(k)} - \boldsymbol{\beta}_0\big) = (G(n_i; p_0) - 1)\boldsymbol{\beta}_0$;

(b) $\lim_{k\to\infty} E\left([\boldsymbol{\beta}_{i,0}^{B(k)} - E(\boldsymbol{\beta}_{i,0}^{B(k)})][\boldsymbol{\beta}_{i,0}^{B(k)} - E(\boldsymbol{\beta}_{i,0}^{B(k)})]'\right) = \sigma^2(G(n_i; p_0))^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}$.

Consider M_c and priors in (3.6), under the sequence $\{L_c^{(k)}\}_{k=1}^{\infty}$ in (3.21), we have:

(c) $\lim_{k\to\infty} E\big(\boldsymbol{\beta}_{c,0}^{B(k)} - \boldsymbol{\beta}_0\big) = (G(n; p_0) - 1)\boldsymbol{\beta}_0$;

(d) $\lim_{k\to\infty} E\left([\boldsymbol{\beta}_{c,0}^{B(k)} - E(\boldsymbol{\beta}_{c,0}^{B(k)})][\boldsymbol{\beta}_{c,0}^{B(k)} - E(\boldsymbol{\beta}_{c,0}^{B(k)})]'\right) = \sigma^2(G(n; p_0))^2(\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}$.

Since the logic is similar to Section 3.3.4, $G(x; p_0)$ is the same function as equation (3.19) with p_0 replacing p_i and t_0 replacing t_i. It follows immediately from Theorem 3.5 that the asymptotic MSE remains the same in terms of $\boldsymbol{\beta}_i$ for M_i and M_c. However, for the common $\boldsymbol{\beta}_0$, Theorem 3.6 indicates that M_c has an asymptotic larger bias and uncertain change in the asymptotic frequentist variance. Therefore, the comparison of asymptotic MSE with respect to the common $\boldsymbol{\beta}_0$ between M_i and M_c is less clear.

3.3.6 The Effective Sample Size (TESS)

This section mainly targets at the application of TESS to our framework. We would use Berger et al. (2014) as a major reference and cite the results directly for the linear regression.

Consider Case 3 in Section 3.2 from Berger et al. (2014), for a simple linear regression (SLR) problem, $Y_i = X_i\beta + \epsilon_i, \epsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2), i = 1, \cdots, n$, with the design matrix $\boldsymbol{X} = (X_1, \cdots, X_n)'$ and $X_i \sim N(k, 1)$ randomly. With facts $E(\sum_{i=1}^{n} X_i^2) =$

$n(k^2+1)$ and $E(\max|X_k|) \approx |k| + (2\log n - 3)^{1/2}$ for large n, the effective sample size is

(a) for $k = 0$ and n is large,

$$n^e \approx \frac{n}{2\log n - 3}; \tag{3.22}$$

(b) for k large compared with $log(n)$, $n^e \approx nk^2/k^2 = n$.

Remark 3.1. *Replacing $N(k,1)$ with $N(k,\tau^2)$ leads to the same TESS.*

Then, we generalize results for SLR to a multiple linear regression problem under some conditions. Consider $\boldsymbol{Y} = \boldsymbol{X\beta} + \epsilon$, where $\boldsymbol{Y} = (Y_1, \cdots, Y_n)' \in \mathbb{R}^n$, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)' \in \mathbb{R}^p$, $\boldsymbol{\epsilon} = (\epsilon_1, \cdots, \epsilon_n)'$ with $\epsilon_i \sim N(0, \sigma^2)$, $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_p) \in \mathbb{R}^{n\times p}$ is the design matrix and the jth column $\boldsymbol{x}_j = (X_{1j}, \cdots, X_{nj})$ is a vector of realizations from a random variable X_j. Consider any scalar linear transformation $\xi = \boldsymbol{\nu}'\boldsymbol{\beta}$ and assume that X_j and $X_{j'}$ are independent and identically distributed with mean 0, we then derive its TESS. Follow the basic expression of TESS in equation (2.1) from Berger et al. (2014), TESS denoted as n^e has the following form:

$$n^e = \frac{|\boldsymbol{\nu}|^2}{\boldsymbol{\nu}'\boldsymbol{C}\boldsymbol{F}_2^{-1}\boldsymbol{C}\boldsymbol{\nu}}, \tag{3.23}$$

where $\boldsymbol{F}_2 = \boldsymbol{X}'\boldsymbol{X}/\sigma^2$ is Fisher information matrix for $\boldsymbol{\beta}$ and $\boldsymbol{C} = \mathrm{diag}(\max_j |X_{ij}|/\sigma)$. Note that TESS in (3.23) is free from scales of σ and design matrix $\boldsymbol{X}$. $\boldsymbol{C}\boldsymbol{F}_2^{-1}\boldsymbol{C}$ in (3.23) can be evaluated through its inverse $\boldsymbol{K} = (\boldsymbol{C}\boldsymbol{F}_2^{-1}\boldsymbol{C})^{-1}$ for convenience. Its diagonal element k_{jj} is:

$$k_{jj} = \frac{\sum_i^n X_{ij}^2}{(\max_i |X_{ij}|)^2}, \tag{3.24}$$

and its off-diagonal element $k_{jj'}$ is:

$$k_{jj'} = \frac{\sum_i X_{ij} X_{ij'}}{\max_i |X_{ij}| \max_i |X_{ij'}|}, j \neq j' \tag{3.25}$$

Since $E(\sum_i X_{ij} X_{ij'}) \approx cov(X_j, X_{j'}) = 0, k_{jj'} \approx 0$. Then, $\boldsymbol{K} \approx diag(k_{11}, \cdots, k_{pp})$ and $\boldsymbol{C}\boldsymbol{F}_2^{-1}\boldsymbol{C} \approx diag(k_{11}^{-1}, \cdots, k_{pp}^{-1})$.

Remark 3.5. *Notice that:*

$$k_{jj}^{-1} \approx (E \max_i |X_{ij}|)^2 / E(\sum_i^n X_{ij}^2) \text{ and } n_j^e \approx (E \max_i |X_{ij}|)^2 / E(\sum_i^n X_{ij}^2),$$

since X_j and $X_{j'}$ are independently and identically distributed, $n_1^e = \cdots = n_p^e$. For example, if $X_j \sim N(0, \sigma^2)$, by (3.22), $n_1^e = \cdots = n_p^e = n/(2log(n) - 3) = n^e$.

Next, we focus on the application of TESS to the ZS prior. For a transformation $\boldsymbol{\xi} = \boldsymbol{V}\boldsymbol{\beta}$, such that $\sigma^2 \boldsymbol{V}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{V} = \boldsymbol{D}$, where $\boldsymbol{D}$ is a diagonal matrix with diagonal elements d_i. As recommended by Berger et al. (2014), the sample size n in the ZS prior should be replaced by the effective sample size for each coordinate of $\boldsymbol{\xi}$ and the prior for $\boldsymbol{\xi}$ has the following form $\boldsymbol{\xi} \sim N_p(0, g\text{diag}(n_1^e d_1, \cdots, n_p^e d_p))$. When X_j and $X_{j'}$ are independently and identically distributed with $E(X_j) = E(X_{j'}) = 0$, by Remark 3.5, it reduces to $\boldsymbol{\xi} \sim N_{p_1}(0, gn^e \text{diag}(d_1, \cdots, d_{p_1}))$.

Since $(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$ by spectral decomposition, where $\boldsymbol{P}$ is an orthogonal matrix comprising of its eigenvectors and $\boldsymbol{D}$ is a diagonal matrix comprising of eigenvalues, if we set $\boldsymbol{V} = \boldsymbol{P}'$, the equivalent prior for $\boldsymbol{\beta}$ with TESS is:

$$\boldsymbol{\beta} \sim N_p(0, g\sigma^2 n^e (\boldsymbol{X}'\boldsymbol{X})^{-1}), g \sim IG(0.5, 0.5).$$

Compared with ZS prior without TESS, the only difference lies in the scale parameter n and n^e.

Remark 3.6. *Notice that TESS highly depends on the distribution and the form of design matrix. If each column of the design matrix is independently identically distributed, TESS for any linear combination of regression coefficients remain the same. Admittedly, this assumption may not be easily achieved in practice. When the design matrix does not arise from normal distribution or has more complicated generation mechanisms (e.g., X_i and X_j are neither independent nor identical; non-standard distributions; no information about the distribution of the design matrix), TESS can still be computed numerically even though an approximation or explicit form is not available.*

At last, we formally state the asymptotic results for ZS prior with TESS in comparison of M_i and M_c. For regression coefficients of interests, notice that the corresponding Bayesian estimators in either (3.14) for M_i or (3.15) for M_c can be written as the product of a posterior expectation of g and a least squares estimate. Substituting the sample size with the effective sample size would only impact Bayesian estimators through the posterior expectation of g. For example, for common coefficients $\boldsymbol{\beta}_0$ in M_i, $\boldsymbol{\beta}^B_{i,0} = F(n_i, p_0, R^2_{0i})\hat{\boldsymbol{\beta}}^L_{i,0}$, where $F(n_i, p_0, R^2_{0i}) = E(t_0|\boldsymbol{y}_i)$ with $t_0 = g_0 n_i/(1+g_0 n_i)$. Therefore, all the procedures in Section 3.3.4 and 3.3.5 can be applied step by step with a simple change of replacing n_i in the prior with the effective sample size. This rationale also applies to the specific coefficients $\boldsymbol{\beta}_i$ and M_c.

With Remark 3.5, ZS prior conditioned on hyperparameter with TESS adjustment for $\boldsymbol{\beta}_0$ and $\boldsymbol{\beta}_i$ are described below. For M_i, we have :

$$\boldsymbol{\beta}_0|g_0, \sigma^2 \sim N_{p_0}(0, g_0 n^e_{i,0}\sigma^2(\boldsymbol{X}'_{0i}\boldsymbol{X}_{0i})^{-1}), \tag{3.26}$$

$$\boldsymbol{\beta}_i \mid g_i, \sigma^2 \sim N_{p_i}(0, g_i n^e_i \sigma^2(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}), \tag{3.27}$$

where $n^e_{i,0}$ indicates TESS for common coefficients $\boldsymbol{\beta}_0$ in M_i, and n^e_i indicates TESS for specific coefficients $\boldsymbol{\beta}_i$ in M_i and M_c. For M_c, we have:

$$\boldsymbol{\beta}_0|g_{0c}, \sigma^2 \sim N_{p_0}(0, g_{0c} n^e_{c,0}\sigma^2(\boldsymbol{X}'_0\boldsymbol{X}_0)^{-1}), \tag{3.28}$$

$$\boldsymbol{\beta}_i \mid g_{ic}, \sigma^2 \sim N_{p_i}(0, g_{ic} n^e_i \sigma^2(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}), \tag{3.29}$$

where $n^e_{c,0}$ indicates TESS for common coefficients $\boldsymbol{\beta}_0$ in M_c. Priors in (3.26) to (3.29) shows that the resulting TESS for $\boldsymbol{\beta}_i$ remains exactly the same due to the same design matrices in M_i and M_c while TESS differs for $\boldsymbol{\beta}_0$ in individual and combined model. Accordingly, the asymptotic results regarding $\boldsymbol{\beta}_i$ are the same M_i and M_c while $\boldsymbol{\beta}_0$ varies. In an analogy to Theorems 3.2 to 3.6, the asymptotic analyses for ZS prior with TESS adjustment is described below.

Theorem 3.7. *When the common coefficient $\boldsymbol{\beta}_0$ is dominant, the asymptotic bias and covariance regarding $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ for M_i are the same as M_c, which is $[G(n^e_i; p_i) - 1]\boldsymbol{\beta}_i$ and $\sigma^2 G^2(n^e_i; p_i)(\boldsymbol{X}'_i\boldsymbol{X}_i)^{-1}$. The asymptotic bias regarding $\boldsymbol{\beta}_0$ for M_i and M_c are both zero, the asymptotic variance regarding $\boldsymbol{\beta}_0$ for M_i and M_c are $\sigma^2(\boldsymbol{X}'_{0i}\boldsymbol{X}_{0i})^{-1}$ and $\sigma^2(\boldsymbol{X}'_{01}\boldsymbol{X}_{01} + \boldsymbol{X}'_{02}\boldsymbol{X}_{02})^{-1}$, respectively. When the specific coefficient $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$ is dominant, the asymptotic bias and covariance matrix $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$ in M_i are the same as*

Table 3.1: Summary of Theorems 3.2 - 3.7. "S", "L","UNC" and "UNS" indicate that the statistics is smaller, larger, unchanged and unsure in M_c compared with M_i.

Dominant Term	Parameter	Statistics	ZS	TESS
Common $\boldsymbol{\beta}_0$	$\boldsymbol{\beta}_0$	Bias	UNC	UNC
		VAR_F	S	S
		MSE	S	S
	$\boldsymbol{\beta}_i$	Bias	S	UNC
		VAR_F	L	UNC
		MSE	UNS	UNC
Specific $\boldsymbol{\beta}_i$	$\boldsymbol{\beta}_0$	Bias	S	S
		VAR_F	UNS	UNS
		MSE	UNS	UNS
	$\boldsymbol{\beta}_i$	Bias	UNC	UNC
		VAR_F	UNC	UNC
		MSE	UNC	UNC

M_c, which are zero and $\sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$. The asymptotic bias and covariance matrix for $\boldsymbol{\beta}_0$ for M_i are $[G(n^e_{i,0};p_0)-1]\boldsymbol{\beta}_0$ and $\sigma^2(G(n^e_{i,0};p_0))^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}$, respectively, and for M_i are $[G(n^e_{c,0};p_0)-1]\boldsymbol{\beta}_0$ and $\sigma^2(G(n^e_{c,0};p_0))^2(\boldsymbol{X}_{01}'\boldsymbol{X}_{01}+\boldsymbol{X}_{02}'\boldsymbol{X}_{02})^{-1}$, respectively.

Theorem 3.7 indicates that the overall asymptotic frequentist variance and MSE are equal or smaller in M_c. In Table 3.1 is a summary of key conclusions from Theorems 3.2 to 3.7, which hopefully serves as a quick reference for interested readers.

3.3.7 Sampling Distributions

For M_i, the following distributions are applied to do the computation:

1. Sample $\sigma^2|g_0, g_i, \boldsymbol{y}_i \sim IG\Big(n_i/2,\ \boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i/2\Big)$;

2. Sample $g_0|\sigma^2, g_i, \boldsymbol{\beta}_0, \boldsymbol{y}_i \sim IG\Big((p_0+1)/2,\ \boldsymbol{\beta}_0'\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i}\boldsymbol{\beta}_0/(2n_i\sigma^2) + 1/2\Big)$;

3. Sample $g_i|\sigma^2, g_0, \boldsymbol{\beta}_i, \boldsymbol{y}_i \sim IG\Big((p_i+1)/2, \boldsymbol{\beta}_i'\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{\beta}_i/(2n_i\sigma^2) + 1/2\Big)$;

4. Sample $\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_0, g_i, \boldsymbol{y}_i \sim N_{p_I}\Big((\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i, \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\Big)$.

For M_c, the following distributions are used to do the computation:

1. Sample $\sigma^2|g_{0c}, g_{1c}, g_{2c}, \boldsymbol{y} \sim IG\Big(n_T/2, \boldsymbol{y}'\big[\boldsymbol{I}_{n_T} - \tilde{\boldsymbol{X}}(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{C}^{-1})^{-1}\tilde{\boldsymbol{X}}'\big]\boldsymbol{y}/2\Big)$;

2. Sample $g_{0c}|\sigma^2, g_{ic}, \boldsymbol{\beta}_0, \boldsymbol{y} \sim IG\Big((p_0+1)/2, \boldsymbol{\beta}_0'\boldsymbol{X}_0'\boldsymbol{X}_0\boldsymbol{\beta}_0/(2n_T\sigma^2) + 1/2\Big)$;

3. Sample $g_{ic}|\sigma^2, g_{0c}, \boldsymbol{\beta}_i, \boldsymbol{y} \sim IG\Big((p_i+1)/2, \boldsymbol{\beta}_i'\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{\beta}_i/(2n_T\sigma^2) + 1/2\Big), i = 1, 2,$;

4. Sample $\tilde{\boldsymbol{\beta}}|\sigma^2, g_{0c}, g_{1c}, g_{2c}, \boldsymbol{y} \sim N_{p_T}\Big((\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{C}^{-1})^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{y}, \sigma^2(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} + \boldsymbol{C}^{-1})^{-1}\Big)$.

Remark 3.7. *An alternative Gibbs sampling algorithm for M_i is as below:*

1. Sample $(g_0|g_i, \boldsymbol{y}_i)$ with ratio-of-uniform by

$$\pi(g_0|g_i, \boldsymbol{y}_i) \propto \frac{g_0^{-\frac{3}{2}}\exp(-\frac{1}{2g_0})|\boldsymbol{C}_i\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{I}_{p_0+p_i}|^{-\frac{1}{2}}}{\Big\{\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i\Big\}^{\frac{n_i}{2}}};$$

2. Sample $(g_i|g_0, \boldsymbol{y}_i)$ with ratio-of-uniform by

$$\pi(g_i|g_0, \boldsymbol{y}_i) \propto \frac{g_i^{-\frac{3}{2}}\exp(-\frac{1}{2g_i})|\boldsymbol{C}_i\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{I}_{p_0+p_i}|^{-\frac{1}{2}}}{\Big\{\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i\Big\}^{\frac{n_i}{2}}};$$

3. Sample $(\sigma^2|g_0, g_i, \boldsymbol{y}_i) \sim IG\Big(n_i/2, \boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i/2\Big)$;

4. Sample $(\tilde{\boldsymbol{\beta}}_i|\sigma^2, g_0, g_i, \boldsymbol{y}_i) \sim N_{p_I}\big((\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i, \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\big)$.

We show the derivation in Appendix (A.2.6). This sampling procedure enables us to sample or analyze directly from the joint distribution of $(g_0, g_i|\boldsymbol{y}_i)$ *and provide better mixing compared with the one we used. However, it takes more time to compute. For example, to obtain 1000 samples, this sampling procedure takes 24 minutes while the method we used only takes 2 minutes.*

Remark 3.8. *In fact, given* M_i *or* M_c*, our case is a special case of Min and Sun (2016), where they considered the independent ZS prior in the linear model with grouped covariates. Without loss of generality, we only present results for* M_1 *as an illustration. Under the commutativity assumption of block projection matrices, the marginal density* $m(\boldsymbol{y}_1|g_0, g_1)$ *is proportional to:*

$$m(\boldsymbol{y}_1|g_0, g_1) \propto (1+g_0)^{-\frac{p_0}{2}}(1+g_1)^{-\frac{p_1}{2}}(1+g_0+g_1)^{-\frac{p_2}{2}}\Big(\boldsymbol{y}_1'\Big(\boldsymbol{I}_{n_1} - \frac{g_0}{1+g_0}\boldsymbol{P}_0 - \frac{g_1}{1+g_1}\boldsymbol{P}_1 + \left(\frac{g_0g_1}{(1+g_0)(1+g_0+g_1)} + \frac{g_0g_1}{(1+g_1)(1+g_0+g_1)}\right)\boldsymbol{P}_0\boldsymbol{P}_1\Big)\boldsymbol{y}_1\Big)^{-\frac{n_1}{2}},$$

where $p_0 =$*rank*$(\boldsymbol{P}_0(\boldsymbol{I}_n - \boldsymbol{P}_1))$, $p_1 =$*rank*$(\boldsymbol{P}_1(\boldsymbol{I}_n - \boldsymbol{P}_0))$ *and* $p_2 =$*rank*$(\boldsymbol{P}_0\boldsymbol{P}_1)$*. Here, for convenience in notations, we set* $\boldsymbol{P}_0 = \boldsymbol{P}_{X_{01}}$ *and* $\boldsymbol{P}_1 = \boldsymbol{P}_{X_1}$*. If we further assume the orthognality of* $\boldsymbol{X}_0'\boldsymbol{X}_1 = \boldsymbol{0}$*, then* $\boldsymbol{P}_0\boldsymbol{P}_1 = \boldsymbol{P}_1\boldsymbol{P}_0 = \boldsymbol{0}$ *with* $p_0 =$*rank*$(\boldsymbol{P}_0) =$*rank*$(\boldsymbol{X}_0)$, $p_1 =$*rank*$(\boldsymbol{P}_1) =$*rank*$(\boldsymbol{X}_1)$*, and* $p_2 = 0$*. The marignal density is simplified into:*

$$m(\boldsymbol{y}_1|g_0, g_1) \propto (1+g_0)^{-\frac{p_0}{2}}(1+g_1)^{-\frac{p_1}{2}}\Big(\boldsymbol{y}_1'\Big(\boldsymbol{I}_n - \frac{g_0}{1+g_0}\boldsymbol{P}_0 - \frac{g_1}{1+g_1}\boldsymbol{P}_1\Big)\boldsymbol{y}_1\Big)^{-\frac{n_1}{2}}, \quad (3.30)$$

which is exactly the same as our case if we replace g_0, g_1 *with* g_0n_1, g_1n_1 *as we adopt a slightly different parameterization for the prior on regression coefficients.*

Proof. See Appendix A.2.7. □

3.4 Numerical Analyses

This section aims at investigating the relative performances of M_i and M_c from the estimation perspective through simulation studies.

3.4.1 Model Comparison of M_i and M_c

We consider two sets of parameters for the regression coefficients with $p_0 = p_1 = p_2 = 3$ as below:

- Set 1: $\boldsymbol{\beta}_0 = (10, 5, 1)'$, $\boldsymbol{\beta}_1 = (0.1, 0.2, 0.1)'$, $\boldsymbol{\beta}_2 = (0.2, 0.1, 0.1)'$, where the common $\boldsymbol{\beta}_0$ is dominant over the specific $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in size;
- Set 2: $\boldsymbol{\beta}_0 = (0.1, 0.2, 0.1)'$, $\boldsymbol{\beta}_1 = (10, 5, 1)'$, $\boldsymbol{\beta}_2 = (0.2, 0.1, 0.1)'$, where the specific $\boldsymbol{\beta}_1$ is dominant over the common $\boldsymbol{\beta}_0$ and the specific $\boldsymbol{\beta}_2$ in size.

For the design matrices, we consider orthogonal $\boldsymbol{X}_{0i}'\boldsymbol{X}_i \neq \mathbf{0}$ and non-orthogonal design $\boldsymbol{X}_{0i}'\boldsymbol{X}_i = \mathbf{0}$ with $(\boldsymbol{X}_{0i}, \boldsymbol{X}_i)$ generated from the normal distribution $N(0, 3)$ [Design 1] and uniform distribution $Unif(-3, 3)$ [Design 2]. For priors, we consider the regular ZS prior as well as TESS prior. We set $\sigma^2 = 1$ and $n_1 = n_2 = 10$. For each combination of coefficients and design matrices, we collect frequentist properties of the Bayesian estimator including its sampling variance, bias and MSE. Each Bayesian estimator is computed through 20,000 samples with 10,000 burn-ins. The frequentist properties are calculated with 500 data sets.

Tables 3.2 - 3.5 presents the $Bias$, VAR_F, and MSE for each grouped parameter. $Bias$ is the summation of absolute value of bias for each element in $\boldsymbol{\beta}_j, j = 0, 1, 2$ and describes the overall absolute difference between the expected value of the Bayesian estimator and its true value. VAR_F shows the overall frequentist variance of the Bayesian estimator for $\boldsymbol{\beta}_j$, which is the summation of diagonal elements of its sampling covariance matrix. Similarly, MSE is reported in groups. The bold number indicates M_c has a smaller value.

Our simulation results consolidate theorems in Subsections 3.3.4 and 3.3.5. Several main findings are summarized as follows. First, when the design matrices are non-orthogonal $\boldsymbol{X}'_{0i}\boldsymbol{X}_i \neq \mathbf{0}$, M_c uniformly outperform M_i in terms of VAR_F and MSE despite of the prior specifications, the distributions of the design matrices and the dominant coefficients. This is within our expectation because it not only enables information borrowing between data sources but also the design matrices. Second, no matter $\boldsymbol{\beta}_0$ or $\boldsymbol{\beta}_1$ is dominant, M_c indicates uniformly better estimations for the common $\boldsymbol{\beta}_0$ than $\boldsymbol{\beta}_1$ or $\boldsymbol{\beta}_2$ across different distributions for design matrices, which is reasonable since $\boldsymbol{\beta}_0$ is shared by two data sources and therefore more information is available for the estimation. Third, overall, ZS and TESS show similar behaviors across all the combinations, especially when the design matrices are from the normal distribution. When comparing M_i and M_c, TESS shows uniformly smaller VAR_F and MSE for $\boldsymbol{\beta}_i$ while ZS is less stable. Also, TESS shows slightly better or equivalent performances in terms of smaller VAR_F and MSE compared with ZS in most cases. At last, we may find that the Bayesian estimator is less biased for the dominant coefficients, which echoes our theoretical results as the Bayesian estimator is asymptotically unbiased for the dominant coefficients.

Tables 3.7 - 3.6 presents the percentages of M_c winning over M_i among 500 simulations in terms of a smaller posterior variance. We have several main findings. First, in general, it is more likely to obtain a smaller posterior variance in M_c for the non-orthogonal design matrix despite of parameters, priors, distributions of the design matrix and dominant coefficients. Second, compared with ZS, TESS shows a smaller posterior variance under an orthogonal design despite of parameters or distributions of the design matrix. It also tends to have a smaller posterior variance for the non-orthogonal design. For example, when the design matrix is from the uniform distribution, TESS outperforms ZS in 7 out of 8 comparisons. Third, common coefficients $\boldsymbol{\beta}_0$ presents higher percentages compared with the specific coefficients despite of other factors. For example, $\boldsymbol{\beta}_0$ reaches at least 90% in 30 out of 32 comparisons with the highest 99.6% while $\boldsymbol{\beta}_i$ only reaches at least 70% in 20 out of 32 comparisons with the highest 97.8%. We also notice that the percentage of smaller posterior variance in M_c in terms of $\boldsymbol{\beta}_2$ is quite low in the comparison of M_2 and M_c when the specific $\boldsymbol{\beta}_1$ is dominant, which is within our expectation and more explanations are offered in Section 3.6.

3.4.2 Sensitivity Analyses for M_c

To evaluate the relative performance of M_c compared with the golden standard M_{gs} in (2.36), we perform sensitivity analyses of M_c under the independent ZS priors in Section 3.3.1. As Section 3.4.1 adopted a balanced design regarding sample size ($n_1 = n_2$) and dimension of coefficients ($p_0 = p_1 = p_2$), we additionally consider the imbalanced design to serve as a complement. We also consider the case where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are dominant over $\boldsymbol{\beta}_0$, which has not been considered in the previous section.

Table 3.2: Comparisons of M_i and M_c under Design 1 and Set 1

Parameters	Design	Orthogonal			Non-orthogonal		
ZS	Statistics	M_1	M_2	M_c	M_1	M_2	M_c
$\boldsymbol{\beta}_0$	$Bias$	0.0301	0.0151	**0.0127**	0.1142	0.0585	**0.0180**
	VAR_F	0.1116	0.1573	**0.0531**	0.4836	0.1867	**0.0884**
	MSE	0.1119	0.1574	**0.0532**	0.4887	0.1882	**0.0885**
$\boldsymbol{\beta}_1$	$Bias$	0.0360	-	**0.0204**	0.0504	-	**0.0473**
	VAR_F	0.1476	-	0.1681	0.1534	-	**0.1532**
	MSE	0.1482	-	0.1683	0.1554	-	**0.1549**
$\boldsymbol{\beta}_2$	$Bias$	-	0.0645	**0.0404**	-	0.0925	**0.0570**
	VAR_F	-	0.1337	0.1519	-	0.1181	0.1246
	MSE	-	0.1355	0.1527	-	0.1211	0.1260
TESS							
$\boldsymbol{\beta}_0$	$Bias$	0.0263	0.0430	**0.0131**	0.1359	0.0802	**0.0504**
	VAR_F	0.1558	0.1722	**0.0591**	0.4622	0.1754	**0.0773**
	MSE	0.1560	0.1729	**0.0592**	0.4687	0.1781	**0.0783**
$\boldsymbol{\beta}_1$	$Bias$	0.0959	-	0.0976	0.0816	-	0.1031
	VAR_F	0.1100	-	**0.1096**	0.1126	-	**0.0992**
	MSE	0.1133	-	**0.1131**	0.1175	-	**0.1059**
$\boldsymbol{\beta}_2$	$Bias$	-	0.1141	0.1144	-	0.1325	**0.1204**
	VAR_F	-	0.1441	**0.1432**	-	0.0896	**0.0812**
	MSE	-	0.1492	**0.1483**	-	0.0960	**0.0873**

Table 3.3: Comparisons of M_i and M_c under Design 1 and Set 2

Parameters	Design	Orthogonal			Non-orthogonal		
ZS	Statistics	M_1	M_2	M_c	M_1	M_2	M_c
$\boldsymbol{\beta}_0$	$Bias$	0.0809	0.0518	**0.0483**	0.1040	0.0697	**0.0643**
	VAR_F	0.0782	0.1151	**0.0430**	0.3355	0.1256	**0.0688**
	MSE	0.0807	0.1163	**0.0440**	0.3416	0.1291	**0.0705**
$\boldsymbol{\beta}_1$	$Bias$	0.0528	-	**0.0499**	0.1245	-	**0.0809**
	VAR_F	0.2150	-	**0.2147**	0.2133	-	**0.1925**
	MSE	0.2161	-	**0.2156**	0.2213	-	**0.1954**
$\boldsymbol{\beta}_2$	$Bias$	-	0.0584	**0.0407**	-	0.0839	**0.0542**
	VAR_F	-	0.1389	0.1516	-	0.1145	0.1207
	MSE	-	0.1405	0.1525	-	0.1172	0.1221
TESS							
$\boldsymbol{\beta}_0$	$Bias$	0.1176	0.0888	**0.0871**	0.1481	0.1044	0.1133
	VAR_F	0.0618	0.0940	**0.0351**	0.2490	0.0938	**0.0482**
	MSE	0.0670	0.0972	**0.0381**	0.2590	0.1011	**0.0535**
$\boldsymbol{\beta}_1$	$Bias$	0.0541	-	**0.0509**	0.1474	-	**0.1139**
	VAR_F	0.2151	-	**0.2147**	0.1957	-	**0.1825**
	MSE	0.2163	-	**0.2156**	0.2066	-	**0.1885**
$\boldsymbol{\beta}_2$	$Bias$	-	0.1101	0.1223	-	0.1191	**0.1167**
	VAR_F	-	0.1037	**0.0943**	-	0.0870	**0.0774**
	MSE	-	0.1086	**0.1003**	-	0.0926	**0.0830**

Table 3.4: Comparisons of M_i and M_c under Design 2 and Set 1

Parameters	Design	Orthogonal			Non-orthogonal		
ZS	Statistics	M_1	M_2	M_c	M_1	M_2	M_c
$\boldsymbol{\beta}_0$	$Bias$	0.0098	0.0207	**0.0125**	0.1288	0.0787	**0.0091**
	VAR_F	0.0296	0.0457	**0.0135**	0.0491	0.0835	**0.0288**
	MSE	0.0297	0.0458	**0.0136**	0.0556	0.0856	**0.0289**
$\boldsymbol{\beta}_1$	$Bias$	0.1017	-	**0.0792**	0.2488	-	**0.0660**
	VAR_F	0.2021	-	0.2313	0.0750	-	0.0761
	MSE	0.2058	-	0.2336	0.0966	-	**0.0777**
$\boldsymbol{\beta}_2$	$Bias$	-	0.0869	**0.0695**	-	0.0983	**0.0264**
	VAR_F	-	0.3500	0.3929	-	0.0592	**0.0498**
	MSE	-	0.3526	0.3947	-	0.0627	**0.0502**
TESS							
$\boldsymbol{\beta}_0$	$Bias$	0.0100	0.0205	0.0124	0.1574	0.1002	**0.0349**
	VAR_F	0.0296	0.0457	**0.0135**	0.0379	0.0619	**0.0194**
	MSE	0.0297	0.0458	**0.0136**	0.0473	0.0654	**0.0200**
$\boldsymbol{\beta}_1$	$Bias$	0.1550	-	0.1553	0.3024	-	**0.1624**
	VAR_F	0.1392	-	**0.1358**	0.0360	-	**0.0337**
	MSE	0.1479	-	**0.1447**	0.0681	-	**0.0433**
$\boldsymbol{\beta}_2$	$Bias$	-	0.1323	0.1334	-	0.1589	**0.0824**
	VAR_F	-	0.2544	**0.2527**	-	0.0306	**0.0294**
	MSE	-	0.2604	**0.2587**	-	0.0391	**0.0330**

Table 3.5: Comparisons of M_i and M_c under Design 2 and Set 2

Parameters	Design	Orthogonal			Non-orthogonal		
ZS	Statistics	M_1	M_2	M_c	M_1	M_2	M_c
$\boldsymbol{\beta}_0$	$Bias$	0.0564	0.0423	**0.0301**	0.0716	0.0860	**0.0410**
	VAR_F	0.0599	0.0344	**0.0211**	0.0882	0.0396	**0.0268**
	MSE	0.0610	0.0350	**0.0214**	0.0912	0.0422	**0.0275**
$\boldsymbol{\beta}_1$	$Bias$	0.0112	-	**0.0104**	0.0364	-	**0.0139**
	VAR_F	0.0850	-	**0.0850**	0.0720	-	**0.0610**
	MSE	0.0851	-	**0.0851**	0.0726	-	**0.0611**
$\boldsymbol{\beta}_2$	$Bias$	-	0.0720	**0.0570**	-	0.0951	**0.0672**
	VAR_F	-	0.0728	0.0790	-	0.0563	0.0606
	MSE	-	0.0750	**0.0804**	-	0.0603	0.0624
TESS							
$\boldsymbol{\beta}_0$	$Bias$	0.1100	0.0849	**0.0792**	0.1276	0.1265	**0.0926**
	VAR_F	0.0442	0.0282	**0.0173**	0.0569	0.0292	**0.0199**
	MSE	0.0487	0.0308	**0.0195**	0.0647	0.0347	**0.0237**
$\boldsymbol{\beta}_1$	$Bias$	0.0120	-	**0.0108**	0.0560	-	**0.0313**
	VAR_F	0.0850	-	**0.0850**	0.0663	-	**0.0597**
	MSE	0.0851	-	**0.0851**	0.0675	-	**0.0603**
$\boldsymbol{\beta}_2$	$Bias$	-	0.1197	0.1311	-	0.1490	**0.1474**
	VAR_F	-	0.0546	**0.0501**	-	0.0384	**0.0368**
	MSE	-	0.0602	**0.0568**	-	0.0476	**0.0450**

Table 3.6: Posterior variance analysis for Design 1 with $\sigma^2 = 1.0$

Comparison	Prior	Parameter	Dominant $\boldsymbol{\beta}_0$		Dominant $\boldsymbol{\beta}_1$	
			Orthogonal	Non-orthogonal	Orthogonal	Non-orthogonal
M_1 vs M_c	ZS	$\boldsymbol{\beta}_0$	93.6%	96.6%	95.2%	98.2%
		$\boldsymbol{\beta}_1$	62.6%	87.6%	75.6%	94.0%
	TESS	$\boldsymbol{\beta}_0$	94.8%	97.4%	95.4%	98.6%
		$\boldsymbol{\beta}_1$	66.4%	86.2%	77.4%	93.6%
M_2 vs M_c	ZS	$\boldsymbol{\beta}_0$	97.6%	99.4%	93.2%	97.4%
		$\boldsymbol{\beta}_2$	68.4%	97.8%	24.2%	77.6%
	TESS	$\boldsymbol{\beta}_0$	98.4%	99.4%	95.2%	98.0%
		$\boldsymbol{\beta}_2$	74.2%	97.0%	35.2%	76.8%

Table 3.7: Posterior variance analysis for Design 2 with $\sigma^2 = 1.0$

Comparison	Prior	Parameter	Dominant $\boldsymbol{\beta}_0$		Dominant $\boldsymbol{\beta}_1$	
			Orthogonal	Non-orthogonal	Orthogonal	Non-orthogonal
M_1 vs M_c	ZS	$\boldsymbol{\beta}_0$	95.2%	99.4%	95.6%	99.6%
		$\boldsymbol{\beta}_1$	64.8%	77.6%	71.0%	77.4%
	TESS	$\boldsymbol{\beta}_0$	96.0%	99.4%	96.4%	99.8%
		$\boldsymbol{\beta}_1$	70.2%	81.4%	72.2%	77.0%
M_2 vs M_c	ZS	$\boldsymbol{\beta}_0$	97.6%	95.0%	92.4%	81.6%
		$\boldsymbol{\beta}_2$	68.0%	72.2%	18.4%	28.4%
	TESS	$\boldsymbol{\beta}_0$	97.8%	96.8%	94.4%	87.2%
		$\boldsymbol{\beta}_2$	69.8%	75.0%	28.4%	37.8%

Table 3.8: Posterior variance analysis for Design 1 with $\sigma^2 = 0.01$

Comparison	Prior	Parameter	Dominant $\boldsymbol{\beta}_0$		Dominant $\boldsymbol{\beta}_1$	
			Orthogonal	Non-orthogonal	Orthogonal	Non-orthogonal
M_1 vs M_c	ZS	$\boldsymbol{\beta}_0$	98.0%	99.2%	99.0%	99.2%
		$\boldsymbol{\beta}_1$	82.4%	98.4%	95.0%	98.0%
	TESS	$\boldsymbol{\beta}_0$	98.0%	99.8%	99.0%	99.2%
		$\boldsymbol{\beta}_1$	84.6%	98.8%	97.0%	98.2%
M_2 vs M_c	ZS	$\boldsymbol{\beta}_0$	99.8%	100.0%	99.6%	100.0%
		$\boldsymbol{\beta}_2$	94.4%	100.0%	89.4%	99.6%
	TESS	$\boldsymbol{\beta}_0$	99.8%	100.0%	99.8%	100.0%
		$\boldsymbol{\beta}_2$	94.6%	99.8%	91.8%	99.8%

Table 3.9: Posterior variance analysis for Design 2 with $\sigma^2 = 0.01$

Comparison	Prior	Parameter	Dominant $\boldsymbol{\beta}_0$		Dominant $\boldsymbol{\beta}_1$	
			Orthogonal	Non-orthogonal	Orthogonal	Non-orthogonal
M_1 vs M_c	ZS	$\boldsymbol{\beta}_0$	98.6%	99.8%	99.0%	99.8%
		$\boldsymbol{\beta}_1$	89.4%	94.0%	95.8%	91.4%
	TESS	$\boldsymbol{\beta}_0$	98.0%	99.8%	99.0%	100.0%
		$\boldsymbol{\beta}_1$	89.8%	94.0%	96.6%	93.0%
M_2 vs M_c	ZS	$\boldsymbol{\beta}_0$	99.4%	98.6%	99.2%	98.4%
		$\boldsymbol{\beta}_2$	89.8%	92.6%	79.8%	88.0%
	TESS	$\boldsymbol{\beta}_0$	99.4%	99.0%	99.4%	98.6%
		$\boldsymbol{\beta}_2$	91.8%	93.8%	89.2%	92.2%

Specifically, the sensitivity analysis is conducted under two scenarios. In Scenario 1, we include four cases where the dimension of $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ is fixed at $p_0 = 4, p_1 = p_2 = 2$ and the sample size ratio n_1/n_2 varies from 0.5 to 2.0 with $n_1 = 30, n_2 = 15;\ n_1 = 30, n_2 = 30;\ n_1 = 30, n_2 = 45;\ n_1 = 30, n_2 = 60$. In Scenario 2, we also have four cases where the sample size is fixed at $n_1 = n_2 = 15$ and the ratio of dimension p_0/p_i varies from 1.0 to 4.0 with $p_0 = 4, p_1 = p_2 = 1$; $p_0 = 4, p_1 = p_2 = 2$; $p_0 = 4, p_1 = p_2 = 3$ and $p_0 = p_1 = p_2 = 4$. We examine each scenario under two sets of coefficients with $\sigma^2 = 0.5$ and the largest model for each set is:

- Set 3: $\boldsymbol{\beta}_0 = (8, 5, 3, 2)'$, $\boldsymbol{\beta}_1 = (0.7, 0.5, 0.6, 0.7)'$, $\boldsymbol{\beta}_2 = (0.6, 0.6, 0.7, 0.6)'$, where the common $\boldsymbol{\beta}_0$ is dominant over the specific $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in size;

- Set 4: $\boldsymbol{\beta}_0 = (0.7, 0.5, 0.6, 0.6)'$, $\boldsymbol{\beta}_1 = (1.5, 2.5, 1.6, 1.5)'$, $\boldsymbol{\beta}_2 = (1.9, 2.1, 1.7, 1.4)'$, where the specific $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are dominant over the common $\boldsymbol{\beta}_0$ in size.

When $p_1,\ p_2 \leq 4$, the coefficients correspond to the first k elements of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ in Sets 1 and 2. The design matrices are generated from the normal distribution $N(0, 1)$. For each combination of four cases from two scenarios and two sets of coefficients, we collect frequentist properties of the Bayesian estimator including its sampling standard deviation, relative bias and relative MSE with the same definition in (2.37) and (2.38). Each Bayesian estimator is computed through 20,000 samples with 10,000 burn-ins. The frequentist properties are calculated with 500 data sets.

Tables 3.10 - 3.13 summarize sensitivity results for two scenarios and two sets of coefficients. Main findings are quiet similar to those in Chapter 2. First, M_c and M_{gs} yield similar frequentist standard deviation while M_{gs} shows a smaller bias and therefore a smaller MSE. Second, despite dominant coefficients, M_c offers improvements

Table 3.10: Sensitivity analysis for Scenario 1 with Set 3

Parameter	Design	$n_1 = 30$, $n_2 = 15$				$n_1 = 30$, $n_2 = 30$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0060	0.0039	0.0053	0.0012	0.0058	0.0052	0.0028	0.0001
	SD_F	0.0595	0.0751	0.0431	0.0482	0.0565	0.0560	0.0380	0.0374
	$RMSE$	0.0047	0.0047	0.0038	0.0028	0.0043	0.0043	0.0026	0.0021
$\boldsymbol{\beta}_1$	$RBias$	0.0568	-	0.0607	0.0115	0.0932	-	0.0895	0.0084
	SD_F	0.0719	-	0.0680	0.0671	0.0742	-	0.0723	0.0495
	$RMSE$	0.0723	-	0.0718	0.0568	0.0998	-	0.0913	0.0417
$\boldsymbol{\beta}_2$	$RBias$	-	0.0815	0.0547	0.0156	-	0.2151	0.1995	0.0090
	SD_F	-	0.0819	0.0787	0.0608	-	0.0719	0.0695	0.0497
	$RMSE$	-	0.0916	0.0789	0.0524	-	0.1680	0.1575	0.0419
Parameter	Design	$n_1 = 30$, $n_2 = 45$				$n_1 = 30$, $n_2 = 60$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0080	0.0048	0.0046	0.0002	0.0033	0.0059	0.0045	0.0001
	SD_F	0.0550	0.0422	0.0324	0.0317	0.0575	0.0334	0.0282	0.0276
	$RMSE$	0.0054	0.0037	0.0035	0.0018	0.0037	0.0043	0.0035	0.0015
$\boldsymbol{\beta}_1$	$RBias$	0.0745	-	0.0503	0.0050	0.1386	-	0.0838	0.0073
	SD_F	0.0831	-	0.0776	0.0456	0.0757	-	0.0695	0.0354
	$RMSE$	0.0918	-	0.0776	0.0382	0.1233	-	0.0839	0.0300
$\boldsymbol{\beta}_2$	$RBias$	-	0.0430	0.0337	0.0052	-	0.0829	0.0773	0.0028
	SD_F	-	0.0536	0.0527	0.0430	-	0.0563	0.0545	0.0427
	$RMSE$	-	0.0568	0.0548	0.0361	-	0.0866	0.0819	0.0356

in terms of VAR_F and MSE compared with M_i, especially for the imbalanced design. Third, M_c and M_{gs} yield similar results for small sample size, which is possibly related to more loss of information for M_c with a larger sample size. For example, the results are quite similar for $n_1 = 30$ and $n_2 = 15$ instead of $n_1 = 30$ and $n_2 = 30$. The differences between M_c and M_{gs} in terms of VAR_F and MSE increase as the sample size increases. Such difference is less obvious for dominant $\boldsymbol{\beta}_i$ rather than dominant $\boldsymbol{\beta}_0$. Fourth, the difference between M_c and M_{gs} in terms of considered quantities are smaller for dominant coefficients rather than specific coefficients.

Table 3.11: Sensitivity analysis for Scenario 2 with Set 3

Parameter	Design Statistics	$p_0 = 4, p_1 = p_2 = 1$				$p_0 = 4, p_1 = p_2 = 2$			
		M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0043	0.0063	0.0040	0.0004	0.0112	0.0036	0.0047	0.0004
	SD_F	0.0816	0.0717	0.0502	0.0487	0.0864	0.0955	0.0599	0.0580
	$RMSE$	0.0052	0.0053	0.0039	0.0027	0.0084	0.0057	0.0042	0.0032
$\boldsymbol{\beta}_1$	$RBias$	0.1820	-	0.1236	0.0330	0.1106	-	0.1047	0.0120
	SD_F	0.1470	-	0.1353	0.0943	0.1376	-	0.1145	0.0733
	$RMSE$	0.2777	-	0.2295	0.1385	0.1502	-	0.1269	0.0618
$\boldsymbol{\beta}_2$	$Bias$	-	0.4377	0.3125	0.0400	-	0.0791	0.0524	0.0137
	SD_F	-	0.1432	0.1360	0.1049	-	0.1126	0.1010	0.0707
	$RMSE$	-	0.4986	0.3859	0.1787	-	0.1096	0.0963	0.0598
Parameter	Design Statistics	$p_0 = 4, p_1 = p_2 = 3$				$p_0 = 4, p_1 = p_2 = 4$			
		M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0079	0.0072	0.0033	0.0003	0.0241	0.0304	0.0148	0.0010
	SD_F	0.0778	0.0807	0.0529	0.0532	0.1003	0.0847	0.0647	0.0661
	$RMSE$	0.0061	0.0059	0.0040	0.0030	0.0148	0.0169	0.0094	0.0037
$\boldsymbol{\beta}_1$	$Bias$	0.0958	-	0.0645	0.0170	0.1158	-	0.0805	0.0061
	SD_F	0.0821	-	0.0805	0.0590	0.0769	-	0.0725	0.0541
	$RMSE$	0.0806	-	0.0636	0.0342	0.0674	-	0.0598	0.0220
$\boldsymbol{\beta}_2$	$Bias$	-	0.1480	0.1246	0.0149	-	0.0614	0.0714	0.0052
	SD_F	-	0.0932	0.0894	0.0708	-	0.0745	0.0725	0.0583
	$RMSE$	-	0.1069	0.0950	0.0386	-	0.0575	0.0517	0.0235

3.5 A Real Data Example

In this section, we analyze a personal medical cost data to illustrate the potential benefits of data combining in practice. It was first analyzed by Lantz (2013) to predict medical expenses, which is an essential task for insurance company to make a profit. Consequently, the insurers spend tremendous time in building models to predict medical expenses so that the medical care offered to beneficiaries can at least be covered by the yearly premiums. Medical expenses could depend on many conditions, which could be rare but costly conditions or more prevalent for certain

Table 3.12: Sensitivity analysis for Scenario 1 with Set 4

Parameter	Design	$n_1 = 30$, $n_2 = 15$				$n_1 = 30$, $n_2 = 30$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0159	0.0207	0.0156	0.0004	0.0164	0.0196	0.0083	0.0005
	SD_F	0.0566	0.0640	0.0509	0.0525	0.0531	0.0508	0.0438	0.0452
	$RMSE$	0.0322	0.0359	0.0320	0.0142	0.0315	0.0344	0.0219	0.0123
$\boldsymbol{\beta}_1$	$RBias$	0.0156	-	0.0136	0.0001	0.0201	-	0.0137	0.0003
	SD_F	0.1316	-	0.1304	0.1201	0.1302	-	0.1282	0.1092
	$RMSE$	0.0434	-	0.0414	0.0150	0.0461	-	0.0410	0.0137
$\boldsymbol{\beta}_2$	$RBias$	-	0.0247	0.0231	0.0002	-	0.0127	0.0063	0.0001
	SD_F	-	0.1523	0.1485	0.1173	-	0.1284	0.1277	0.1086
	$RMSE$	-	0.0489	0.0477	0.0147	-	0.0359	0.0281	0.0136
Parameter	Design	$n_1 = 30$, $n_2 = 45$				$n_1 = 30$, $n_2 = 60$			
	Statistics	M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0113	0.0105	0.0070	0.0002	0.0080	0.0075	0.0059	0.0003
	SD_F	0.0530	0.0457	0.0410	0.0417	0.0541	0.0436	0.0406	0.0413
	$RMSE$	0.0257	0.0243	0.0213	0.0113	0.0226	0.0226	0.0187	0.0111
$\boldsymbol{\beta}_1$	$RBias$	0.0256	-	0.0241	0.0001	0.0040	-	0.0031	0.0000
	SD_F	0.1309	-	0.1284	0.1023	0.1245	-	0.1210	0.0940
	$RMSE$	0.0498	-	0.0472	0.0128	0.0207	-	0.0193	0.0118
$\boldsymbol{\beta}_2$	$RBias$	-	0.0104	0.0107	0.0001	-	0.0115	0.0106	0.0001
	SD_F	-	0.1182	0.1173	0.1033	-	0.1124	0.1118	0.1000
	$RMSE$	-	0.0349	0.0341	0.0129	-	0.0323	0.0310	0.0125

group in a population. This data focused on the latter. The response variable is the total medical expenses charged to the insurance plan for the calendar year and covariates if interests include age of the primary beneficiary (age), policy holder's gender (sex), body mass index (BMI), smoker (whether the insured regularly smokes tobacco), region (beneficiary's place of residence in the U.S.).

To reflect the data combining situation, we first divide the data set into two data sets according to the variable "region". Specifically, instead of four levels of region (i.e., northeast, northwest, southeast, southwest), we consider two levels (east, west),

Table 3.13: Sensitivity analysis for Scenario 2 with Set 4

Parameter	Design Statistics	$p_0 = 4, p_1 = p_2 = 1$				$p_0 = 4, p_1 = p_2 = 2$			
		M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.0872	0.1448	0.0557	0.0076	0.1642	0.2072	0.1298	0.0094
	SD_F	0.0679	0.0580	0.0447	0.0463	0.0700	0.0706	0.0481	0.0517
	$RMSE$	0.0645	0.0846	0.0370	0.0197	0.0906	0.1154	0.0761	0.0220
$\boldsymbol{\beta}_1$	$RBias$	0.2731	-	0.1188	0.0055	0.0535	-	0.0597	0.0030
	SD_F	0.1208	-	0.1082	0.0872	0.1339	-	0.1207	0.0862
	$RMSE$	0.2847	-	0.1390	0.0585	0.0623	-	0.0524	0.0217
$\boldsymbol{\beta}_2$	$RBias$	-	0.2168	0.0583	0.0036	-	0.0914	0.0544	0.0022
	SD_F	-	0.1407	0.1237	0.1030	-	0.1052	0.1031	0.0811
	$RMSE$	-	0.2292	0.0874	0.0542	-	0.0708	0.0463	0.0203
Parameter	Design Statistics	$p_0 = 4, p_1 = p_2 = 3$				$p_0 = 4, p_1 = p_2 = 4$			
		M_1	M_2	M_c	M_{gs}	M_1	M_2	M_c	M_{gs}
$\boldsymbol{\beta}_0$	$RBias$	0.1376	0.0945	0.0996	0.0079	0.1068	0.1165	0.0916	0.0094
	SD_F	0.0696	0.0618	0.0512	0.0550	0.0952	0.0637	0.0493	0.0572
	$RMSE$	0.0845	0.0598	0.0558	0.0233	0.0882	0.0696	0.0550	0.0244
$\boldsymbol{\beta}_1$	$RBias$	0.0244	-	0.0106	0.0005	0.1098	-	0.0956	0.0017
	SD_F	0.0685	-	0.0673	0.0642	0.0859	-	0.0810	0.0707
	$RMSE$	0.0711	-	0.0688	0.0115	0.0598	-	0.0547	0.0100
$\boldsymbol{\beta}_2$	$RBias$	-	0.0873	0.0857	0.0017	-	0.0443	0.0462	0.0026
	SD_F	-	0.0892	0.0883	0.0647	-	0.0698	0.0663	0.0586
	$RMSE$	-	0.0659	0.0628	0.0125	-	0.0298	0.0297	0.0084

and assign observation in east to data source 1 and west to data source 2. We also simulate a couple of cases with different combinations of common and specific covariates to offer more perspectives of the relative performances of M_i and M_c. M_1, M_2 and M_c has 688, 650 and 1,338 observations, respectively. For Case 1, we consider sex and smoker are shared by M_1 and M_2. BMI is only collected by M_1 and age is collected only by M_2. For Case 2, BMI and smoker are considered as shared while sex is available for M_1 and age is available for M_2.

Tables 3.14 and 3.15 summarize findings including, posterior mean (M_p), poste-

Table 3.14: M_p, VAR_p, and 95% CI along with its width in Case 1.

Parameter	Statistics	M_1	M_c	M_2
β_{01}	M_p (VAR_p)	1.6657 (0.0032)	1.7310 (**0.0017**)	1.8148 (0.0040)
	95 % CI	(1.5573, 1.7754)	(1.6484, 1.8140)	(1.6924, 1.9388)
	width	0.2181	**0.1655**	0.2463
β_{02}	M_p (VAR_p)	-0.4170 (0.0015)	-0.4301 (**0.0008**)	-0.4399 (0.0017)
	95 % CI	(-0.4921, -0.3403)	(-0.4841, -0.3758)	(-0.5201, -0.3599)
	width	0.1518	**0.1083**	0.1602
β_{11}	M_p (VAR_p)	0.2486 (0.0014)	0.2483 (**0.0014**)	-
	95 % CI	(0.1742, 0.3225)	(0.1758, 0.3209)	-
	width	0.1483	**0.1451**	
β_{21}	M_p (VAR_p)	-	0.1311 (**0.0002**)	0.1282 (0.0003)
	95 % CI	-	(0.1000, 0.1620)	(0.0963, 0.1605)
	width		**0.0619**	0.0642

rior variance (VAR_p), 95% credible intervals (CI) and its corresponding width. From these tables, compared with M_i, we may find that posterior variances and width of 95 % CI for each parameter is smaller in M_c, which implies that M_c offers more precise Bayesian estimates. In addition, no matter Case 1 or Case 2, common parameters benefits more from data combining since more reductions have been observed regarding posterior variances and widths in contrast with specific parameters. At last, all models from Table 3.14 and Table 3.15 indicates that smoker, female, older people are more likely to be charged more while lower BMI is associated with lower charges, which align with our common senses.

3.6 Discussion

In this chapter, we take a focused investigation on independent g-priors from an estimation perspective under two cases. For Case 1, where (σ^2, g) is known, we mainly evaluate the posterior variances of the Bayesian estimators in M_i and M_c, and

Table 3.15: M_p, VAR_p, and 95% CI along with its width in Case 2.

Parameter	Statistics	M_1	M_c	M_2
	M_p (VAR_p)	1.6650 (0.0032)	1.6439 (**0.0017**)	1.6184 (0.0040)
β_{01}	95 % CI	(1.5559, 1.7754)	(1.5618, 1.7273)	(1.4946, 1.7440)
	width	0.2195	**0.1654**	0.2494
	M_p (VAR_p)	0.2491 (0.0014)	0.2852 (**0.0009**)	0.3406 (0.0024)
β_{02}	95 % CI	(0.1759, 0.3240)	(0.2262, 0.3437)	(0.2434, 0.4369)
	width	0.1481	**0.1175**	0.1935
	M_p (VAR_p)	-0.4163 (0.0015)	-0.4166 (**0.0014**)	
β_{11}	95 % CI	(-0.4900, -0.3407)	(-0.4899, -0.3444)	-
	width	0.1493	**0.1455**	
	M_p (VAR_p)		0.0558 (**0.0002**)	0.0545 (0.0003)
β_{21}	95 % CI	-	(0.0246, 0.0864)	(0.0230, 0.0867)
	width		**0.0617**	0.0637

some frequentist properties with special cases. For the second case, where (σ^2, g) is unknown, we research more on the frequentist properties of the Bayesian estimator through the lens of asymptotic analysis. This asymptotic is first defined by Som et al. (2016) and often referred to as fixed p and fixed n asymptotic (or conditional informational asymptotic). Utilizing this defined asymptotic sequence, we derive the asymptotic mean and covariance of the Bayesian estimator in M_i and M_c under two situations. One considers where the model is driven by common coefficients and the other considers where the model is driven by specific coefficients. Our theoretical results not only echo the essential least squares (ELS) estimation framework in Som et al. (2016) under the hyper-g prior but also extend the framework to study of conditional asymptotic frequentist properties. Inspired by Berger et al. (2014) and Min and Sun (2016), we further adopt TESS to offer an adjustment to the scale in the ZS-prior and investigate its potential in improving the estimation. Our findings reveal that M_c contributes to a smaller risk in terms of MSE even if there is no information borrowing in most cases (e.g. block orthogonal design matrices). Incorporating TESS in the ZS prior is very likely to improve the estimates. More importantly, we bring ZS prior together with TESS into the estimation scope and quantify their potential benefits of data combining. Our extensive simulation studies and real data example also consolidate our theories.

In fact, our framework could be extended to other shrinkage estimations such as shrinkage prior or robust prior and plenty of future directions worth exploring based on our work. First and foremost, M_i and M_c is under the assumption of independent and identically distributed (iid) error terms. In practice, it might be more realistic to consider non-iid settings where we allow different errors for different data sources and therefore such generalizations of our framework are needed. Second, although our straightforward combining strategy makes theoretical pursuit possible, alternative data combining methods should be explored and compared. One example is introducing external data sources to impute missing covariates (Jackson et al., 2009). Another example is modeling the between-study covariance matrix to enable more information sharing depending on a specific situation (Siegel et al., 2020), which is a common method in meta-analysis. In fact, this option will be explored in our next chapter. Third, the derivation of TESS is offered by Berger et al. (2014) for the purpose of model selection and is applied directly for estimation in our case. Specifically, we remove the scale by the observation with maximum information. Alternative scaling options or a general definition to obtain a suitable form for the purpose of estimation is highly recommended. Fourth, under our data combining framework, extensions to the generalized linear mixed model (Li and Clyde, 2018) under g-prior along with TESS should be studied to accommodate practical considerations.

Chapter 4

Female Breast Cancer Prevalence in Missouri

Chapters 2 and 3 investigate the data combining with linear models in (1.2) and primarily access the performance of the Bayesian estimator in M_i and M_c under the classical and independent g-prior or ZS prior. However, in practice, more complex models are frequently needed rather than the linear regression, and the two data sources may be of different nature. As a result, it is critical to evaluate our data combining strategy in a broader context.

Among many fields, analyses for cancer statistics play an important role as it is the second leading cause of death according to Centers for Disease Control and Prevention (CDC). Breast cancer is the most common cancer in women in the United States except for skin cancers. This makes it essential to evaluate its overall burden in the population and cancer prevalence is generally used to achieve such purpose. In this chapter, we evaluate county-level female breast cancer prevalence in Missouri via different variants of our data combining strategy. In fact, for many diseases, including FBC, county-level data sources for calculating prevalence estimates are limited due to small sample sizes. To the best of our knowledge, Missouri Cancer Registry (MCR) and County-level Study (CLS) are two available data sources in Missouri to conduct such analysis. On the one hand, different data sources could have common variables such as county attributes. On the other hand, different data sources have their own limitations. For example, regarding prevalence estimates, the survey data from the Behavioral Risk Factor Surveillance System (BRFSS)-based CLS suffers from non-response and possible recall bias and lacks some cancer information of interest (e.g., stage at diagnosis). Meanwhile, administrative data from MCR suffers from limited time period of data collection (only the prevalence of cancer survivors diagnosed since 1996 can be directly measured via MCR data) and lacks risk factor information (such as whether the person is obese or told of high cholesterol level by health professional). Therefore, combining available county-level data sources in Missouri is a promising approach to help us model the relationship between the FBC prevalence and covariates of interests, and provide more precise estimates of the corresponding effects. Additionally, understanding the relationship between the prevalence estimates and covariates, such as risk factors, could provide useful information for the public to prevent disease in advance and for health care planners to allocate resources.

The remainder of this chapter is organized as follows. We first provide some background for the two types of cancer prevalence, two data sources for prevalence with distinct characteristics, and the data source for county attributes. Second, we present several candidate data combining strategies according to observed data sources. We wrap up with a discussion of issues and potential future directions.

4.1 Background

According to National Institutes of Health (NIH), prevalence is defined as a proportion of people alive on a certain date in a population who previously had a diagnosis of the disease. In the context of cancer statistics, there are two types of prevalence: (1) Limited-Duration Prevalence (LDP), which represents the proportion of people alive on a certain day who had a diagnosis of the disease within a past period; (2) Complete Prevalence (CP), which represents the proportion of people alive on a certain day who previously had a diagnosis of the disease. Prevalence, LDP or CP, is used to evaluate existing cases or the overall burden of a certain disease. It differs from incidence, which indicates newly diagnosed cases in a defined population. Although prevalence cannot provide as much information as incidence from the perspective of cancer etiology, it could provide information regarding health care resources and be helpful in health care planning. One may also notice that prevalence is studied in limited settings since reliable cancer prevalence estimates might come from long-term cancer registries instead of population survey.

4.2 Data Source

Sections 4.2.1 and 4.2.2 present data sources for FBC prevalence and specific covariates according to different data sources. Section 4.2.3 describes the data source for common covariates.

4.2.1 Missouri Cancer Registry and Research Center (MCR-ARC)

Every cancer incidence is required to be reported to the Missouri Cancer Registry in accordance with Missouri Statutes (192.650-192.657 RSMo), then the information is edited and consolidated by MCR-ARC staff. MCR-ARC has high quality (at least 95% of expected incidence cases) data through December 31, 2018. Several key characteristics are as follows.

1. MCR-ARC data is population-based, which collects all cancer incidence in Missouri since 1996.

2. Only the prevalence of cancer survivors diagnosed since 1996 can be directly measured via MCR-ARC data. Hence, the prevalence estimate is based on a limited time period and called LDP. In our case, we use 20 years limited-duration prevalence so that LDP and CP could be as close as possible.

3. MCR-ARC data has cancer-specific information such as stage information, which is not available in CLS.

4. MCR-ARC data is edited by professional staff, which implies that cancer related concepts might be different than those without professional training.

4.2.2 2016 Missouri County-level Study (CLS)

Missouri county-level study (CLS) is a self-reported survey based on landline and cell telephones. Its target is to produce accurate county-level estimates. CLS has been conducted in years 2007, 2011 and 2016. In 2016, it aimed at completing approximately 52,000 landline and cell telephone calls for adults (aged 18 or older) throughout the year. Prevalence estimates were generated for the 114 Missouri counties and the city of St.Louis. We need to point out several features for this data source.

1. Since CLS focused on providing accurate county-level estimation, for each county, the sample size used to perform statistical analysis is relatively larger compared with other surveys, such as BRFSS. The specific goal is described as below.
   - 400 each in the 105 smallest counties;
   - 800 (400 urban/400 rural) in Buchanan, Boone, Cole, Greene, and Jasper Counties; 800 in St. Charles County (400 eastern and 400 western); 800 in Jefferson County (400 northern and 400 southern);
   - 1200 in Jackson County (800 in Kansas City, 400 in Independence and 400 in Eastern Jackson County); 1200 in the City of St. Louis (400 each in 3 strata);
   - 2000 in St. Louis County (400 each in 5 strata).
2. The prevalence percent estimates from CLS refers to complete prevalence for the reason that a participant in the survey is asked questions such as "Have you ever been diagnosed as cancer", which differs from when an individual is asked "Have you ever been diagnosed as cancer in the past xx years".
3. The CLS prevalence percent estimates were weighted with the raking method to be representative for the Missouri adult, non-institutionalized population of the area covered.
4. The CLS collects information about cancer-related risk factors, such as obesity and smoke, which are not available for other data sources. In our case,

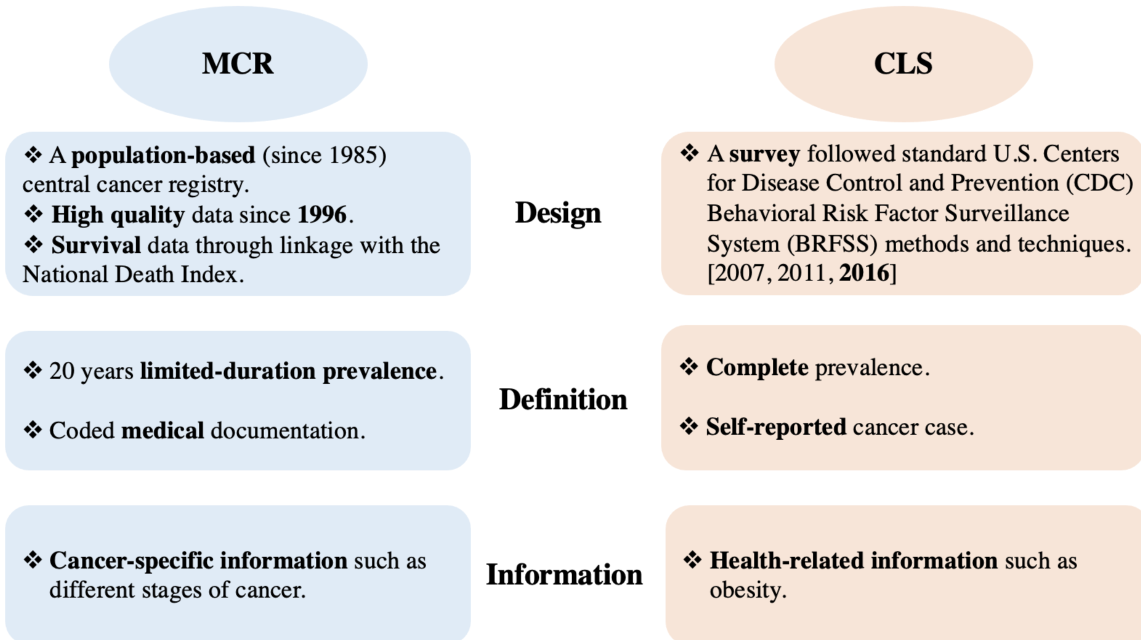

Figure 4.1: Summary of characteristics for MCR and CLS

we use cholesterol information, which is well-known for its relationship with cardiovascular disease. Its specific association with FBC is still under investigation from a clinical perspective (Garcia-Estevez and Moreno-Bueno, 2019; Wei et al., 2021). From a surveillance viewpoint, we intend to study the relationship between FBC prevalence and cholesterol level.

5. As a self-reported survey, it suffered from non-response and possible recall bias. Especially, when it comes to the cancer study, the CLS additionally lacks some cancer information of interest, such as stage at diagnosis.

For an easier reference, Figure 4.1 summarizes the comparison of characteristics for MCR and CLS. Figure 4.2 demonstrates the concepts of LDP and CP using 2016 FBC prevalence data from MCR and CLS.

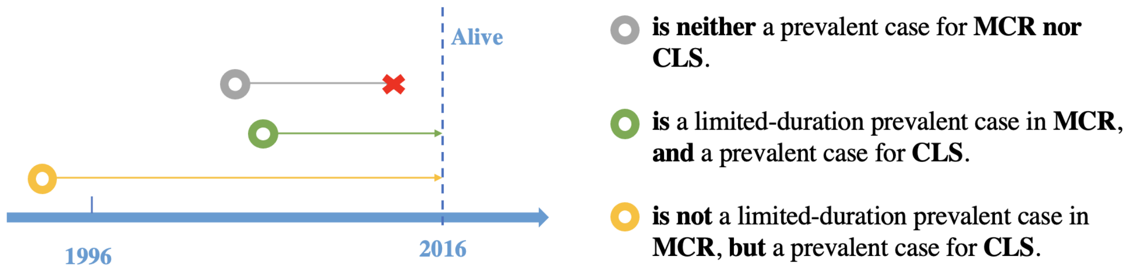

Figure 4.2: An illustration for LDP and CP

4.2.3 Others

For common covariates, we consider county attributes since we expect this demographic information to be the same for a geographical region despite data sources. County attributes are obtained from American Community Survey (ACS) 5-year files, which provide more reliable data for small population compared yearly file. Variables of interests such as percentages of poverty level (below, at or above) are aggregated from the ACS 2014-2018 data file. As 2016 is the middle of this time span, we expect it to be more accurate than other time spans or ACS yearly files.

4.2.4 Data Overview

To better understand our various modeling strategies, this section visualizes the distribution of FBC prevalence (response variable) and covariates of interests. Tables 4.1 - 4.3 show summary statistics for covariates and prevalence proportions (PP), prevalence counts, and population size, respectively. Figures 4.3 - 4.4 display histograms for prevalence proportions and covariates of interests, respectively. Two models are discussed in this chapter. The first is the linear mixed model where we assume the logit transformed FBC prevalence proportions are observed. The second is the gen-

Table 4.1: Summary statistics for covariates and prevalence proportions

Variable		**Statistics**					
		Min	**1st quantile**	**Median**	**Mean**	**3rd quantile**	**Max**
Age	65+	0.1435	0.2371	0.2590	0.2619	0.2904	0.3920
Poverty	At or above	0.6917	0.8041	0.8319	0.8289	0.8619	0.9392
MCR PP	Early	0.0080	0.0132	0.0146	0.0149	0.0163	0.0256
	Late	0.0032	0.0058	0.0067	0.0069	0.0077	0.0114
CLS PP	Had high cholesterol	0.0081	0.0347	0.0507	0.0536	0.0680	0.1580
	No high cholesterol	0.0031	0.0160	0.0268	0.0333	0.0442	0.1087

Table 4.2: Summary statistics for prevalence counts

Prevalent counts		**Statistics**					
		Min	**1st quantile**	**Median**	**Mean**	**3rd quantile**	**Max**
MCR	**Early**	12.0	64.5	114.0	323.4	201.0	8014.0
	Late	9.0	29.5	47.0	146.4	94.5	3537.0
CLS	**Had high cholesterol**	7.0	46.5	101.0	324.8	227.5	6096.0
	No high cholesterol	3.0	38.5	109.0	347.7	232.0	8978.0

Table 4.3: Summary statistics for population size

County-level population size		**Statistics**					
		Min	**1st quantile**	**Median**	**Mean**	**3rd quantile**	**Max**
MCR		823	3989	7237	21073	15209	416081
CLS	**Had high cholesterol**	230	1134	2279	5537	4028	100073
	No high cholesterol	490	1982	3318	10804	7245	243540

eralized linear mixed model, where the prevalence counts and population size are considered to be observed.

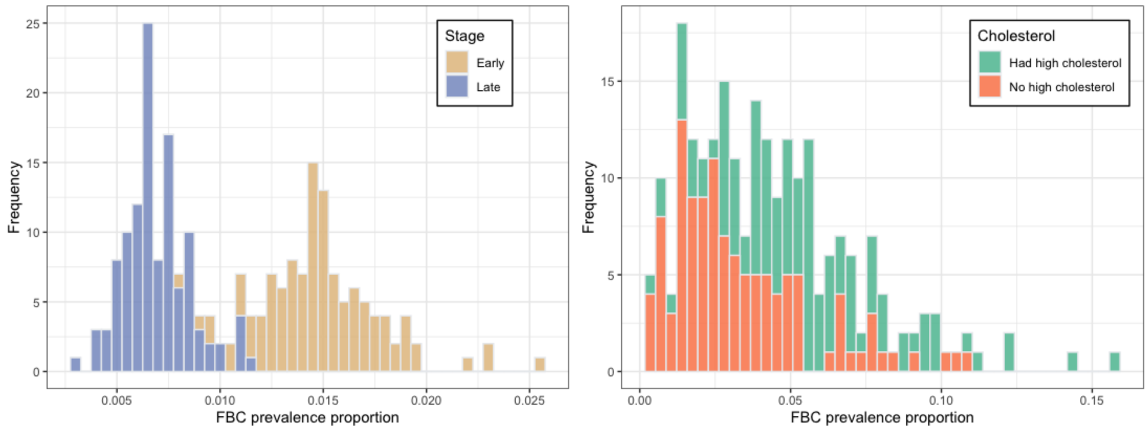

Figure 4.3: Histograms of prevalence proportions for MCR (left) and CLS (right).

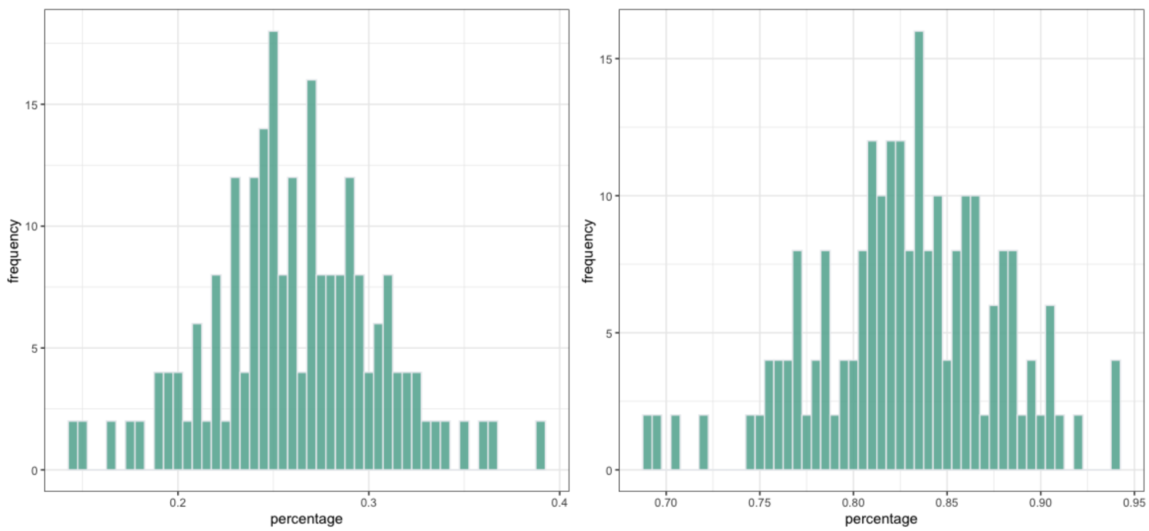

Figure 4.4: Histograms for percentages of women age over 65 (left) and live at or above poverty level (right).

4.3 Linear Mixed Model (LMM)

4.3.1 Individual Model for MCR or CLS

Let p^i_{jk} denote the FBC prevalence proportion the i-th data source, j-th county and k-th category, and then the observed response variable v^i_{jk} is the logit transformed

p^i_{jk}, i.e., $v^i_{jk} = logit(p^i_{jk}) = log(p^i_{jk}/(1-p^i_{jk}))$. As displayed in Table 4.1, there are no zero prevalence estimates for both data sources. We specify M_i as below:

$$v^i_{jk} = \mu^i + \beta_{01}x_{1j} + \beta_{02}x_{2j} + \gamma^i_k + z_j + \epsilon^i_{jk}, \tag{4.1}$$

where

- $i = 1, 2$, $j = 1, 2, \cdots, J$ and $k = 1, \cdots, K_i$;
  - $i = 1$ corresponds to the data from MCR;
  - $i = 2$ corresponds to the data from CLS;
- μ^i is the overall mean for data source i;
- x_{1j} is the percentage of age group 65+ for county j, with coefficient β_{01};
- x_{2j} is the percentage of at or above poverty level for county j, with coefficient β_{02};
- γ^i_k is the special effect for data source i;
  - γ^1_k is the stage effect with two categories (early[localized]/late[regional,distant]) and $K_1 = 2$;
  - γ^2_k is the cholesterol effect with two categories (no high cholesterol/had high cholesterol) and $K_2 = 2$;
  - $\sum_{k=1}^{K_i} \gamma^i_k = 0$;
- z_j is the random effect, which accounts for county spatial effect. In our case, $J = 115$.

- ϵ_{jk}^i is the random error and $\epsilon_{jk}^i \overset{i.i.d.}{\sim} N(0, \sigma^2)$.

To rewrite M_i in (4.1) as the matrix representation, we define a $JK_i \times J$ design matrix $\boldsymbol{X}_z = \boldsymbol{I}_J \otimes \boldsymbol{1}_{K_i}$, with a vector $\boldsymbol{z} = (z_1, z_2, \cdots, z_J)'$, where $\boldsymbol{I}_J$ is an identity matrix of size J, $\boldsymbol{1}_{K_i}$ is a vector with all ones of size K_i and $\otimes$ is the Kronecker product. Similarly, let the vector $\boldsymbol{\beta}_0 = (\beta_{01}, \beta_{02})'$ and $\boldsymbol{x}_j = (x_{1j}, x_{2j})'$, then $\beta_{01}x_{1j} + \beta_{02}x_{2j} = \boldsymbol{x}_j'\boldsymbol{\beta}_0$ and the design matrix for $\boldsymbol{\beta}_0$ is $\boldsymbol{X}_{\beta_0}^i = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_J)' \otimes \boldsymbol{1}_{K_i}$; let the vector $\boldsymbol{\gamma}^i = (\gamma_1^i, \gamma_2^i, \cdots, \gamma_{K_i}^i)$ and its design matrix $\boldsymbol{X}_\gamma^i = \boldsymbol{1}_J \otimes \boldsymbol{I}_{K_i}$; finally, $\boldsymbol{v}_j^i = (v_{j1}^i, \cdots, v_{jK_i}^i)'$ and $\boldsymbol{v}^i = ((\boldsymbol{v}_1^i)', \cdots, (\boldsymbol{v}_J^i)')'$. The model in (4.1) is equivalent to:

$$\boldsymbol{v}^i = \mu^i \boldsymbol{1}_{JK_i} + \boldsymbol{X}_{\beta_0}^i \boldsymbol{\beta}_0 + \boldsymbol{X}_\gamma^i \boldsymbol{\gamma}^i + \boldsymbol{X}_z \boldsymbol{z} + \boldsymbol{\epsilon}^i, \tag{4.2}$$

where $\boldsymbol{\epsilon}^i$ is a vector of dimension JK_i with distribution $N_{JK_i}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{JK_i})$. Here, $\boldsymbol{\beta}_0$ are interpreted as county attributes, $\boldsymbol{\gamma}^i$ is interpreted as data-source special effect and $\boldsymbol{z}$ is considered as county spatial effect.

4.3.2 Combined Model for MCR and CLS

To specify M_c for combined data of MCR and CLS, we first clarify the common coefficients and specific coefficients, and two candidate models are considered accordingly. The Candidate Model 1, denoted by M_{c_1}, assumes that no systematic difference exist between two data sources, namely, $\mu^1 = \mu^2$. This case considers the common coefficients for two data sources are the overall mean, county attributes, and spatial effects. Then, the specific coefficients correspond to stage effects and cholesterol information. Candidate Model 2, denoted by M_{c_2}, assumes that there is a systematic difference between two data sources. Thus, the common coefficients correspond to county attributes, while the specific coefficients correspond to different overall means from M_1, and special coefficients related to stage effects and cholesterol effects. As the stage and cholesterol effects are categorical variables and we intend to include an overall mean effect in the model, we rewrite the stage and cholesterol effects as its treatment means' form to quantify the data source effect later. As a result, although M_{c_2} does not appear the same as M_c, they have essentially the same structure.

Candidate Model 1

Following notations in Section 4.3.1, M_{c_1} for the combined data is:

$$\begin{pmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{JK_1} & \boldsymbol{X}_{\beta_0}^1 \\ \mathbf{1}_{JK_2} & \boldsymbol{X}_{\beta_0}^2 \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\beta}_0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_{\gamma}^1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{X}_{\gamma}^2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\gamma}^1 \\ \boldsymbol{\gamma}^2 \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_z \\ \boldsymbol{X}_z \end{pmatrix} \boldsymbol{z} + \begin{pmatrix} \boldsymbol{\epsilon}^1 \\ \boldsymbol{\epsilon}^2 \end{pmatrix}, \quad (4.3)$$

where $\mu, \boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}^k$ are the fixed effects, $\boldsymbol{z}$ is the random effect. For this particular application, $\boldsymbol{X}_{\gamma}^1 = \boldsymbol{X}_{\gamma}^2$ and $\boldsymbol{X}_{\beta_0}^1 = \boldsymbol{X}_{\beta_0}^2$. In general, our framework allows these design matrices to be different.

Formulation in (4.3) has some potential issues. First, for county attributes $\boldsymbol{\beta}_0$, we center its design matrix $\boldsymbol{X}_{\beta_0}$ with a centering matrix $\boldsymbol{C}_{JK} = \boldsymbol{I}_{JK} - \mathbf{1}\mathbf{1}'/JK$ and denote $\boldsymbol{X}_{\beta_0}^{i\star} = \boldsymbol{C}_{JK}\boldsymbol{X}_{\beta_0}^i$, where $K = K_1 + K_2$. This enables the same meaning for the overall mean in all models. In a model/variable setting, this is a frequently used setting so that a common prior can be specified for common coefficients. Second, since the special effect $\boldsymbol{\gamma}^k$ is a categorical variable, model in (4.3) suffers from identification problems, which has been discussed in vast literature in analysis of variance models (Rouder et al., 2012; Wang, 2017). For special effects $\boldsymbol{\gamma}^i$, firstly, we use the sum-to-zero constraints to relieve the identification issue.

$\boldsymbol{\gamma}^1 = (\gamma_1^1, \gamma_2^1)'$ reduces to γ^1 and $\boldsymbol{\gamma}^2 = (\gamma_1^2, \gamma_2^2)'$ reduces to γ^2. The corresponding design matrix reduces to $\boldsymbol{X}_\gamma^{i\star} = \mathbf{1}_I \otimes (1, -1)'$, and then a QR decomposition is applied $\boldsymbol{X}_\gamma^{i\star}\gamma^i = \boldsymbol{Q}^i\boldsymbol{R}^i\gamma^i = \boldsymbol{Q}^i\gamma^{i\star}$, where $\boldsymbol{Q}^{i\prime}\boldsymbol{Q}^i = \boldsymbol{I}_{K_i-1}$. We found that QR decomposition contributes to better mixing property or the reduction of correlation among posterior samples regarding different parameters based on practice, especially when the random error is not i.i.d. A typical method to perform a QR decomposition is the Gram-Schmidt process. Besides, considering that the R matrix, which is an upper triangular matrix, is invertible in QR decomposition, we can always transform our estimates back to its original scales for reasonable interpretations.

After reparameterization, M_{c_1} in (4.3) is equivalent to:

$$\begin{pmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{JK_1} & \boldsymbol{X}_{\beta_0}^{1\star} \\ \mathbf{1}_{JK_2} & \boldsymbol{X}_{\beta_0}^{2\star} \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\beta}_0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{Q}^1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{Q}^2 \end{pmatrix} \begin{pmatrix} \gamma^{1\star} \\ \gamma^{2\star} \end{pmatrix} + \begin{pmatrix} \boldsymbol{X}_z \\ \boldsymbol{X}_z \end{pmatrix} \boldsymbol{z} + \begin{pmatrix} \boldsymbol{\epsilon}^1 \\ \boldsymbol{\epsilon}^2 \end{pmatrix}, \quad (4.4)$$

which is the final form for our data analysis.

Following this reparameterization, M_i is reparameterized as:

$$\boldsymbol{v}^i = \mathbf{1}_{JK_i}\mu^k + \boldsymbol{X}_{\beta_0}^{i\star}\boldsymbol{\beta}_0 + \boldsymbol{Q}^i\gamma^{i\star} + \boldsymbol{X}_z\boldsymbol{z} + \boldsymbol{\epsilon}^i, i = 1, 2. \quad (4.5)$$

Candidate Model 2

A study effect γ^3 with design matrix $\boldsymbol{X}_\gamma^{3\star} = (\mathbf{1}'_{JK_1}, -\mathbf{1}'_{JK_2})'$ is added to quantify potential differences between two data sources. After the QR decomposition, $\boldsymbol{X}_\gamma^{3\star}\gamma^3 =$

$\boldsymbol{Q}^3\boldsymbol{R}^3\gamma^3 = \boldsymbol{Q}^3\gamma^{3\star}$, where $\boldsymbol{Q}^{3\prime}\boldsymbol{Q}^3 = 1$. Then, M_{c_2} is specified as:

$$\begin{pmatrix} \boldsymbol{v}^1 \\ \boldsymbol{v}^2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{1}_{JK_1} & \boldsymbol{X}^{1\star}_{\beta_0} \\ \boldsymbol{1}_{JK_2} & \boldsymbol{X}^{2\star}_{\beta_0} \end{pmatrix} \begin{pmatrix} \mu \\ \boldsymbol{\beta}_0 \end{pmatrix} + \begin{pmatrix} \boldsymbol{Q}^1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}^2 \end{pmatrix} \begin{pmatrix} \gamma^{1\star} \\ \gamma^{2\star} \end{pmatrix} + \boldsymbol{Q}^3\gamma^{3\star} + \begin{pmatrix} \boldsymbol{X}_z \\ \boldsymbol{X}_z \end{pmatrix} \boldsymbol{z} + \begin{pmatrix} \boldsymbol{\epsilon}^1 \\ \boldsymbol{\epsilon}^2 \end{pmatrix}, \tag{4.6}$$

where $\gamma^{3\star}$ is used to account for the potential systematic differences between two data sources described in Section 4.2,

4.3.3 Prior Distributions

The non-informative prior is used for overall mean $\mu \propto 1$, independent ZS priors are used for $\boldsymbol{\beta}_0$ and γ^i. The independent form of ZS prior is also recommended by Rouder et al. (2012) for g-prior in terms of fix effects, where each factor is modeled with a separate g parameter. Prior distributions are summarized as below:

$$\boldsymbol{\beta}_0 | g_0 \sim N_{p_0}\left(\boldsymbol{0}, g_0\sigma^2(\boldsymbol{X}^{\star}_{\beta_0}{}'\boldsymbol{X}^{\star}_{\beta_0})^{-1}\right), \tag{4.7}$$

$$\boldsymbol{\gamma}^{i\star} | g_i \sim N_{K_i - 1}(0, g_i\sigma^2), \boldsymbol{\gamma}^{3\star} | g_3 \sim N(0, g_3\sigma^2). \tag{4.8}$$

For the random spatial effect $\boldsymbol{z}$, many advanced techniques have been developed to suit different spatial structures or considerations (Besag et al., 1991; Dean et al., 2001; Lee and Mitchell, 2012; Leroux et al., 2000; MacNab, 2022; Simpson et al., 2017). As it is not our primary focus, we adopted one popular method, the conditional autoregressive (CAR) model, to provide insights on how random effects are impacted in our data combining framework. CAR model is shown to have guaranteed posterior propriety (Sun et al., 2004; Woodard et al., 1999), and it assumes that counties with shared boundaries are spatially correlated with the following form:

$$\boldsymbol{z}|\delta, \rho \sim N_I(0, \delta(\boldsymbol{I} - \rho \boldsymbol{C})^{-1}), \rho \in \left(\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}\right), \tag{4.9}$$

where $\boldsymbol{C}$ is the adjacency matrix, with element $c_{ij} = 1$ if county i and county j are adjacent, and 0 otherwise. $\lambda_{min}, \lambda_{max}$ are minimum and maximum eigenvalues of $\boldsymbol{C}$, respectively.

For the scale parameter δ in the distribution of spatial effect $\boldsymbol{z}$ and random error σ^2, Inverse-Gamma (IG) distribution is used:

$$f(\sigma^2|a, b) \propto \frac{1}{(\sigma^2)^{a+1}} exp(-\frac{b}{\sigma^2}), \sigma^2 > 0, \tag{4.10}$$

$$f(\delta|a_0, b_0) \propto \frac{1}{(\delta)^{a_0+1}} exp(-\frac{b_0}{\delta}), \delta > 0. \tag{4.11}$$

A uniform prior is used for ρ to ensure $\delta(\boldsymbol{I} - \rho \boldsymbol{C})^{-1}$ is positive definite:

$$\rho \sim Unif\left(\frac{1}{\lambda_{min}}, \frac{1}{\lambda_{max}}\right) = Unif\left(\rho_{min}, \rho_{max}\right). \tag{4.12}$$

IG distributions are used for g_0, g_1, g_2 and g_3:

$$f(g_0) \propto \frac{1}{g_0^{3/2}} exp(-\frac{JK}{2g_0}), g_0 > 0, \tag{4.13}$$

$$f(g_i) \propto \frac{1}{g_i^{3/2}} exp(-\frac{1}{2g_i}), g_i > 0. \tag{4.14}$$

For σ^2, we set $a = b = 0$, that's to say, $f(\sigma^2) \propto 1/\sigma^2$. For δ, we set $a_0 = b_0 = 1$.

For computation, a Markov Chain Monte Carlo (MCMC) method was used to generate samples of posterior distributions. Data aggregation was carried out by

SEER*Stat and SAS software for MCR and CLS, respectively. Sampling algorithms were implemented in R. We fit M_1, M_2, M_{c_1}, and M_{c_2}, separately. For each model, we used 50,000 samples after discarding the first 20,000 ones. We collect posterior mean M_p, posterior standard deviation (SD_p) and 95% credible interval (CI) for each model. Since true values for regression coefficients are not available for real data, we adopt the mean absolute difference (MAD) between the observed value v^i_{jk} and estimated value $\hat{v}^i_{jk}$ and correlation between v^i_{jk} and $\hat{v}^i_{jk}$, denoted as COR in Tables. A model with a lower MAD and a higher COR indicates a better model.

4.3.4 Results

Figures 4.5-4.8 shows the estimated against observed responses for MCR, CLS, M_{c_1}, and M_{c_2}, respectively. Figure 4.9 maps SD_p for $\boldsymbol{z}$ to visualize the impacts of data combining on the random effects. M_p, SD_p, 95% CIs, MAD and COR for key parameters calculated from individual models and combined models are summarized in Tables 4.4 and 4.5, respectively. To compare results from Tables 4.4 and 4.5, main findings are presented from three perspectives. First, we compare overall performance of using the combined data and individual data. Second, we examine two different combining methods (M_{c_1} and M_{c_2}). Third, based on M_{c_2}, we interpret associations between FBC prevalence and covariates.

First, combining data is not all always beneficial for both data sources due to different variability inherited in data sources. However, data combining, even if the most naive one such as M_{c_1}, could still be advantageous for the data source with large variability such as CLS data if a researcher is interested in smaller SD_p and shorter CI. We would recommend researchers to use administrative data to improve the precision of estimates based on survey data. Second, when systematic difference appears in data sources, a study effect should be included to improve the overall model performance, which can be consolidated by the exclusion of 0 for the 95%CI of γ^3. For example, for CLS, the inclusion of a study effect γ^3 lead to a decrease of MAD from 0.4313 to 0.3411, and an increase of COR from 0.7597 to 0.8633.

At last, as M_{c_2} is the best model in terms of smaller MAD and larger COR, we use results from M_{c_2} to interpret the relationship between the covariates and FBC prevalence in its logit form. For county attributes $\boldsymbol{\beta}_0$ (percentages of age 65+, people below poverty level), posterior means are not far from zero. For γ^1 stage factor, the posterior mean is 2.2024 with 95% CI exclude 0, which indicate the localized stage is associated with more prevalent cases, which can be explained by higher survival rates for early stage cancer. For γ^2, the posterior mean is -0.5098 with 95% CI excludes 0, which indicates people who had no cholesterol is associated with lower FBC prevalence.

Additionally, γ^3, the study effect, has a posterior mean of -0.5290 with 95% CI exclude 0. This is the evidence that, given other covariates, FBC prevalence of MCR is below overall mean while CLS is above. Reasons behind the lower FBC prevalence for MCR could be: (1) FBC prevalence for MCR is LDP, which shows existing cases in a shorter time period compared with CP in CLS; (2) Medical definition difference about "cancer" in MCR and CLS. MCR only collects malignant cancer cases except for brain cancer while an interviewee might consider both benign and in Situ tumor as cancer cases in CLS; (3) MCR does not get all cases that should be reported. For county effects $\boldsymbol{z}$, the posterior distribution for spatial smoothing parameter ρ is slightly right-skewed but centered around zero. This indicates the spatial structure

Table 4.4: M_p and SD_p for individual models with LMM

Data Source	MCR		CLS	
Parameter	$M_p(SD_p)$	95% CI	$M_p(SD_p)$	95% CI
μ	-4.5470 (0.0230)	(-4.5934, -4.5026)	-3.2943 (0.0556)	(-3.4004, -3.1809)
β_{01}	7.9498 (3.7834)	(1.5330, 16.0593)	-0.4982 (1.2526)	(-3.0794, 1.9536)
β_{02}	-3.1575 (1.5083)	(-6.432, -0.5878)	-0.9286 (0.5720)	(-2.1626, 0.0666)
$\gamma^{1\star}$	-5.2476 (0.2937)	(-5.8307, -4.6701)	-	-
$\gamma^{2\star}$	-	-	3.8808 (0.6931)	(2.4874, 5.2311)
γ^1	2.4448 (0.1368)	(2.1758, 2.7164)	-	-
γ^2	-	-	-0.4169 (0.0745)	(-0.5620, -0.2672)
σ^2	0.0701 (0.0078)	(0.0567, 0.0867)	0.4188 (0.0547)	(0.3280, 0.5390)
δ	0.3803 (0.1542)	(0.1616, 0.7599)	0.3158 (0.0966)	(0.1632, 0.5336)
ρ	0.0886 (0.0947)	(-0.1739, 0.1714)	-0.0401 (0.1047)	(-0.2514, 0.1363)
MAD	0.1975		0.4313	
COR	0.8377		0.7597	

in (4.9) may not be helpful in our model. De Oliveira (2012) pointed out that the uniform prior in (4.12) assigns little mass when there is substantial spatial correlation, and much mass when there is weak or no spatial correlation.

4.4 Generalized Linear Mixed Model (GLMM)

Consider possible loss in the accuracy during the logit transformation in Section 4.3 for prevalent percentages, here, we model via the observed prevalent counts and population size directly. As displayed in Table 4.2, the data do not have zero responses.

4.4.1 Model Specifications

Following the same notations as in Section 4.3, for the i-th data source, j-th county, and k-th category, we consider the prevalent counts y^i_{jk} as random variable and the

Table 4.5: M_p and SD_p for M_{c_1} and M_{c_2} from LMM

Data Source	M_{c_1}		M_{c_2}	
Parameter	$M_p(SD_p)$	95% CI	$M_p(SD_p)$	95% CI
μ	-4.5597 (0.0668)	(-4.6897, -4.4261)	-4.2758 (0.0369)	(-4.3479, -4.2032)
β_{01}	0.0998 (1.2608)	(-2.4440, 2.6056)	-0.3090 (1.1436)	(-2.6468, 1.9451)
β_{02}	0.2091 (0.4270)	(-0.5893, 1.1244)	0.0749 (0.3768)	(-0.6729, 0.8411)
$\gamma^{1\star}$	-6.5078 (0.5597)	(-7.6049, -5.4284)	-5.9169 (0.5255)	(-6.9347, -4.8697)
$\gamma^{2\star}$	4.9069 (0.5829)	(3.7510, 6.0398)	3.6664 (0.5527)	(2.5639, 4.7422)
$\gamma^{3\star}$	-	-	15.2490 (1.0166)	(13.2970, 17.2361)
γ^1	3.0319 (0.2608)	(2.5290, 3.5430)	2.2024 (0.2448)	(2.2687, 3.2308)
γ^2	-0.5272 (0.0626)	(-0.6489, -0.4030)	-0.5098 (0.0594)	(-0.5095, -0.2755)
γ^3	-	-	-1.6052 (0.1070)	(-1.8144, -1.3998)
σ^2	0.3074 (0.0255)	(0.2620, 0.3616)	0.2667 (0.0207)	(0.2290, 0.3096)
δ	1.1207 (0.2351)	(0.7214, 1.6417)	0.7466 (0.1584)	(0.4683, 1.0881)
ρ	0.1716 (0.0015)	(0.1675, 0.1732)	-0.0798 (0.0855)	(-0.2447, 0.0825)
MAD	0.3724		0.3411	
COR	0.8464		0.8633	

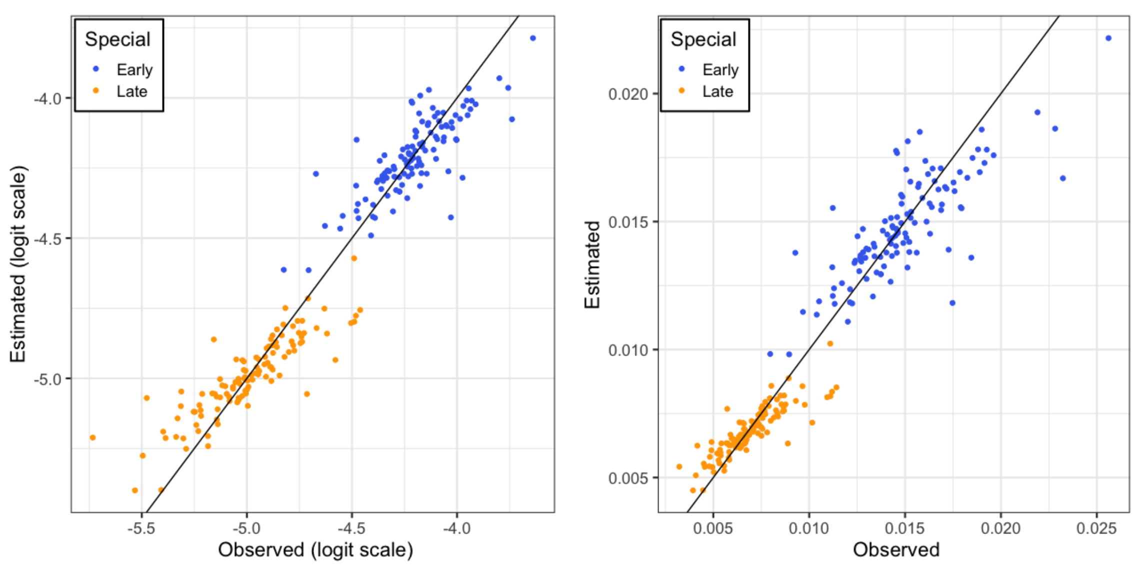

Figure 4.5: Estimated against the observed responses in logit (left) and percentage (right) form from MCR with LMM.

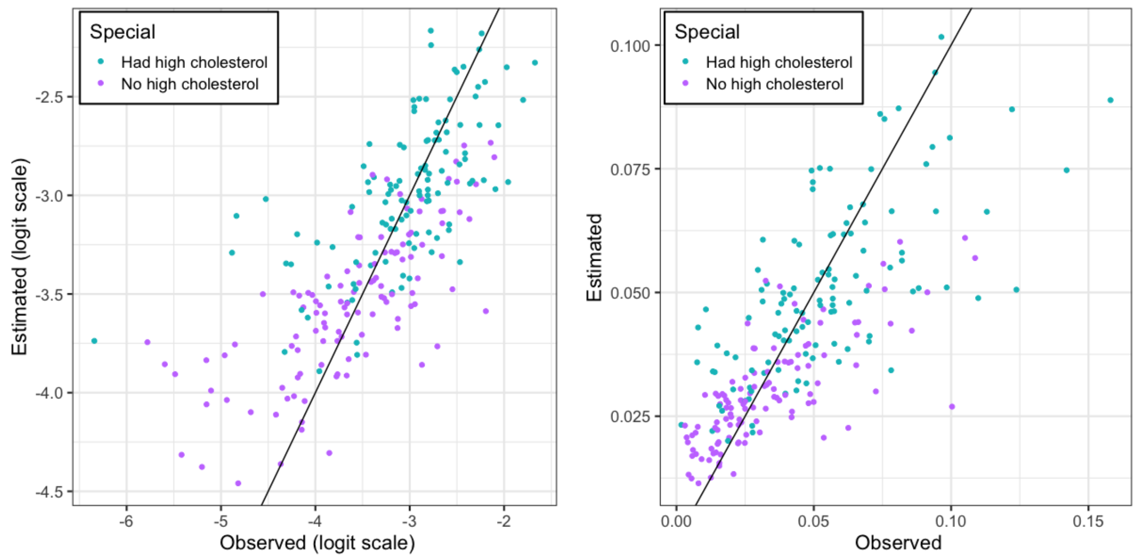

Figure 4.6: Estimated against the observed responses in logit (left) and percentage (right) form from CLS with LMM.

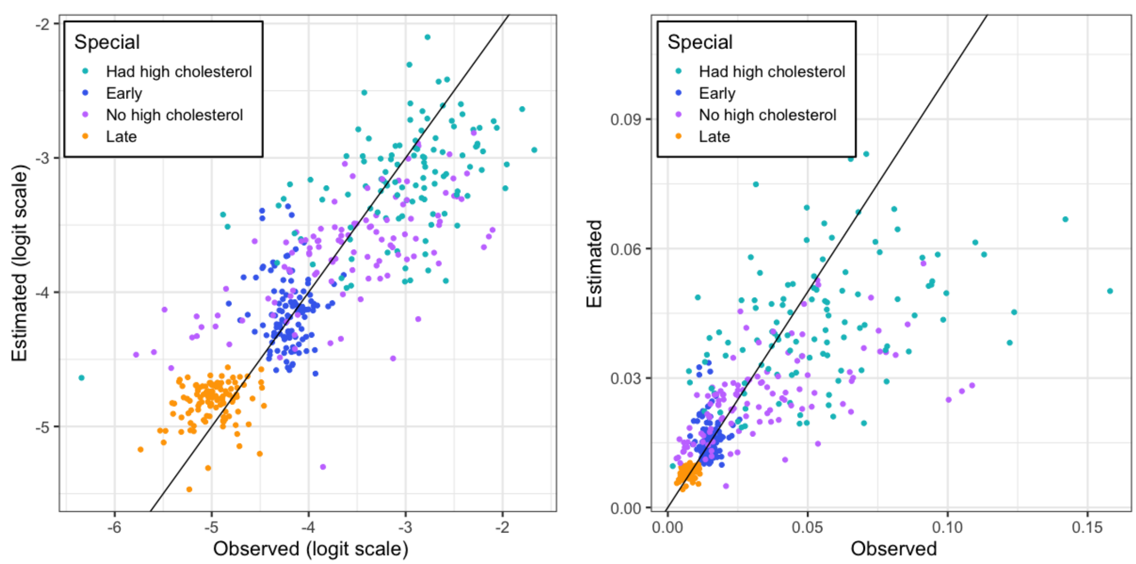

Figure 4.7: Estimated against the observed responses in logit (left) and percentage (right) form from M_{c_1} with LMM.

population size n^i_{jk} as known values. Recall that $i = 1$ indicates data from MCR and two categories are early vs late stage. $i = 2$ indicates data from CLS and

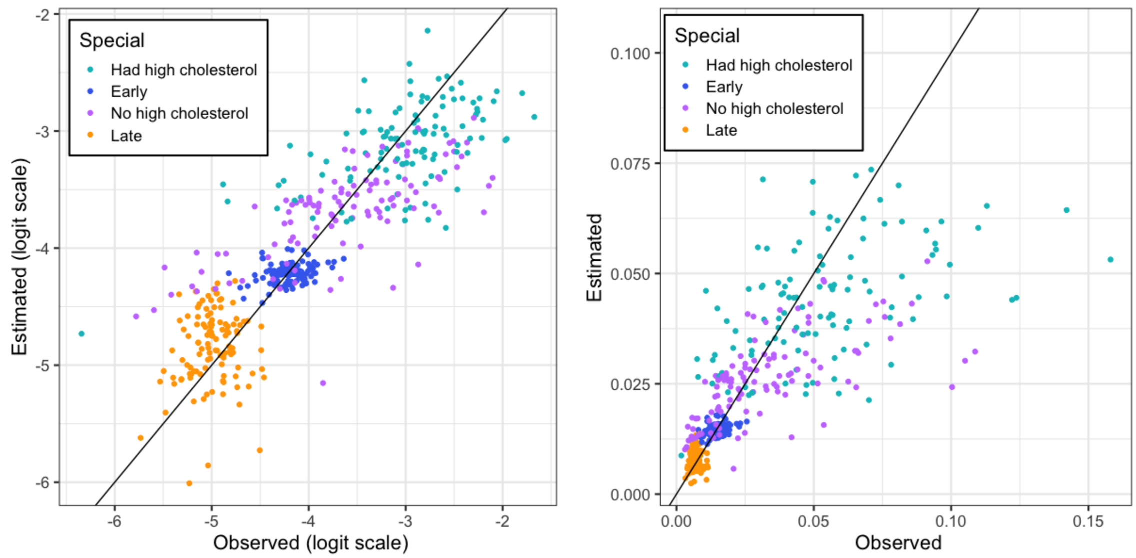

Figure 4.8: Estimated against the observed responses in logit (left) and percentage (right) form from M_{c_2} with LMM.

two categories are whether a person is told of high cholesterol level or not by health professional. There is no zero prevalence count in the data. To ensure the consistency of notations, we use n^i_{jk} for $i = 1, 2$. Notice that, when $n^1_{i1} = n^1_{i2} = n^1_i$.

For the individual data source i, we assume that

$$y^i_{jk} \sim Bin(n^i_{jk}, p^i_{jk}), v^i_{jk} = logit(p^i_{jk}) = log(\frac{p^i_{jk}}{1-p^i_{jk}}) \tag{4.15}$$

$$v^i_{jk} = \mu^i + \beta_{01}x_{1j} + \beta_{02}x_{2j} + \gamma^i_k + z_j + \epsilon^i_{jk}. \tag{4.16}$$

The matrix form of model in (4.16) is rewritten as

$$\boldsymbol{v}^i = \mu^i \mathbf{1}_{JK_i} + \boldsymbol{X}^i_{\beta_0}\boldsymbol{\beta}_0 + \boldsymbol{X}^i_{\gamma}\boldsymbol{\gamma}^i + \boldsymbol{X}_z\boldsymbol{z} + \boldsymbol{\epsilon}^i. \tag{4.17}$$

The same reparameterization strategy as in Section 4.3.2 is applied to models in this

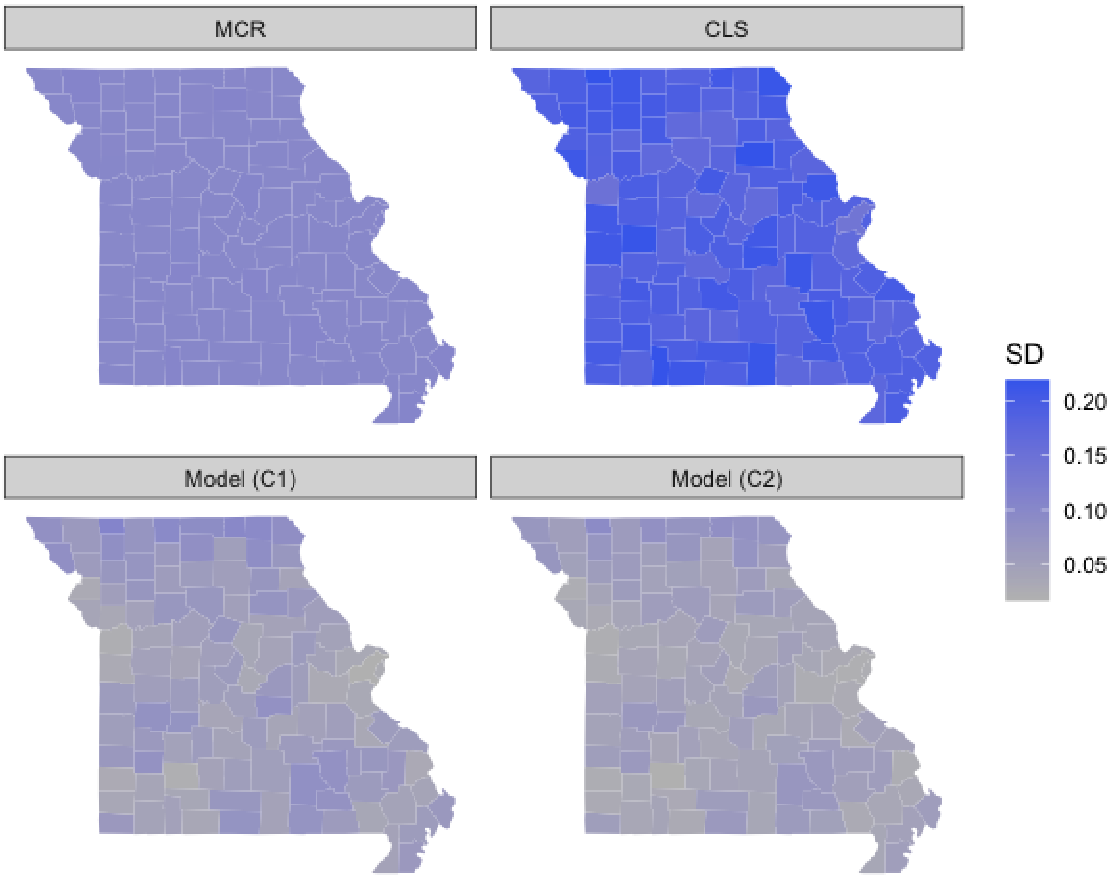

Figure 4.9: Map of SD_p of $\boldsymbol{z}$ for M_1, M_2, M_{c_1}, and M_{c_2} with LMM.

section. Additionally, since we incorporate the aggregate-level county attributes in the model, there is potential confounding in the model. In a meta-analytical framework for aggregate-level data, it is more frequently called "ecological fallacy" (Chen et al., 2020; Cooper and Patall, 2009). In the context of spatial analysis, the confounding issue is referred to as “spatial confounding” ,and the models in (4.15) and (4.16) are referred to as the spatial generalized linear mixed model (SGLMM).

Specifically, the phenomenon of multicollinearity among spatial covariates and the spatial random effect is referred to as “spatial confounding” (Paciorek, 2010). When a researcher is interested in the interpretation of the relationship between spatial covariates and a response, the spatial confounding can have a significant effect on regression parameters in SGLMM. Although no universal solution exists, one popular method to relieve this problem is the restricted spatial regression (RSR) (Hanks et al., 2015; Hodges and Reich, 2010; Hughes and Haran, 2013), which constrains the random effects to be orthogonal to fixed effects. It has been shown that, conditioned on the spatial effects, RSR is a reparameterization of the SGLMM. For example,

$$\begin{aligned}
\boldsymbol{\eta} &= \boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0 + \boldsymbol{X}_z\boldsymbol{z} \\
&= \boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0 + \boldsymbol{P}_{\boldsymbol{X}_{\beta_0}}\boldsymbol{X}_z\boldsymbol{z} + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{\beta_0}})\boldsymbol{X}_z\boldsymbol{z} \\
&= \boldsymbol{X}_{\beta_0}[\boldsymbol{\beta}_0 + (\boldsymbol{X}'_{\beta_0}\boldsymbol{X}_{\beta_0})^{-1}\boldsymbol{X}'_{\beta_0}\boldsymbol{X}_z\boldsymbol{z}] + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{\beta_0}})\boldsymbol{X}_z\boldsymbol{z} \\
&= \boldsymbol{X}_{\beta_0}\tilde{\boldsymbol{\beta}}_0 + (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{\beta_0}})\boldsymbol{X}_z\boldsymbol{z},
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_0 = \boldsymbol{\beta}_0 + (\boldsymbol{X}'_{\beta_0}\boldsymbol{X}_{\beta_0})^{-1}\boldsymbol{X}'_{\beta_0}\boldsymbol{X}_z\boldsymbol{z}$ represents marginal regression coefficients and $\boldsymbol{\beta}_0$ represents conditional regression coefficients. They also recommended that, without strong belief that regression coefficients are orthogonal to the random effects, SGLMM is a better option. The samples of $\tilde{\boldsymbol{\beta}}_0$ can be obtained through the relationship between $\tilde{\boldsymbol{\beta}}_0$ and $\boldsymbol{\beta}_0$ when performing MCMC.

With (4.15)-(4.17), following Section 4.3, two data combining methods are considered. If the data is combined with (4.4), we denote the model as M_{c_1}. If the data is combined with (4.6), the model is referred to as M_{c_2}.

4.4.2 Prior Distributions

Prior specifications are the same as Section 4.3. Non-informative prior is used for overall mean, $\mu \propto 1$. For other parameters, we use priors in (4.7)-(4.13). A MCMC method, such as Gibbs sampling, was used to generate samples of posterior distributions. We only present full conditional distributions for M_{c_2}, and one could easily obtain the full conditional distributions for the rest of candidate models. Since the full conditional distributions for ρ and σ^2 (See 6 and 8 in Appendix A.3) are not standard distributions, Adaptive Rejection Metropolis Sampling (ARMS) by Gilks et al. (1995) was used to obtain the posterior samples. The following hyper parameter values are used: $a_1 = b_1 = 0, a_0 = b_0 = 1$. We fit models using MCR, CLS and the combined data, separately. Therefore, results for four models are reported in total. For each model, we used 50,000 samples after discarding the first 20,000.

4.4.3 Results

Tables 4.6 and 4.7 summarize M_p, SD_p, and 95% CIs for individual models and the combined model. Figures 4.10, 4.11, 4.12, and 4.13 present the estimated responses against observed in both logit scale and percentage scale. Figure 4.14 presents the posterior standard deviations for the spatial effects $\boldsymbol{z}$. To further visualize the estimation of prevalence estimates, Figure 4.15 gives an example of early stage prevalence in a map form. Finally, for a convergence check, Figure 4.16 displays the trace plots for key parameters with M_{c_2}. The trace plots for other models along with those for

LMM show similar patterns and are not presented to save space.

There are several main findings. First, M_{c_1}, M_{c_2}, and M_i nearly reach consistent conclusions for most of the parameters in terms of whether CI covers 0 except the age group β_{01}. Second, M_{c_1} and M_{c_2} offer a smaller SD_p for coefficients compared with CLS but not MCR. This is within our expectation as we have discussed in Section 4.3. Third, M_{c_2} yields more precise estimates compared with M_{c_1} in terms of SD_p. This echoes the result that the 95% CI of $\gamma^{3\star}$ excludes 0. Fourth, for other parameters (σ^2, ρ, δ), M_{c_1} and M_{c_2} show a smaller SD_p compared with both MCR and CLS. In addition, compared with LMM, the estimated prevalence are more close to the observed, which can be reflected by Figures 4.10-4.13. At last, we found that, the estimates for $\boldsymbol{\beta}_0$, with and without the spatial confounding adjustment, are more close in M_{c_1} and M_{c_2} rather than M_1 and M_2. It might be interesting to investigate whether incorporating more data sources could help us relieve the confounding issue if RSR is a proper formulation. We also found that, the marginal estimates for $\boldsymbol{\beta}_0$ have smaller posterior variance compared with the conditional estimates.

4.5 Discussion

This chapter takes a primary investigation of combining FBC prevalence from MCR and CLS under several candidate models with random effects incorporated. Specially, we examine how our data combining framework impact the relationship between the FBC prevalence and covariates of interests. The take home message is that it is essential to understand the nature of data when we intend to perform data synthesis. For example, in our case, we intend to combine data sources with very different

Table 4.6: M_p and SD_p for MCR and CLS from GLMM

Data Source	MCR		CLS	
Parameter	$M_p(SD_p)$	95% CI	$M_p(SD_p)$	95% CI
μ	-4.5955 (0.0092)	(-4.6135, -4.5777)	-3.3218 (0.0433)	(-3.4073, -3.2366)
β_{01}	1.9588 (0.4929)	(0.9858, 2.9131)	0.9689 (1.2727)	(-1.5186, 3.4630)
β_{02}	1.2094 (0.4466)	(0.3651, 2.1274)	2.8414 (1.1571)	(0.5788, 5.0884)
$\tilde{\beta_{01}}$	2.0215 (0.1998)	(1.6237, 2.4114)	1.0718 (0.9908)	(-0.8496, 3.0105)
$\tilde{\beta_{02}}$	1.4003 (0.1794)	(1.0428, 1.7543)	2.7997 (0.9028)	(1.0247, 4.5938)
γ^1	0.3992 (0.0075)	(0.3843, 0.4138)	-	-
γ^2	-	-	-0.2867 (0.0429)	(-0.3712, -0.2034)
γ^3	-	-	-	-
σ^2	0.0028 (0.0011)	(0.0011, 0.0055)	0.4106 (0.0486)	(0.3243, 0.5138)
δ	0.0361 (0.0054)	(0.0270, 0.0480)	0.1390 (0.0339)	(0.0840, 0.2153)
ρ	0.0625 (0.0498)	(-0.0475, 0.1448)	-0.0852 (0.0897)	(-0.2539, 0.0882)

Table 4.7: M_p and SD_p for M_{c_1} and M_{c_2} from GLMM

Data Source	M_{c_1}		M_{c_2}	
Parameter	$M_p(SD_p)$	95% CI	$M_p(SD_p)$	95% CI
μ	-3.9669 (0.0397)	(-4.0448, -3.8892)	-3.9573 (0.0226)	(-4.0020, -3.9126)
β_{01}	1.4714 (1.0908)	(-0.6891, 3.6008)	1.4948 (0.7751)	(-0.0221, 3.0080)
β_{02}	2.1262 (0.9960)	(0.1520, 4.0807)	2.0930 (0.7101)	(0.7022, 3.4934)
$\tilde{\beta_{01}}$	1.4862 (0.9073)	(-0.2949, 3.2688)	1.5374 (0.5194)	(0.5078, 2.5498)
$\tilde{\beta_{02}}$	2.1239 (0.8268)	(0.5000, 3.7298)	2.0996 (0.4754)	(1.1648, 3.0295)
γ^1	0.3878 (0.0559)	(0.2792, 0.4970)	0.3942 (0.0322)	(0.3312, 0.4572)
γ^2	-0.2891 (0.0551)	(-0.3974, -0.1807)	-0.2843 (0.0320)	(-0.3469, -0.2222)
γ^3	-	-	-0.6490 (0.0225)	(-0.6925, -0.6048)
σ^2	0.6944 (0.0489)	(0.6048, 0.7980)	0.2184 (0.0174)	(0.1866, 0.2549)
δ	0.0787 (0.0169)	(0.0511, 0.1171)	0.0721 (0.0135)	(0.0496, 0.1023)
ρ	-0.0313 (0.0819)	(-0.1986, 0.1132)	-0.0290 (0.0747)	(-0.1815, 0.1051)

features in terms of data collection, data measurements, case definition, information collected, and what statistics are used to publish these data. These features enable

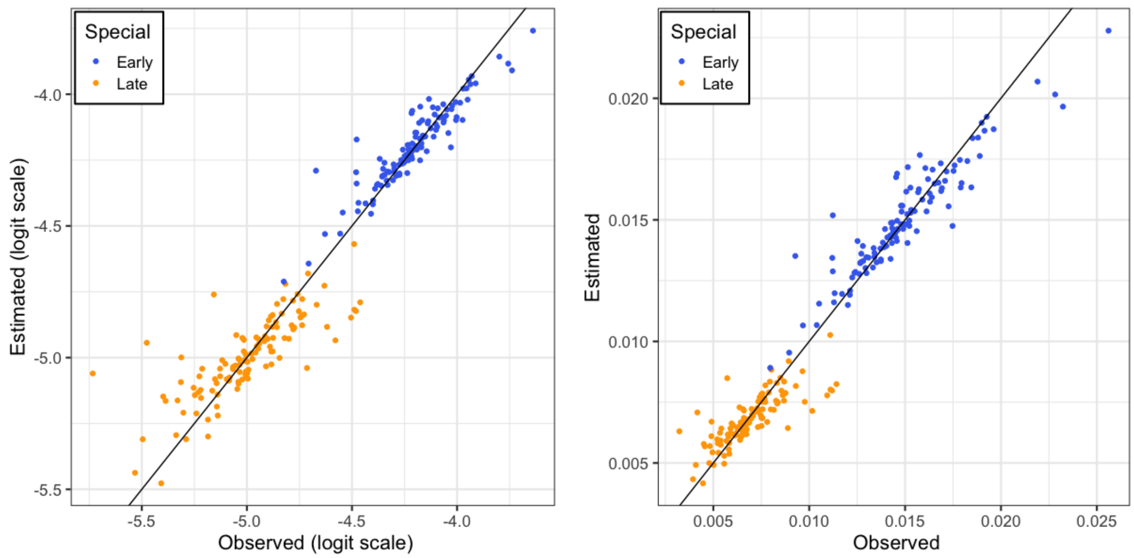

Figure 4.10: Estimated against the observed responses in logit (left) and percentage (right) form from MCR with GLMM.

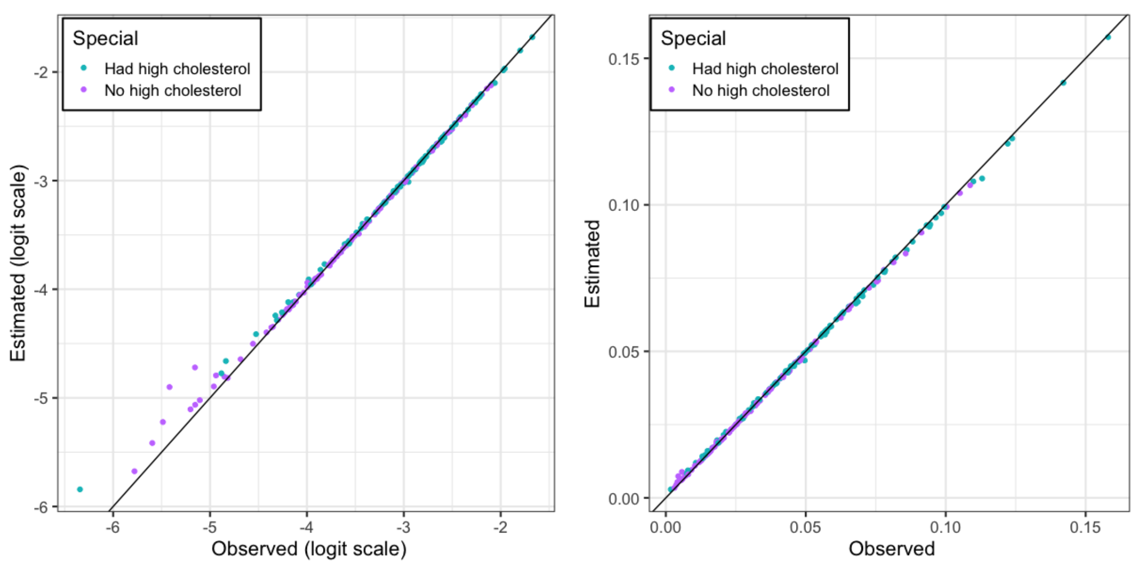

Figure 4.11: Estimated against the observed responses in logit (left) and percentage (right) form from CLS with GLMM.

us to better target appropriate statistical methods and interpret results.

The results with LMM or GLMM also offer implications to health care planners or

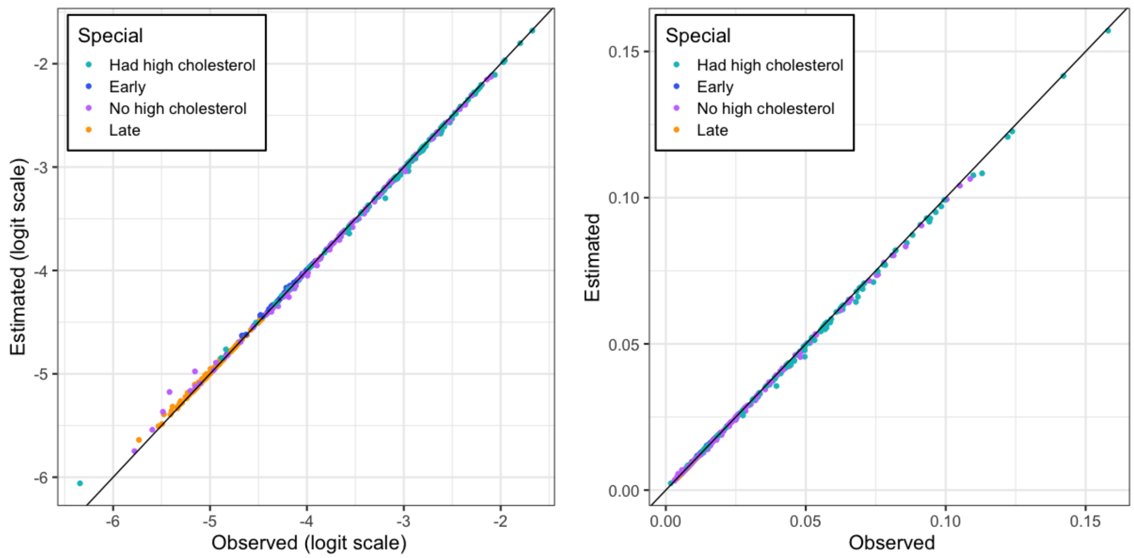

Figure 4.12: Estimated against the observed responses in logit (left) and percentage (right) form from M_{c_1} with GLMM.

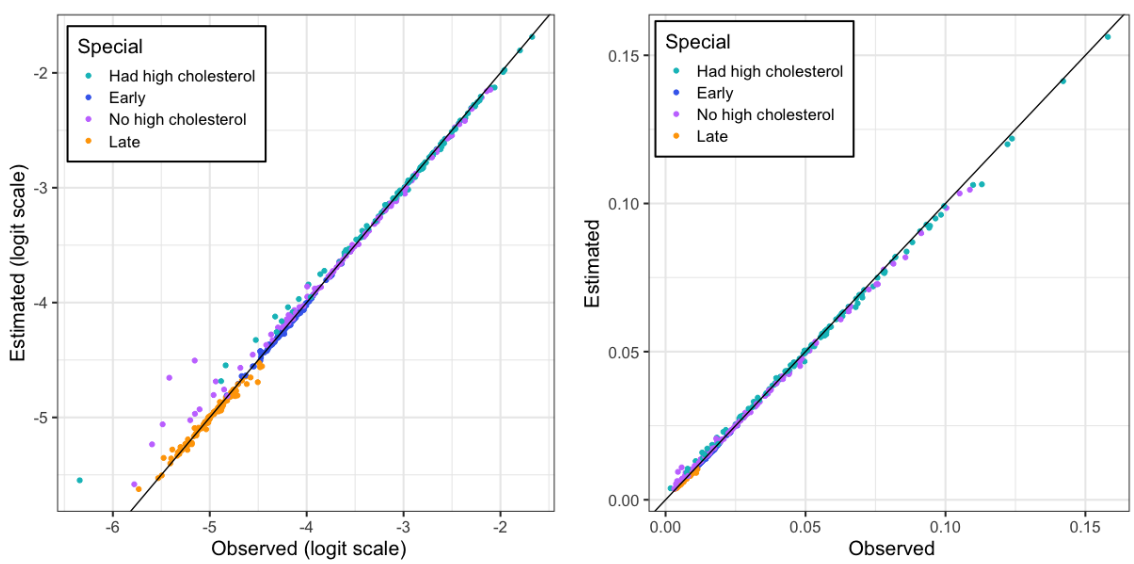

Figure 4.13: Estimated against the observed responses in logit (left) and percentage (right) form from M_{c_2} with GLMM.

policy makers. For one thing, all models indicate that people aged over 65 and at or above poverty level is associated with a higher FBC prevalence from the population

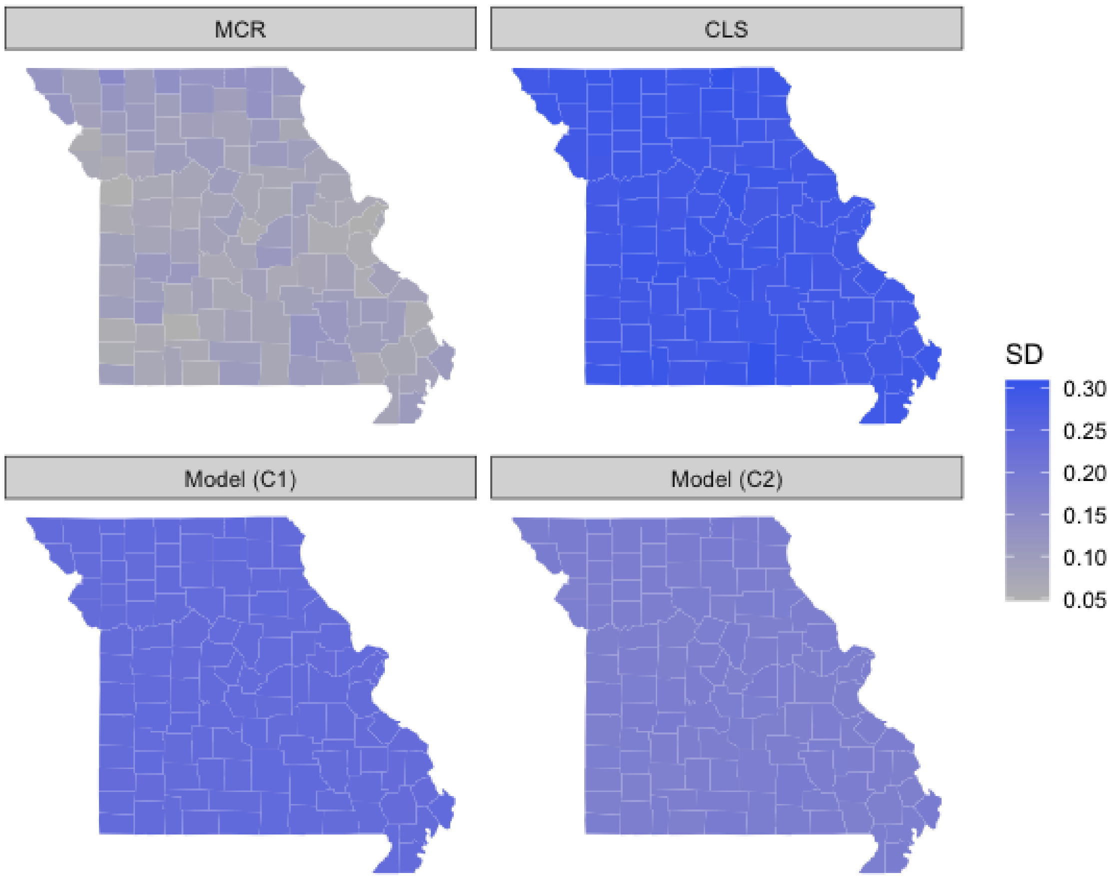

Figure 4.14: Map of SD_p of $\boldsymbol{z}$ for M_1, M_2, M_{c_1}, and M_{c_2} with GLMM

perspective. While the higher prevalence is more related to a higher incidence, people below the poverty level is more related to a higher mortality due to limited access to affordable health care resources. It is helpful to provide accessible medical resources such as early screening for people below the poverty level. For another thing, there

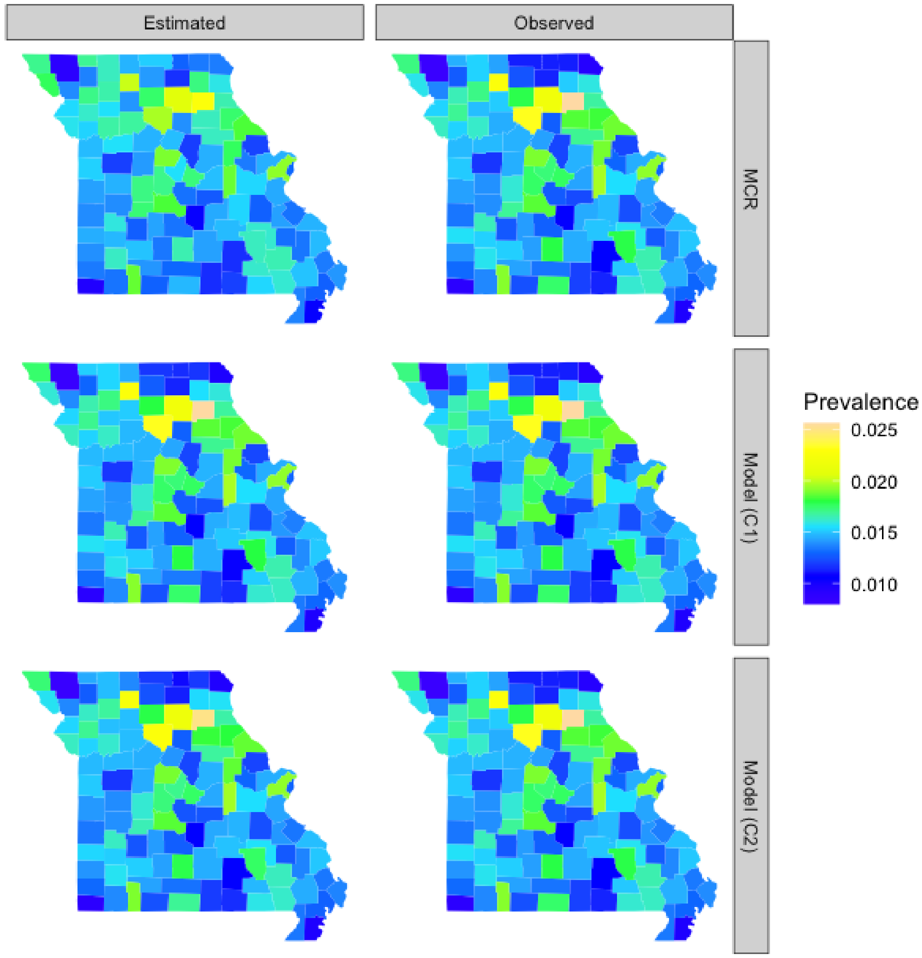

Figure 4.15: Estimated (left) FBC prevalence against the observed (right) for the early stage from M_1, M_{c_1}, and M_{c_2} with GLMM.

is a positive association between the FBC prevalence and early stage, which puts emphasis on cancer prevalence since late stage is more related to higher mortality. Although the specific dynamic between the FBC prevalence and cholesterol level remain unclear due to complicated factors, our results still show that a higher FBC prevalence is associated with a higher cholesterol level from the population perspective. This indicates that appropriate control over the cholesterol level through diet or exercise maybe a helpful way to reduce the FBC prevalence.

There are several aspects worth exploring in the future. To start with, with the proposed data combining framework, the incorporation of random effects and the generalization to counts outcomes cause theoretical challenge on the analyses of posterior variance and frequentist properties. We may adopt approximation method for the theoretical support in some special cases. This makes extensive simulation studies with different structures of random effects, different sample sizes, and different size of coefficients necessary to offer a more complete picture of the behaviors of the Bayesian estimator and posterior variances in terms of the random effects. Second, other data combining strategies should be studied and compared. For example, in the FBC prevalence setting, another interesting framework is to combine data according to different geographical regions and study the correlations among multiple factors. Third, for the spatial structure of the random effects, many spatial structures besides CAR model have been developed, and one key question need to be answered is how to deal with the spatial confounding. This issue has also been identified in the meta-regression or meta-analysis when the aggregate-level data is used. Several ways have been proposed to relieve this problem. For example, one may use individual level data, or one may consider a marginal estimate for the coefficients by RSR. We adopted the later. Alternatively, Page et al. (2017) models the correlation between the spatial effects and the spatial related covariates, which is a very interesting topic to pursue next. Fourth, M_{c_1} or M_{c_2} assumes MCR and CLS shares the same spatial correlation. However, according to the posterior mean of ρ from MCR and CLS, one indicates a positive correlation and the other indicates a negative correlation, and therefore it is more realistic to allow two spatial correlations for the combined data (Du, 2018; Kim et al., 2001; Schmaltz, 2012).

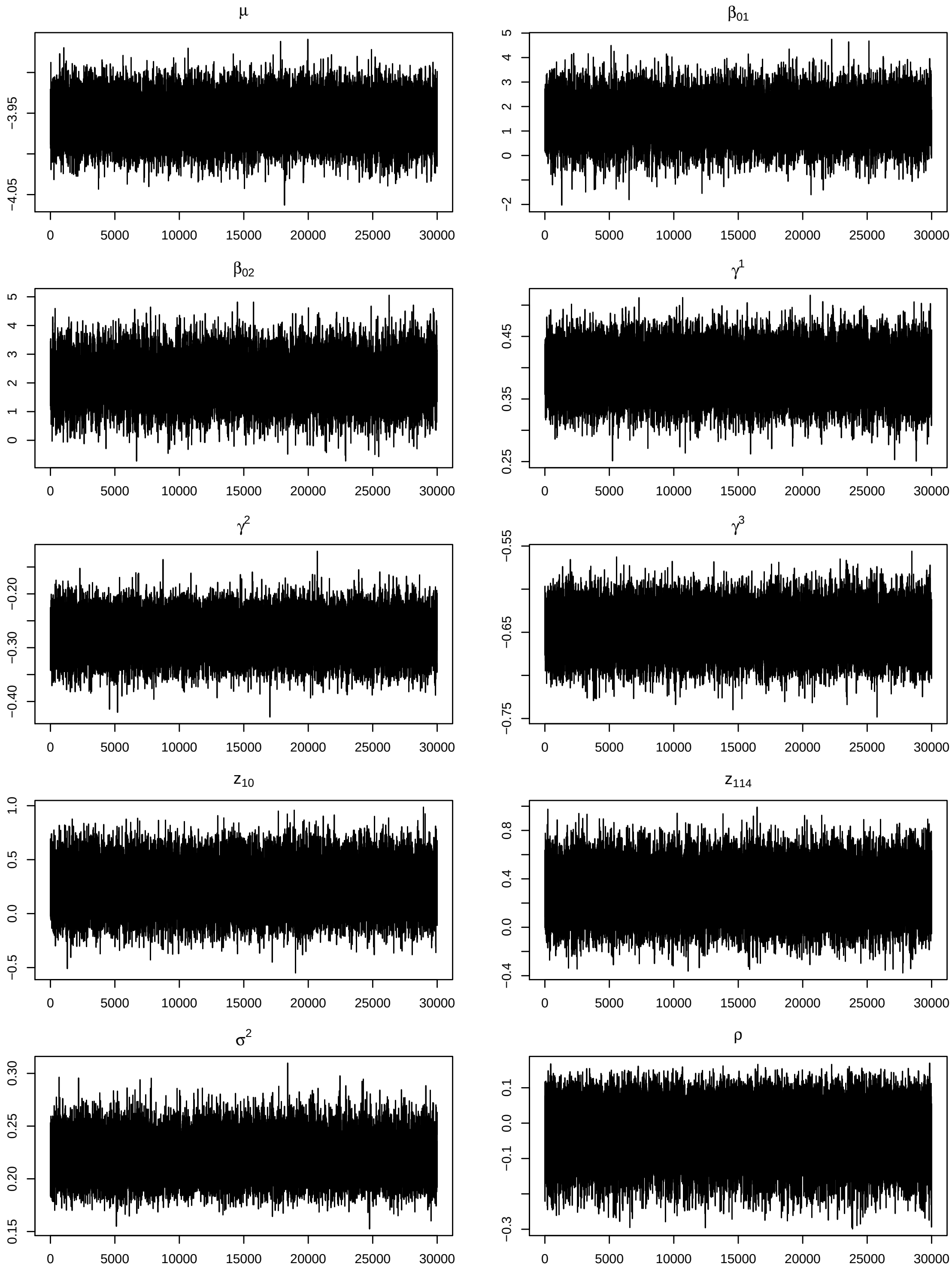

Figure 4.16: Trace plots of selected parameters for M_{c_2} with GLMM

Chapter 5

Summary and Concluding Remarks

This dissertation aims at a theoretical and numerical evaluation of g-prior, ZS prior, and shrinkage prior under M_i and M_c defined in Chapter 1 in terms of posterior variances and frequentist properties of the Bayesian estimators. We also generalize our data combining framework from continuous outcomes and fixed effects model to counts outcomes and mixed effects model through an application on the county-level female breast cancer prevalence. Our methods and results can be extended to data combining with more than two data sources using other g-type priors.

As data combining has become a common practice for researchers, the contributions in this dissertation are three-fold. First, while common data synthesis methods focus on the inference of an overall mean or multiple correlated factors, our data combining framework enables a specification of both shared and model-specific covariates. Compared with the graphical model, our formulation offers a balance between the theoretical justification and model complexity, which is highly needed to draw conclusions of suitability of data combining. We found that M_c performs better M_i when the sample size is small and $\boldsymbol{\beta}_i$ is not dominant. M_c offers more stable estimates, especially when the focus is on $\boldsymbol{\beta}_0$. Second, we explore the performance of g-prior and ZS prior from the estimation perspective, which has been ignored due to its desirable properties in the model selection. Our results indicate that g prior or ZS prior, especially the independent version, can be a good candidate for estimation in terms of a smaller risk compared with least squares estimates. Third, it provides insights of how much "strength" can be borrowed via such data combination. In Chapters 2 and 3, our work formally compares the posterior variance and frequentist properties in M_i and M_c. While Chapter 2 mainly adopt a Laplace approximation approach, Chapter 3 takes a conditional asymptotic approach and addresses its convergence in the analysis of frequentist properties. The conditional asymptotic analyses results implies that, independent ZS prior offers unbiased estimates for large coefficients and substantial shrinkage for small coefficients, which is desired from the estimation perspective (Berger, 1985).

However, several issues deserve further comments and investigations. First, for both M_i and M_c, we assume that σ^2 are common for both sources, which may be not be true in practice. Although allowing different σ^2 for both sources in Chapter 4 contributes little to improve model performance, it is of interest to study a more general covariance structure for ϵ depending on a specific research question. Second, despite the straightforward theoretical justification with M_c, the relative performance of M_c and alternative data combining strategies need to be evaluated. For example, for the application in Chapter 4, it is also feasible to integrate data through geographical region. Then, the difficulties lies in the incorporation of the correlation between stage and cholesterol information, as well as the systematic difference among MCR and CLS.

Third, our framework reveals that, when there are more specific coefficients and the specific coefficients are dominant in size, the benefits of our data combining strategy is not much, and alternatives need to be explored. For example, from the missing data perspective, whether imputations from external data sources for "missing" covariates could improve the estimation. Fourth, for the specification of g-prior, we utilize zero as its mean for the shrinkage purpose and alignment of model selection. If the priority lies in the estimation and shrinkage is not a primary consideration, one may generalize the mean zero to an unknown parameter and specify a prior accordingly to adjust the magnitude of the shrinkage effects.

Appendix A

Theorems and Lemmas

A.1 Proofs of Theorems, Lemmas, Remarks and Facts in Chapter 2

A.1.1 Proof of Theorem 2.1

By the Woodbury matrix identity,

$$(\boldsymbol{A}+\boldsymbol{U}\boldsymbol{C}\boldsymbol{V})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}\boldsymbol{A}^{-1},$$

where $\boldsymbol{A}, \boldsymbol{U}, \boldsymbol{C}$ and $\boldsymbol{V}$ all denote matrices of the conformable sizes. If we let $\boldsymbol{U} = \boldsymbol{B}', \boldsymbol{C} = \boldsymbol{I}, \boldsymbol{V} = \boldsymbol{B}$, then $(\boldsymbol{A}+\boldsymbol{B}'\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{B}'(\boldsymbol{I}+\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}')^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$. In our case, we set $\boldsymbol{A}_i = \boldsymbol{X}_{0i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}$, and $\boldsymbol{B}_j = (\boldsymbol{I}_{n_j} - \boldsymbol{P}_j)\boldsymbol{X}_{0j}$, where $i = 1, 2, j =$

1, 2 and $i \neq j$, then

$$
\begin{aligned}
& VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \\
=& \frac{g\sigma^2}{1+g}(\boldsymbol{A}_i + \boldsymbol{B}_j'\boldsymbol{B}_j)^{-1} - \frac{g_i\sigma^2}{1+g_i}\boldsymbol{A}_i^{-1} \\
=& \frac{g\sigma^2}{1+g}\{\boldsymbol{A}_i^{-1} - \boldsymbol{A}_i^{-1}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\} - \frac{g_i\sigma^2}{1+g_i}\boldsymbol{A}_i^{-1} \\
=& \left(\frac{g}{1+g} - \frac{g_i}{1+g_i}\right)\sigma^2\boldsymbol{A}_i^{-1} - \frac{g\sigma^2}{1+g}\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}.
\end{aligned}
$$

Hence, $VAR(\boldsymbol{\beta}_0|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$ is equivalent to

$$
\left[1 - \frac{g_i(1+g)}{(1+g_i)g}\right]\boldsymbol{I}_{p_0} \leq \boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}. \tag{A.1}
$$

Since $\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}} < \boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}$, the eigenvalues of $\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}$ are less than 1 or 0. Inequality (A.1) is equivalent to

$$
\left[1 - \frac{g_i(1+g)}{(1+g_i)g}\right] \leq \lambda_{min}(\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}) \in (0, 1).
$$

Notice that $\lambda_{min}(\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}) \neq 0$ because it is of full rank p_0. If we further assume the ordered eigenvalues of $\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'$ is $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_J$, where $J = n_j$, then

$$
\begin{aligned}
\lambda_{min}[\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}] &= \lambda_{min}[(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'] \\
&= \frac{\lambda_1}{1+\lambda_1}.
\end{aligned}
$$

Since $\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'$ depends on the rank of $\boldsymbol{B}_j$, $\lambda_1 = 0$ if and only if $\boldsymbol{B}_j$ is not of full column rank.

For the specific regression coefficient $\boldsymbol{\beta}_i$, suppose $\boldsymbol{Q}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{X}_{0i}$, $\boldsymbol{M}_i = [(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} + \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i']$, and $\boldsymbol{N} = \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'[\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j']^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i'$, then

$$\begin{aligned} & VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \\ &= \left(\frac{g}{1+g} - \frac{g_i}{1+g_i}\right)\sigma^2\boldsymbol{M}_i - \frac{g}{1+g}\sigma^2\boldsymbol{N}. \end{aligned} \tag{A.2}$$

Equation (A.2) can be simplified based on the rank of $\boldsymbol{Q}_i$ since the rank of $\boldsymbol{N}$ depends on $\boldsymbol{Q}_i$.

When $\boldsymbol{X}_i'\boldsymbol{X}_{0i} = \boldsymbol{0}_{n_i\times p_0}$ or $p_i > p_0$, $\boldsymbol{Q}_i$ is either equal to $\boldsymbol{0}_{p_i\times p_0}$ or of full column rank and $\boldsymbol{N}$ is either $\boldsymbol{0}_{p_i\times p_i}$ or not of full rank, and therefore

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

is equivalent to

$$g_i \geq g.$$

When $\boldsymbol{X}_i'\boldsymbol{X}_{0i} \neq \boldsymbol{0}_{n_i\times p_0}$ and $p_i \leq p_0$, $\boldsymbol{Q}_i$ is of full row rank and $\boldsymbol{N}$ is of full rank.

$$VAR(\boldsymbol{\beta}_i|\sigma^2, g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|\sigma^2, g_i, \boldsymbol{y}_i, M_i) \leq 0$$

is equivalent to

$$\left(1 - \frac{g_i(1+g)}{g(1+g_i)}\right)\boldsymbol{I}_{p_i} \leq \boldsymbol{M}_i^{-\frac{1}{2}}\boldsymbol{N}\boldsymbol{M}_i^{-\frac{1}{2}},$$

which is equivalent to

$$1 - \frac{g_i(1+g)}{g(1+g_i)} \leq \lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}).$$

Furthermore, let $\boldsymbol{X}$ be a matrix and $\lambda_i(\boldsymbol{X})$ denote its *i-th* eigenvalue, where $\lambda_i \in \{\lambda_1, \cdots, \lambda_{n-1}, \lambda_n\}$ and $\lambda_1 \geq \cdots \geq \lambda_n$. Notice that

$$\boldsymbol{N} < \boldsymbol{Q}_i \boldsymbol{A}_i^{-1} \boldsymbol{B}_j' [\boldsymbol{B}_j \boldsymbol{A}_i^{-1} \boldsymbol{B}_j']^{-1} \boldsymbol{B}_j \boldsymbol{A}_i^{-1} \boldsymbol{Q}_i' \leq \boldsymbol{Q}_i \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{Q}_i', \boldsymbol{M}_i^{-1} < (\boldsymbol{Q}_i \boldsymbol{A}_i^{-1} \boldsymbol{Q}_i')^{-1},$$

$$\begin{aligned}
\lambda_i(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}) &< \lambda_i(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{Q}_i \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{Q}_i' \boldsymbol{M}_i^{-\frac{1}{2}}) \\
&= \lambda_i(\boldsymbol{Q}_i \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{Q}_i' \boldsymbol{M}_i^{-1}) \\
&= \lambda_i(\boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{Q}_i' \boldsymbol{M}_i^{-1} \boldsymbol{Q}_i \boldsymbol{A}_i^{-\frac{1}{2}}) \\
&< \lambda_i(\boldsymbol{A}_i^{-\frac{1}{2}} \boldsymbol{Q}_i' (\boldsymbol{Q}_i \boldsymbol{A}_i^{-1} \boldsymbol{Q}_i')^{-1} \boldsymbol{Q}_i \boldsymbol{A}_i^{-\frac{1}{2}}). \qquad \text{(A.3)}
\end{aligned}$$

Notice that equation (A.3) is a projection matrix and hence the minimum eigenvalue of $\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}$ is controlled by a projection matrix, whose eigenvalues can be either 0 or 1, which indicates $\lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}) \in (0, 1)$.

A.1.2 Proof of Remark 2.2

Consider posterior variances for $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i | \boldsymbol{y}_i, M_i)$ as well as $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_i | \boldsymbol{y}, M_c)$. Recall that $\boldsymbol{\Sigma}_i = g_i \sigma^2 (\tilde{\boldsymbol{X}}_i{}' \tilde{\boldsymbol{X}}_i)^{-1}/(1+g_i)$ in equation (2.2) and $\boldsymbol{\Sigma} = g\sigma^2(\tilde{\boldsymbol{X}}' \tilde{\boldsymbol{X}})^{-1}/(1+g)$ in equation (2.5), where

$$\tilde{\boldsymbol{X}}_i{}'\tilde{\boldsymbol{X}}_i = \begin{pmatrix} \boldsymbol{X}_{0i}'\boldsymbol{X}_{0i} & \boldsymbol{X}_{0i}'\boldsymbol{X}_i \\ \boldsymbol{X}_i'\boldsymbol{X}_{0i} & \boldsymbol{X}_i'\boldsymbol{X}_i \end{pmatrix} \text{ and } \tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}} = \begin{pmatrix} \boldsymbol{X}_{01}'\boldsymbol{X}_{01} + \boldsymbol{X}_{02}'\boldsymbol{X}_{02} & \boldsymbol{X}_{01}'\boldsymbol{X}_1 & \boldsymbol{X}_{02}'\boldsymbol{X}_2 \\ \boldsymbol{X}_1'\boldsymbol{X}_{01} & \boldsymbol{X}_1'\boldsymbol{X}_1 & 0 \\ \boldsymbol{X}_2'\boldsymbol{X}_{02} & 0 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{pmatrix}.$$

With the formula of inverse block diagonal matrix

$$\begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & -(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \\ -\boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} & \boldsymbol{D}^{-1} + \boldsymbol{D}^{-1}\boldsymbol{C}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1}\boldsymbol{B}\boldsymbol{D}^{-1} \end{pmatrix},$$

where all inverses exist and $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ are suitable matrices. When $i = 1$,

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{X}_{01}'\boldsymbol{X}_{01} + \boldsymbol{X}_{02}'\boldsymbol{X}_{02} & \boldsymbol{X}_{01}'\boldsymbol{X}_1 \\ \boldsymbol{X}_1'\boldsymbol{X}_{01} & \boldsymbol{X}_1'\boldsymbol{X}_1 \end{pmatrix}, \boldsymbol{B} = \begin{pmatrix} \boldsymbol{X}_{02}'\boldsymbol{X}_2 \\ 0 \end{pmatrix},$$

$$\boldsymbol{C} = \left(\boldsymbol{X}_2'\boldsymbol{X}_{02}, \boldsymbol{0}\right), \boldsymbol{D} = \left(\boldsymbol{X}_2'\boldsymbol{X}_2\right)^{-1}.$$

Then the posterior covariance matrix for $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_1)$ is

$$\begin{aligned}(\boldsymbol{A}-\boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C})^{-1} &= \begin{pmatrix} \boldsymbol{X}_{01}'\boldsymbol{X}_{01} + \boldsymbol{X}_{02}'(\boldsymbol{I}_{n_2} - \boldsymbol{P}_2)\boldsymbol{X}_{02} & \boldsymbol{X}_{01}'\boldsymbol{X}_1 \\ \boldsymbol{X}_1'\boldsymbol{X}_{01} & \boldsymbol{X}_1'\boldsymbol{X}_1 \end{pmatrix}^{-1} \\ &\leq \begin{pmatrix} \boldsymbol{X}_{01}'\boldsymbol{X}_{01} & \boldsymbol{X}_{01}'\boldsymbol{X}_1 \\ \boldsymbol{X}_1'\boldsymbol{X}_{01} & \boldsymbol{X}_1'\boldsymbol{X}_1 \end{pmatrix}^{-1}.\end{aligned}$$

When $i = 2$, we only need to multiply $\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}$ by

$$\boldsymbol{M} = \begin{pmatrix} \boldsymbol{I}_{p_0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{p_2} \\ 0 & \boldsymbol{I}_{p_1} & 0 \end{pmatrix}$$

on the left and $\boldsymbol{M}'$ on the right. A similar procedure can be performed for $(\boldsymbol{\beta}_0, \boldsymbol{\beta}_2)$.

A.1.3 Proof of Fact 2.3

We show the derivations for marginal distributions in (2.9). For brevity, let's consider the linear regression model $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$ and σ^2 is unknown. Here, the known design matrix $\boldsymbol{X}$ is $n \times p$, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is unknown regression coefficients. Conventional g-prior for the regression coefficient $\boldsymbol{\beta}$ and Jeffrey prior for σ^2 are specified as

$$\begin{aligned} \boldsymbol{\beta}|\sigma^2, g &\sim N(\boldsymbol{0}, \sigma^2 g(\boldsymbol{X}'\boldsymbol{X})^{-1}); \\ f(\sigma^2) &\propto \frac{1}{\sigma^2} \end{aligned}$$

Proof. The posterior distribution for $(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y})$ is

$$\begin{aligned} f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \propto &(\sigma^2)^{-\frac{n}{2}} exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta})\} \\ &|\sigma^2 g(\boldsymbol{X}'\boldsymbol{X})^{-1}|^{-\frac{1}{2}} exp\{-\frac{1}{2}\boldsymbol{\beta}'(\sigma^2 g(\boldsymbol{X}'\boldsymbol{X})^{-1})^{-1}\boldsymbol{\beta}\}\frac{1}{\sigma^2}. \end{aligned}$$

The marginal distribution for $(\boldsymbol{\beta}|\boldsymbol{y})$ is

$$\begin{aligned} f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \int_{\sigma^2} f(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) d\sigma^2 \\ &\propto \int_{\sigma^2} (\sigma^2)^{-\frac{n+p+2}{2}} exp\{-\frac{1}{2\sigma^2}[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + \frac{(\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}\boldsymbol{\beta})}{g}]\} d\sigma^2 \\ &\propto \{\frac{1}{2}[(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta}) + \frac{(\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{X}\boldsymbol{\beta})}{g}]\}^{-\frac{n+p}{2}}. \end{aligned}$$

Set $\boldsymbol{\mu} = (g^{-1}+1)^{-1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \boldsymbol{\Lambda}^{-1} = n(1+g^{-1})(\boldsymbol{X}'\boldsymbol{X})$ and $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$,

$$\begin{aligned} f(\boldsymbol{\beta}|\boldsymbol{y}) &\propto \{\frac{(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})}{n} + \boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}\}^{-\frac{n+p}{2}} \\ &\propto \{\frac{(\boldsymbol{\beta}-\boldsymbol{\mu})'\boldsymbol{\Lambda}^{-1}(\boldsymbol{\beta}-\boldsymbol{\mu})}{n\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}} + 1\}^{-\frac{n+p}{2}}. \end{aligned}$$

Hence, the marginal distribution for $\boldsymbol{\beta}|\boldsymbol{y}$ is multivatiate t distribution with

$$t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \text{ where } \boldsymbol{\Sigma}^{-1} = \frac{\boldsymbol{\Lambda}^{-1}}{\boldsymbol{y}'(\boldsymbol{I}-\boldsymbol{P})\boldsymbol{y}}.$$

□

A.1.4 Proof of Theorem 2.2

Here, we use the same notations and similar techniques in Proof A.1.1. Recall that $\boldsymbol{A}_i = \boldsymbol{X}_{0i}'(\boldsymbol{I}_{n_i} - \boldsymbol{P}_i)\boldsymbol{X}_{0i}$, and $\boldsymbol{B}_j = (\boldsymbol{I}_{n_j} - \boldsymbol{P}_j)\boldsymbol{X}_{0j}$, where $i, j = 1, 2, i \neq j$, then

$$\begin{aligned} & VAR(\boldsymbol{\beta}_0|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|g_i, \boldsymbol{y}_i, M_i) \\ =& a(\boldsymbol{A}_i + \boldsymbol{B}_j'\boldsymbol{B}_j)^{-1} - a_i\boldsymbol{A}_i^{-1} \\ =& a\{\boldsymbol{A}_i^{-1} - \boldsymbol{A}_i^{-1}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\} - a_i\boldsymbol{A}_i^{-1} \end{aligned}$$

$$=(a-a_i)\boldsymbol{A}_i^{-1} - a\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}.$$

Hence, $VAR(\boldsymbol{\beta}_0|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_0|, g_i, \boldsymbol{y}_i, M_i) \leq 0$ is equivalent to

$$\left(1 - \frac{a_i}{a}\right)\boldsymbol{I}_{p_0} \leq \boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}. \tag{A.4}$$

As it has been proved that in Proof A.1.1, $\lambda_{min}(\boldsymbol{A}_i^{-\frac{1}{2}}\boldsymbol{B}_j'(\boldsymbol{I}_{n_j}+\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j')^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-\frac{1}{2}}) \in (0, 1)$.

Next, we show the results for specific regression coefficients $\boldsymbol{\beta}_i$.

$$\begin{aligned} VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) = &(a-a_i)\sigma^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} + (a-a_i)\sigma^2\boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i' \\ &- a\sigma^2\boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'[\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j']^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i'. \end{aligned}$$

As $\boldsymbol{Q}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{X}_{0i},\ \boldsymbol{M}_i = [(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1} + \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i']$, and $\boldsymbol{N} = \boldsymbol{Q}_i\boldsymbol{A}_i^{-1}\boldsymbol{B}_j'[\boldsymbol{I}_{n_j} + \boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{B}_j']^{-1}\boldsymbol{B}_j\boldsymbol{A}_i^{-1}\boldsymbol{Q}_i'$, then

$$VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) = (a-a_i)\sigma^2\boldsymbol{M}_i - a\sigma^2\boldsymbol{N}. \tag{A.5}$$

Notice that equation (A.5) can be simplified according to the rank of $\boldsymbol{Q}_i$.

When $\boldsymbol{X}_i'\boldsymbol{X}_{0i} = \boldsymbol{0}_{n_i\times p_0}$ or $p_i > p_0$, $\boldsymbol{Q}_i$ is either equal to $\boldsymbol{0}_{p_i\times p_0}$ or of full column rank and $\boldsymbol{N}$ is either $\boldsymbol{0}_{p_i\times p_i}$ or not of full rank. Then, $VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) \leq 0$ is equivalent to $a_i \geq a$.

When $\boldsymbol{X}_i'\boldsymbol{X}_{0i} \neq \boldsymbol{0}_{n_i\times p_0}$ and $p_i \leq p_0$, $\boldsymbol{Q}_i$ is of full row rank and $\boldsymbol{N}$ is of full rank.

$$VAR(\boldsymbol{\beta}_i|g, \boldsymbol{y}, M_c) - VAR(\boldsymbol{\beta}_i|g_i, \boldsymbol{y}_i, M_i) \leq 0$$

is equivalent to

$$\left(1 - \frac{a_i}{a}\right) \boldsymbol{I}_{p_i} \leq \boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}},$$

which is also equivalent to

$$1 - \frac{a_i}{a} \leq \lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}).$$

As in Proof A.1.1, $\lambda_{min}(\boldsymbol{M}_i^{-\frac{1}{2}} \boldsymbol{N} \boldsymbol{M}_i^{-\frac{1}{2}}) \in (0, 1)$ holds.

A.1.5 Proof of Theorem 2.3

We need the following lemmas.

Lemma A.1. *If the ratio of two densities $f_1(x)$ and $f_2(x)$ is increasing in x, then $E_{f_1}(x) \geq E_{f_2}(x)$.*

Proof. See Lemma 6.1 in Shao (2003). □

Lemma A.2.

$$\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt = (1-z)^{c-a-b} \int_0^1 t^{c-b-1}(1-t)^{b-1}(1-tz)^{a-c}dt,$$

where $c > b > 0, |z| < 1$.

Proof. When $c > b > 0$, we have the following equation from Bailey (1935)

$${}_2F_1(a, b; c; z)Beta(b, c-b) = \int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}dt,$$

where ${}_2F_1(a,b;c;z)$ is the Gaussian hypergeometric function with

$$ {}_2F_1(a,b;c;z) = \frac{1}{Beta(b,c-b)} \int_0^1 x^{b-1}(1-x)^{c-b-1}(1-xz)^{-a}dx, c > b > 0. $$

${}_2F_1(a,b;c;z)$ is convergent for $|z| < 1$ with $c > b > 0$ and for $z = \pm 1$ only if $c > a+b$ and $b > 0$. By Euler's transformation ${}_2F_1(a,b;c;z) = (1-z)^{c-a-b}{}_2F_1(c-a,c-b;c;z)$,

$$ \begin{aligned} {}_2F_1(a,b;c;z)Beta(b,c-b) &= (1-z)^{c-a-b}{}_2F_1(c-a,c-b;c;z)Beta(b,c-b) \\ &= (1-z)^{c-a-b}\int_0^1 t^{c-b-1}(1-t)^{b-1}(1-tz)^{a-c}dt. \end{aligned} $$

□

Recall that the marginal posterior of g has the following form

$$ \pi(g|\boldsymbol{y}) \propto (1+g)^{\frac{n-p}{2}} g^{-\frac{3}{2}} \exp(-\frac{n}{2g}) \left(1 - \frac{g\tilde{R}^2}{1+g}\right)^{-\frac{n}{2}}. $$

Then, we set $t = g/(1+g)$ to offer an easier analysis and then the density becomes

$$ \pi(t|\boldsymbol{y}) \propto (1-t)^{\frac{p-1}{2}} t^{-\frac{3}{2}} \exp(-\frac{n}{2t})(1-t\tilde{R}^2)^{-\frac{n}{2}}. $$

Since an explicit evaluation of $\pi(t|\boldsymbol{y})$ is not feasible, alternatively, we consider the density $h(t|\boldsymbol{y}) \propto (1-t)^{p/2}(1-t\tilde{R}^2)^{-n/2}$. Notice that the ratio $\pi(t|\boldsymbol{y})/h(t|\boldsymbol{y}) = t^{-3/2}(1-t)^{-1/2}\exp(-n/(2t))$ is the increasing function with respect to t. Hence, by Lemma A.1, we have $E_\pi(t|\boldsymbol{y}) \geq E_h(t|\boldsymbol{y})$. It is obvious that $E_\pi(t|\boldsymbol{y}) \leq 1$ and we only need to evaluate $E_h(t|\boldsymbol{y})$. Specifically,

$$E_h(t|\boldsymbol{y}) = \frac{\int_0^1 t(1-t)^{\frac{p}{2}}(1-t\tilde{R}^2)^{-\frac{n}{2}}dt}{\int_0^1 (1-t)^{\frac{p}{2}}(1-t\tilde{R}^2)^{-\frac{n}{2}}dt} = \frac{2}{4+p}\frac{{}_2F_1(\frac{n}{2},2;\frac{p}{2}+3;\tilde{R}^2)}{{}_2F_1(\frac{n}{2},1;\frac{p}{2}+2;\tilde{R}^2)}. \tag{A.6}$$

Then, by Lemma A.2, (A.6) is represented as

$$\frac{(1-\tilde{R}^2)^{\frac{p-n}{2}+1}\int_0^1 t^{\frac{p}{2}}(1-t)(1-t\tilde{R}^2)^{\frac{n-p}{2}-3}dt}{(1-\tilde{R}^2)^{\frac{p-n}{2}+1}\int_0^1 t^{\frac{p}{2}}(1-t\tilde{R}^2)^{\frac{n-p}{2}-2}dt} = \frac{\int_0^1 t^{\frac{p}{2}}(1-t)(1-t\tilde{R}^2)^{\frac{n-p}{2}-3}dt}{\int_0^1 t^{\frac{p}{2}}(1-t\tilde{R}^2)^{\frac{n-p}{2}-2}dt}. \tag{A.7}$$

As $\tilde{R}^2 \to 1$, (A.7) $\to 1$. For the posterior variance, we only need to prove $E(g^2/(1+g)^2|\boldsymbol{y}) \to 1$, which is equivalent to show $E(t^2|\boldsymbol{y}) \to 1$. Similarly, by Lemma A.2, we have

$$E(t^2|\boldsymbol{y}) = \frac{\int_0^1 t^2(1-t)^{\frac{p}{2}}(1-t\tilde{R}^2)^{-\frac{n}{2}}dt}{\int_0^1 (1-t)^{\frac{p}{2}}(1-t\tilde{R}^2)^{-\frac{n}{2}}dt} = \frac{\int_0^1 t^{\frac{p}{2}}(1-t)^2(1-t\tilde{R}^2)^{\frac{n-p}{2}-4}dt}{\int_0^1 t^{\frac{p}{2}}(1-t\tilde{R}^2)^{\frac{n-p}{2}-2}dt}, \tag{A.8}$$

which approaches to 1 as $\tilde{R}^2 \to 1$.

A.2 Proofs of Theorems, Lemmas, Remarks and Facts in Chapter 3

A.2.1 Proof of Theorem 3.1

Define $\boldsymbol{V}_1 = \boldsymbol{X}_{01}'[(1+g_0^{-1})\boldsymbol{I}_{n_1} - (1+g_1^{-1})^{-1}\boldsymbol{P}_1]\boldsymbol{X}_{01}$ and $\boldsymbol{V}_2 = \boldsymbol{X}_{02}'[(1+g_0^{-1})\boldsymbol{I}_{n_2} - (1+g_2^{-1})^{-1}\boldsymbol{P}_2]\boldsymbol{X}_{02}$, then

$$VAR(\boldsymbol{\beta}_0|\sigma^2, g_0, g_1, \boldsymbol{y}_1, M_1) = \sigma^2\boldsymbol{V}_1^{-1},$$

$$VAR(\boldsymbol{\beta}_0|\sigma^2, g_0, g_1, g_2, \boldsymbol{y}, M_c) = \sigma^2(\boldsymbol{V}_1 + \boldsymbol{V}_2)^{-1}.$$

For the comparison for common regression coefficient $\boldsymbol{\beta}_0$,

$$\begin{aligned}
&VAR(\boldsymbol{\beta}_0|\sigma^2, g_0, g_1, \boldsymbol{y}_1, M_1) - VAR(\boldsymbol{\beta}_0|\sigma^2, g_0, g_1, g_2, \boldsymbol{y}, M_c)\\
=&\sigma^2[\boldsymbol{V}_1^{-1} - (\boldsymbol{V}_1 + \boldsymbol{V}_2)^{-1}]\\
=&\sigma^2\{\boldsymbol{V}_1^{-1} - [\boldsymbol{V}_1^{-1} - (\boldsymbol{V}_1 + \boldsymbol{V}_1\boldsymbol{V}_2^{-1}\boldsymbol{V}_1)^{-1}]\}\\
=&\sigma^2(\boldsymbol{V}_1 + \boldsymbol{V}_1\boldsymbol{V}_2^{-1}\boldsymbol{V}_1)^{-1}.
\end{aligned} \tag{A.9}$$

Notice that $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ have the same structure. Recall that $\boldsymbol{P}_1 = \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$ and we assume that $\boldsymbol{X}_1$ is of full column rank. $\boldsymbol{P}_1$ is idempotent and of rank p_1. According to the eigen-decomposition theorem, there exists an orthogonal matrix $\boldsymbol{T}$, where the columns of $\boldsymbol{T}$ is composed of eigenvectors of $\boldsymbol{P}_1$, and $\boldsymbol{\Lambda} = diag(1, 1, \cdots, 0)$, where $\boldsymbol{\Lambda}$ if a diagonal matrix of rank p_1 and dimension n_1, such that $\boldsymbol{P}_1 = \boldsymbol{T}'\boldsymbol{\Lambda}\boldsymbol{T}$. Then,

$$\begin{aligned}
\boldsymbol{V}_1 &= \boldsymbol{T}\{\boldsymbol{X}_{01}'[(1 + g_0^{-1})\boldsymbol{I}_{n_1} - (1 + g_1^{-1})^{-1}\boldsymbol{P}_1]\boldsymbol{X}_{01}\}\boldsymbol{T}'\\
&= \boldsymbol{X}_{01}'[(1 + g_0^{-1})\boldsymbol{I}_{n_1} - (1 + g_1^{-1})^{-1}\boldsymbol{\Lambda}]\boldsymbol{X}_{01}.
\end{aligned}$$

The elements of $[(1 + g_0^{-1})\boldsymbol{I}_{n_1} - (1 + g_1^{-1})^{-1}\boldsymbol{\Lambda}]$ are either $g_0^{-1} + (g_1 + 1)^{-1} > 0$ or $(1 + g_{01})^{-1} > 0$, which indicates $\boldsymbol{V}_1$ is positive definite. Similarly, $\boldsymbol{V}_2$ is positive definite. Hence, (A.9) is positive definite.

For specific regression coefficients $\boldsymbol{\beta}_1$, recall that $\boldsymbol{Q}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{X}_{0i}$, then

$$VAR(\boldsymbol{\beta}_1|\sigma^2, g_0, g_1, \boldsymbol{y}_1, M_1) - VAR(\boldsymbol{\beta}_1|\sigma^2, g_0, g_1, g_2, \boldsymbol{y}, M_c)$$

$$
\begin{aligned}
=&\sigma^2\{[(g_1^{-1}+1)^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}+(g_1^{-1}+1)^{-2}\boldsymbol{Q}_1\boldsymbol{V}_1^{-1}\boldsymbol{Q}_1'] - [(g_1^{-1}+1)^{-1}(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1} \\
&+(g_1^{-1}+1)^{-2}\boldsymbol{Q}_1(\boldsymbol{V}_1+\boldsymbol{V}_2)^{-1}\boldsymbol{Q}_1]'\} \\
=&\sigma^2\{(g_1^{-1}+1)^{-2}\boldsymbol{Q}_1[\boldsymbol{V}_1^{-1}-(\boldsymbol{V}_1+\boldsymbol{V}_2)^{-1}]\boldsymbol{Q}_1'\}. \qquad \text{(A.10)}
\end{aligned}
$$

As $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ are proved to be positive definite, (A.10) is positive semi-definite, where $\boldsymbol{Q}_i$ has dimension of $p_i \times p_0$. The same holds for $\boldsymbol{\beta}_2$.

A.2.2 Proof of Lemma 3.1

Proof. Without loss of generality, we use $1/R_{01}^{2(k)}$ in M_1 as an example. With the defined sequence $\{L^{(k)}\}_{k=1}^{\infty}$ for M_1, we have:

$$
\begin{aligned}
\frac{1}{R_{01}^{2(k)}} &= \frac{\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}'\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}+\boldsymbol{\beta}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{\epsilon}_1'\boldsymbol{\epsilon}_1}{\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}+\boldsymbol{\epsilon}_1'\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon}_1} \\
&= 1+\frac{\boldsymbol{\beta}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{\epsilon}_1'(\boldsymbol{I}_{n_1}-\boldsymbol{P}_{X_{01}})\boldsymbol{\epsilon}_1}{\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}'\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}+\boldsymbol{\epsilon}_1'\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon}_1}.
\end{aligned}
$$

Since $\boldsymbol{\epsilon}_1'(\boldsymbol{I}_{n_1}-\boldsymbol{P}_{X_{01}})\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_1'\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon}_1$ are independent, denote $\{\boldsymbol{X}_{01},\boldsymbol{\beta}_0^{(k)},\boldsymbol{X}_1,\boldsymbol{\beta}_1\}$ as $\cdot$ and we have:

$$
\begin{aligned}
&E(1/R_{01}^{2(k)}|\cdot) \\
=&1+E\big[\boldsymbol{\beta}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}_1+\boldsymbol{\epsilon}_1'(\boldsymbol{I}_{n_1}-\boldsymbol{P}_{X_{01}})\boldsymbol{\epsilon}_1|\cdot\big]E\big[(\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}'\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}+\boldsymbol{\epsilon}_1'\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon}_1)^{-1}|\cdot\big] \\
=&1+(\sigma^{-2}\boldsymbol{\beta}_1'\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{\beta}_1+n_1-p_0)E\big[(\sigma^{-2}\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}'\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}+\sigma^{-2}\boldsymbol{\epsilon}_1'\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon}_1)^{-1}|\big]
\end{aligned}
$$

Let $a = \sigma^{-2}\boldsymbol{\beta}_0^{(k)\prime}\boldsymbol{X}_{01}^\prime\boldsymbol{X}_{01}\boldsymbol{\beta}_0^{(k)}$ and we may find that $Q = \sigma^{-2}\boldsymbol{\epsilon}^\prime\boldsymbol{P}_{X_{01}}\boldsymbol{\epsilon} \sim \chi^2_{p_0}$. Then, we only need to calculate $E[(a+Q)^{-1}]$. We have

$$
\begin{aligned}
E[(a+Q)^{-1}] &= \int_0^{+\infty} \frac{1}{a+q}\frac{(\frac{1}{2})^{\frac{p_0}{2}}}{\Gamma(\frac{p_0}{2})} q^{\frac{p_0}{2}-1}\exp(-\frac{q}{2})dq \\
&= \int_0^{+\infty}\int_0^{+\infty} \exp(-(a+q)t)\frac{(\frac{1}{2})^{\frac{p_0}{2}}}{\Gamma(\frac{p_0}{2})} q^{\frac{p_0}{2}-1}\exp(-\frac{q}{2})dtdq \\
&= \int_0^{+\infty}\int_0^{+\infty} \exp(-(a+q)t)\frac{(\frac{1}{2})^{\frac{p_0}{2}}}{\Gamma(\frac{p_0}{2})} q^{\frac{p_0}{2}-1}\exp(-\frac{q}{2})dqdt \\
&= (\frac{1}{2})^{\frac{p_0}{2}}\int_0^{+\infty} \exp(-at)(\frac{1}{2}+t)^{-\frac{p_0}{2}}dt \\
&= (\frac{1}{2})^{\frac{p_0}{2}}\exp(\frac{a}{2})\int_{\frac{1}{2}}^{+\infty} \exp(-at)t^{-\frac{p_0}{2}}dt \\
&= (\frac{1}{2})^{\frac{p_0}{2}}\exp(\frac{a}{2})a^{\frac{p_0}{2}-1}\Gamma(1-\frac{p_0}{2},\frac{a}{2}),
\end{aligned}
$$

where $\Gamma(m,x) = \int_x^{+\infty} t^{m-1}\exp(-t)dt$ is the upper incomplete Gamma function. If $k \to \infty, ||\boldsymbol{\beta}_0^{(k)}||^2 \to \infty$ and therefore $a \to \infty$, by the L'Hospital rule, we have

$$
\begin{aligned}
\lim_{a\to+\infty}\frac{\int_{\frac{a}{2}}^{+\infty}\exp(-s)s^{-\frac{p_0}{2}}ds}{\exp(-\frac{a}{2})a^{1-\frac{p_0}{2}}} &= -\lim_{a\to+\infty}\frac{\frac{1}{2}\exp(-\frac{a}{2})(\frac{a}{2})^{-\frac{p_0}{2}}}{-\frac{1}{2}\exp(-\frac{a}{2})a^{1-\frac{p_0}{2}} + (1-\frac{p_0}{2})\exp(-\frac{a}{2})a^{-\frac{p_0}{2}}} \\
&= \lim_{a\to+\infty}\frac{(\frac{1}{2})^{1-\frac{p_0}{2}}}{\frac{1}{2}a-(1-\frac{p_0}{2})} \\
&= 0.
\end{aligned}
$$

□

A.2.3 Proof of Lemma 3.2

To prove this argument, we begin with the following lemmas.

Recall that directly dealing with the term $(1 - t_0 R_{0i}^2 - t_i R_i^2)^{-n_i/2}$ in (3.11) or $(1 - t_{0c} R_0^2 - t_{1c} R_1^2 - t_{2c} R_2^2)^{-n/2}$ in (3.13) might be difficult. As an alternative, inspired by Som et al. (2015), we employ the joint density of (t_0, t_i) with the following form

$$h(t_0, t_i|\boldsymbol{y}_i) \propto (1 - t_0)^{\frac{p_0}{2}} (1 - t_i)^{\frac{p_i}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} (1 - t_i R_i^2)^{-\frac{n_i}{2}}. \tag{A.11}$$

Notice that $h(t_0, t_i|\boldsymbol{y}_i)$ is equivalent to applying the independent scaled Pareto priors with the parameterization (g_0, g_i).

Remark A.1. *With the joint density of $h(t_0, t_i|\boldsymbol{y}_i)$, we have*

$$\begin{aligned} E_h(t_0|\boldsymbol{y}_i) =& \frac{\int_0^1 \int_0^1 t_0 (1 - t_0)^{\frac{p_0}{2}} (1 - t_i)^{\frac{p_i}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} (1 - t_i R_i)^{-\frac{n_i}{2}} dt_i dt_0}{\int_0^1 \int_0^1 (1 - t_0)^{\frac{p_0}{2}} (1 - t_i)^{\frac{p_i}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} (1 - t_i R_i)^{-\frac{n_i}{2}} dt_i dt_0} \\ =& \frac{\int_0^1 t_0 (1 - t_0)^{\frac{p_0}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} dt_0}{\int_0^1 (1 - t_0)^{\frac{p_0}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} dt_0} = \frac{2}{4 + p_0} \frac{{}_2F_1(\frac{n_i}{2}, 2; \frac{p_0}{2} + 3; R_{0i}^2)}{{}_2F_1(\frac{n_i}{2}, 1; \frac{p_0}{2} + 2; R_{0i}^2)}. \end{aligned} \tag{A.12}$$

Similarly,

$$\begin{aligned} E_h(t_i|\boldsymbol{y}_i) =& \frac{\int_0^1 \int_0^1 t_i (1 - t_0)^{\frac{p_0}{2}} (1 - t_i)^{\frac{p_i}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} (1 - t_i R_i^2)^{-\frac{n_i}{2}} dt_0 dt_i}{\int_0^1 \int_0^1 (1 - t_0)^{\frac{p_0}{2}} (1 - t_i)^{\frac{p_i}{2}} (1 - t_0 R_{0i}^2)^{-\frac{n_i}{2}} (1 - t_i R_i^2)^{-\frac{n_i}{2}} dt_0 dt_i} \\ =& \frac{\int_0^1 t_i (1 - t_i)^{\frac{p_i}{2}} (1 - t_i R_i^2)^{-\frac{n_i}{2}} dt_i}{\int_0^1 (1 - t_i)^{\frac{p_i}{2}} (1 - t_i R_i^2)^{-\frac{n_i}{2}} dt_i} = \frac{2}{4 + p_i} \frac{{}_2F_1(\frac{n_i}{2}, 2; \frac{p_i}{2} + 3; R_i^2)}{{}_2F_1(\frac{n_i}{2}, 1; \frac{p_i}{2} + 2; R_i^2)}. \end{aligned} \tag{A.13}$$

For M_i, we denote $E_h(t_0|\boldsymbol{y}_i) = H(n_i, p_0, R_{0i}^2)$ and $E_h(t_i|\boldsymbol{y}_i) = H(n_i, p_i, R_i^2)$. Similarly, for M_c, denote $E_h(t_0|\boldsymbol{y}) = H(n, p_0, R_0^2)$ and $E_h(t_i|\boldsymbol{y}) = H(n, p_i, R_i^2)$. For simplicity, we only state results for M_i in Lemma A.3 and results can be extended to M_c without efforts. For an easy demonstration, and for $i = 1, 2$, we further denote $F(n_i, p_0, R_{0i}^2) = E(t_0|\boldsymbol{y}_i, M_i)$, $F(n_i, p_i, R_i^2) = E(t_i|\boldsymbol{y}_i, M_i)$, $F(n, p_0, R_0^2) = E(t_{0c}|\boldsymbol{y}, M_c)$, and $F(n, p_i, R_i^2) = E(t_{ic}|\boldsymbol{y}, M_c)$

Lemma A.3. *Notice that $F(n_i, p_0, R_{0i}^2)$ is the target posterior expectation with density $f(t_0|\boldsymbol{y}_i)$, $H(n_i, p_0, R_{0i}^2)$ is the alternative posterior expectation with density $h(t_0|\boldsymbol{y}_i)$, $F(n_i, p_i, R_{0i}^2)$ is the target posterior expectation with density $f(t_i|\boldsymbol{y}_i)$, and $H(n_i, p_i, R_{0i}^2)$ is the alternative posterior expectation with density $h(t_i|\boldsymbol{y}_i)$. By Lemma A.1, we have*

$$F(n_i, p_0, R_{0i}^2) \geq H(n_i, p_0, R_{0i}^2) \text{ and } F(n_i, p_i, R_i^2) \geq H(n_i, p_i, R_i^2). \tag{A.14}$$

Proof. With the alternative density in (A.11), the ratio of the target and alternative marginal densities is as below

$$\begin{aligned}\frac{f(t_0|\boldsymbol{y}_i)}{h(t_0|\boldsymbol{y}_i)} &\propto \frac{\int_0^1 f(t_0, t_i|\boldsymbol{y}_i)dt_i}{\int_0^1 h(t_0, t_i|\boldsymbol{y}_i)dt_i} \\ &\propto t_0^{-\frac{3}{2}}(1-t_0)^{-\frac{1}{2}}\exp(-\frac{n_i}{2t_0})\int_0^1 t_i^{-\frac{3}{2}}(1-t_i)^{-\frac{1}{2}}\exp(-\frac{n_i}{2t_i})(1-\frac{t_iR_i^2}{1-t_0R_{0i}^2})^{-\frac{n_i}{2}}dt_i.\end{aligned}$$

It is easy to show that both

$$t_0^{-\frac{3}{2}}(1-t_0)^{-\frac{1}{2}}\exp(-\frac{n_i}{2t_0}) \text{ and } \int_0^1 \exp(-\frac{n_i}{2t_i})(1-\frac{t_iR_i^2}{1-t_0R_{0i}^2})^{-n_i/2}dt_i$$

are increasing with respect to t_0 and therefore $f(t_0|\boldsymbol{y}_i)/h(t_0|\boldsymbol{y}_i)$ is increasing in t_0. By Lemma A.1, we can conclude that

$$F(n_i, p_0, R_{0i}^2) \geq H(n_i, p_0, R_{0i}^2).$$

The rest of inequalities can be proved in a similar way.

□

Lemma A.4. *Under the sequence $\{L_i^{(k)}\}$ defined in (3.17), if $n_i > p_0+2$ and $R_{0i}^{2(k)} \rightarrow$*

1, $H(n_i, p_0, R_{0i}^{2(k)}) \to 1$ and therefore $F(n_i, p_0, R_{0i}^{2(k)}) \to 1$.

Proof. Recall that

$$H(n_i, p_0, R_{0i}^{2(k)}) = \frac{\int_0^1 t_0(1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^2)^{-\frac{n_i}{2}}dt_0}{\int_0^1 (1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^2)^{-\frac{n_i}{2}}dt_0}. \tag{A.15}$$

By Lemma A.2, (A.15) is represented as

$$\begin{aligned} H(n_i, p_0, R_{0i}^{2(k)}) =& \frac{(1-R_{0i}^{2(k)})^{\frac{p_0-n_i}{2}+1}\int_0^1 t_0^{\frac{p_0}{2}}(1-t_0)(1-t_0R_{0i}^{2(k)})^{\frac{n_i-p_0}{2}-3}dt_0}{(1-R_{0i}^{2(k)})^{\frac{p_0-n_i}{2}+1}\int_0^1 t_0^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{\frac{n_i-p_0}{2}-2}dt_0} \\ =& \frac{\int_0^1 t_0^{\frac{p_0}{2}}(1-t_0)(1-t_0R_{0i}^{2(k)})^{\frac{n_i-p_0}{2}-3}dt_0}{\int_0^1 t_0^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{\frac{n_i-p_0}{2}-2}dt_0}. \end{aligned} \tag{A.16}$$

As $R_{0i}^{2(k)} \to 1$, (A.16) $\to 1$, which is equiavalent to $H(n_i, p_0, R_{0i}^{2(k)}) \to 1$. Together with Lemma (A.3), we can conclude that $F(n_i, p_0, R_{0i}^{2(k)}) \to 1$ as $R_{0i}^{2(k)} \to 1$. □

Lemma A.5. *$H(n_i, p_0, R_{0i}^{2(k)})$ is non-decreasing with respect to $R_{0i}^{2(k)}$.*

Proof. The derivative of $H(n_i, p_0, R_{0i}^{2(k)})$ with respect to $R_{0i}^{2(k)}$ is

$$\begin{aligned} \frac{\partial H(n_i, p_0, R_{0i}^{2(k)})}{\partial R_{0i}^{2(k)}} &= \frac{\frac{n_i}{2}\int_0^1 t_0^2(1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{-\frac{n_i}{2}-1}dt_0}{\int_0^1 (1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{-\frac{n_i}{2}}dt_0} \\ &- \frac{\frac{n_i}{2}\int_0^1 t_0(1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{-\frac{n_i}{2}-1}dt_0 \int_0^1 t_0(1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{-\frac{n_i}{2}}dt_0}{(\int_0^1 (1-t_0)^{\frac{p_0}{2}}(1-t_0R_{0i}^{2(k)})^{-\frac{n_i}{2}}dt_0)^2} \end{aligned} \tag{A.17}$$

If we set $s(t_0) = (1-t_0)^{p_0/2}(1-t_0R_{0i}^{2(k)})^{-n_i/2}$ and $h(t_0) = t_0/(1-t_0R_{0i}^{2(k)})$, then

$$\text{(A.17)} = \frac{\frac{n_i}{2}\int_0^1 t_0 s(t_0)h(t_0)dt_0}{\int_0^1 s(t_0)dt_0} - \frac{\frac{n_i}{2}\int_0^1 s(t_0)h(t_0)dt_0\int_0^1 t_0 s(t_0)dt_0}{(\int_0^1 s(t_0)dt_0)^2}$$

and (A.17) ≥ 0 is equivalent to

$$\int_0^1 t_0 h(t_0)s(t_0)dt_0 \int_0^1 s(t_0)dt_0 \geq \int_0^1 h(t_0)s(t_0)dt_0 \int_0^1 t_0 s(t_0)dt_0. \tag{A.18}$$

Since both $h(t_0)$ and t_0 are increasing with respect to t_0, (A.18) holds by the Chebyshev's algebraic inequality (See Proposition 2.1 in Egozcue et al. (2009)). Therefore, $H(n_i, p_0, R_{0i}^{2(k)})$ is increasing with respect to $R_{0i}^{2(k)}$. Although we prove the monotonity in $R_{0i}^{2(k)}$ under the defined sequence, the conclusion holds for any random variable u satisfying the function $H(n_i, p_0, u)$. □

Next, we prove Lemma 3.2.

Proof. By Lemma A.3, we have $F(n_i, p_0, R_{0i}^{2(k)}) \geq H(n_i, p_0, R_{0i}^{2(k)})$ and therefore we only need to prove $\big(H(n_i, p_0, R_{0i}^{2(k)}) - 1\big)||\boldsymbol{\beta}_0^{(k)}|| \to 0$. We first consider the eigendecomposition as $\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i} = \boldsymbol{S}\boldsymbol{\Lambda}\boldsymbol{S}^{-1}$, where $\boldsymbol{\Lambda}$ is a diagonal matrix comprising of the corresponding eigenvalues with λ_1 and λ_{p_0} being its minimum and maximum elements, respectively. We have

$$\begin{aligned} R_{min}^{2(k)} =& \frac{\lambda_1||\boldsymbol{\beta}_0^{(k)}||^2 + \boldsymbol{\epsilon}_i'\boldsymbol{P}_{X_{0i}}\boldsymbol{\epsilon}_i}{\lambda_1||\boldsymbol{\beta}_0^{(k)}||^2 + \boldsymbol{\beta}_i'\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i} \leq R_{0i}^{2(k)} \\ &\leq \frac{\lambda_{p_0}||\boldsymbol{\beta}_0^{(k)}||^2 + \boldsymbol{\epsilon}_i'\boldsymbol{P}_{X_{0i}}\boldsymbol{\epsilon}_i}{\lambda_{p_0}||\boldsymbol{\beta}_0^{(k)}||^2 + \boldsymbol{\beta}_i'\boldsymbol{X}_i'\boldsymbol{X}_i\boldsymbol{\beta}_i + \boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i} = R_{max}^{2(k)}. \end{aligned}$$

By Lemma A.5, we have

$$\begin{aligned} &\big(H(n_i, p_0, R_{min}^{2(k)}) - 1\big)||\boldsymbol{\beta}_0^{(k)}|| \\ \leq&\big(H(n_i, p_0, R_{0i}^{2(k)}) - 1\big)||\boldsymbol{\beta}_0^{(k)}|| \leq \big(H(n_i, p_0, R_{max}^{2(k)}) - 1\big)||\boldsymbol{\beta}_0^{(k)}||. \end{aligned} \tag{A.19}$$

For $\big(H(n_i, p_0, R^{2(k)}_{min}) - 1\big)||\boldsymbol{\beta}^{(k)}_0||$, by L'Hospital's rule,

$$\lim_{k\to\infty} \big(H(n_i, p_0, R^{2(k)}_{min}) - 1\big)||\boldsymbol{\beta}^{(k)}_0|| = \lim_{k\to\infty} \frac{\frac{\partial(H(n_i,p_0,R^{2(k)}_{min})-1)}{\partial||\boldsymbol{\beta}^{(k)}_0||}}{\frac{\partial||\boldsymbol{\beta}^{(k)}_0||^{-1}}{\partial||\boldsymbol{\beta}^{(k)}_0||}}, \tag{A.20}$$

where

$$\frac{\partial(H(n_i, p_0, R^{2(k)}_{min}) - 1)}{\partial||\boldsymbol{\beta}^{(k)}_0||} = \frac{\partial(H(n_i, p_0, R^{2(k)}_{min}) - 1)}{\partial R^{2(k)}_{min}} \frac{\partial R^{2(k)}_{min}}{\partial||\boldsymbol{\beta}^{(k)}_0||}. \tag{A.21}$$

Then, we calculate the limitation of $\partial(H(n_i, p_0, R^{2(k)}_{min}) - 1)/\partial R^{2(k)}_{min}$, which has the same form as (A.17) with $R^{2(k)}_{0i}$ replaced by $R^{2(k)}_{min}$. By Lemma A.2, we have

$$\frac{\partial(H(n_i, p_0, R^{2(k)}_{min}) - 1)}{\partial R^{2(k)}_{min}} = \frac{n_i(AD - BC)}{2(1 - R^{2(k)}_{min})D^2}, \text{ where} \tag{A.22}$$

$$A = \int_0^1 t_0^{\frac{p_0}{2}} (1 - t_0)^2 (1 - t_0 R^{2(k)}_{min})^{\frac{n_i - p_0}{2} - 3} dt_0, B = \int_0^1 t_0^{\frac{p_0}{2}} (1 - t_0)(1 - t_0 R^{2(k)}_{min})^{\frac{n_i - p_0}{2} - 3} dt_0,$$
$$C = \int_0^1 t_0^{\frac{p_0}{2}} (1 - t_0)(1 - t_0 R^{2(k)}_{min})^{\frac{n_i - p_0}{2} - 2} dt_0, D = \int_0^1 t_0^{\frac{p_0}{2}} (1 - t_0 R^{2(k)}_{min})^{\frac{n_i - p_0}{2} - 2} dt_0.$$

Then,

$$\lim_{R^{2(k)}_{min}\to 1} \frac{AD - BC}{(1 - R^{2(k)}_{min})D^2} = \lim_{R^{2(k)}_{min}\to 1} \frac{1}{D^2} \lim_{R^{2(k)}_{min}\to 1} \frac{AD - BC}{(1 - R^{2(k)}_{min})}, \tag{A.23}$$

where, as $R^{2(k)}_{min} \to 1$,

$$A \to Beta\big(p_0/2 + 1, (n_i - p_0)/2\big), B \to Beta\big(p_0/2 + 1, (n_i - p_0)/2 - 1\big),$$
$$C \to Beta\big(p_0/2 + 1, (n_i - p_0)/2\big), D \to Beta\big(p_0/2 + 1, (n_i - p_0)/2 - 1\big).$$

By the L'Hospital's rule, we have

$$\begin{aligned}\lim_{R^{2(k)}_{min}\to 1}\frac{AD-BC}{(1-R^{2(k)}_{min})} &= -\lim_{R^{2(k)}_{min}\to 1}(A'D+AD'-B'C-BC')\\ &= \frac{\Gamma(\frac{p_0}{2}+1)\Gamma(\frac{p_0}{2}+2)\Gamma(\frac{n_i-p_0}{2}-1)\Gamma(\frac{n_i-p_0}{2}-2)}{\Gamma(\frac{n_i}{2})\Gamma(\frac{n_i}{2}+1)},\end{aligned}$$

where

$$\begin{aligned}A' &\to -[(n_i-p_0)/2-3]Beta\big(p_0/2+2,(n_i-p_0)/2-1\big),\\ B' &\to -[(n_i-p_0)/2-3]Beta\big(p_0/2+2,(n_i-p_0)/2-2\big),\\ C' &\to -[(n_i-p_0)/2-2]Beta\big(p_0/2+2,(n_i-p_0)/2-1\big),\\ D' &\to -[(n_i-p_0)/2-2]Beta\big(p_0/2+2,(n_i-p_0)/2-2\big),\end{aligned}$$

and " $'$ " refers to the derivative in terms of $R^{2(k)}_{min}$. Therefore,

$$(\text{A.22}) \to \frac{p_0+2}{n_i-p_0-4}. \tag{A.24}$$

At last, as $||\boldsymbol{\beta}_0^{(k)}||\to\infty$, we may find that the limitation as follows

$$\begin{aligned}&\lim_{||\boldsymbol{\beta}_0^{(k)}||\to\infty}\frac{\partial R^{2(k)}_{min}/\partial||\boldsymbol{\beta}_0^{(k)}||}{\partial||\boldsymbol{\beta}_0^{(k)}||^{-1}/\partial||\boldsymbol{\beta}_0^{(k)}||}\\ =&\lim_{||\boldsymbol{\beta}_0^{(k)}||\to\infty}-\frac{2\lambda_1\big(\|\boldsymbol{X}_i\boldsymbol{\beta}_i\|^2+\boldsymbol{\epsilon}_i'(\boldsymbol{I}-\boldsymbol{P}_{X_{0i}})\boldsymbol{\epsilon}_i\big)||\boldsymbol{\beta}_0^{(k)}||^3}{\big(\lambda_1||\boldsymbol{\beta}_0^{(k)}||^2+\|\boldsymbol{X}_i\boldsymbol{\beta}_i\|^2+\boldsymbol{\epsilon}_i'\boldsymbol{\epsilon}_i\big)^2}=0.\end{aligned} \tag{A.25}$$

By (A.19), (A.20), (A.24) and (A.25), we may conclude $\big[H(n_i,p_0,R^{2(k)}_{0i})-1\big]||\boldsymbol{\beta}_0^{(k)}||\to 0$ in probability, which yields $\big[F(n_i,p_0,R^{2(k)}_{0i})-1\big]||\boldsymbol{\beta}_0^{(k)}||\to 0$ in probability through the squeeze theorem. Notice that

$$\boldsymbol{\beta}^{B(k)}_{i,0}-\boldsymbol{\beta}_0^{(k)} = F(n_i,p_0,R^{2(k)}_{0i})\hat{\boldsymbol{\beta}}^{L(k)}_{i,0}-\boldsymbol{\beta}_0^{(k)}$$

$$=\big(F(n_i,p_0,R_{0i}^{2(k)})-1\big)\boldsymbol{\beta}_0^{(k)}+F(n_i,p_0,R_{0i}^{2(k)})(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i. \quad \text{(A.26)}$$

For the first term on the right hand of (A.26), since $||(F(n_i,p_0,R_{0i}^{2(k)})-1)\boldsymbol{\beta}_0^{(k)}|| \leq |(F(n_i,p_0,R_{0i}^{2(k)})-1)|||\boldsymbol{\beta}_0^{(k)}||$ and $|(F(n_i,p_0,R_{0i}^{2(k)})-1)|||\boldsymbol{\beta}_0|| \to 0$, $||(F(n_i,p_0,R_{0i}^{2(k)})-1)\boldsymbol{\beta}_0^{(k)}|| \to 0$ holds. For the second term, we have $F(n_i,p_0,R_{0i}^{2(k)})(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i \to (\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i$ in probability by the Slutsky's theorem. Recall that $X^{(k)} \to X$ in probability and $Y^{(k)} \to Y$ in probability imply $(X^{(k)},Y^{(k)}) \to (X,Y)$ in probability. Specifically, $X^{(k)}+Y^{(k)} \to X+Y$ in probability according to the continuous mapping theorem. Therefore, we conclude $\boldsymbol{\beta}_{i,0}^{B(k)}-\boldsymbol{\beta}_0^{(k)} \to (\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i$ in probability. □

A.2.4 Proof of Lemma 3.3

Proof. Recall that

$$G(n;p_i)=\frac{\int_0^1 t_i^{-\frac{1}{2}}(1-t_i)^{\frac{p_i-1}{2}}\exp(-\frac{n}{2t_i})dt_0}{\int_0^1 t_i^{-\frac{3}{2}}(1-t_i)^{\frac{p_i-1}{2}}\exp(-\frac{n}{2t_i})dt_i}.$$

If we define

$$k(x)=\int_0^1 t_i^{-\frac{1}{2}-x}(1-t_i)^{\frac{p_i-1}{2}}\exp(-\frac{n}{2t_i})dt_i,$$

then $G(n;p_i)=k(0)/k(1)$ and the derivative of $G(n;p_i)$ with respect to n is

$$G'(n;p_i)=\frac{-k(1)k(1)/2+k(2)k(0)/2}{k^2(1)}. \quad \text{(A.27)}$$

To prove $G'(n;p_i)\geq 0$, we equivalently show $k(1)k(1)\leq k(0)k(2)$, which can be established by proving $k(x)$ log-convex. Next, we show $k(x)$ is log-convex. Consider the lth derivative of $k(x)$ as

$$k^{(l)}(x) = \int_0^1 (ln\frac{1}{t_i})^l t_0^{-\frac{1}{2}-x}(1-t_i)^{\frac{p_i-1}{2}} \exp(-\frac{n}{2t_i})dt_i, \tag{A.28}$$

and define the inner product as

$$\langle f, g\rangle = \int_0^1 f(t_i)g(t_i)t_i^{-\frac{1}{2}-x}(1-t_i)^{\frac{p_i-1}{2}} \exp(-\frac{n}{2t_i})dt_i, \forall x > 0.$$

It is easy to verify its the linearity, conjugate symmetry and postive definiteness. Suppose $f(x) = ln(1/x)$ and $g(x) = 1$, with the Cauchy's inequality, we have

$$\langle f, g\rangle^2 \leq \langle f, f\rangle\langle g, g\rangle \Leftrightarrow (k^{(1)}(x))^2 \leq k^{(2)}(x)k(x),$$

which implies $(log(k(x)))'' \geq 0$. □

A.2.5 Proof of Theorem 3.2

Proof. We first present results about the bias. For the first term on the right hand of (A.26), by Lemma 3.2, for the defined sequence, there exists a subsequence (m_k) such that $(F(n_i, p_0, R_{0i}^{2(m_k)}) - 1)||\boldsymbol{\beta}_0^{(m_k)}|| \to 0$ almost surely if $n_i - p_0 > 4$. Furthermore, we consider $(1 - F(n_i, p_0, R_{0i}^{2(m_k)}))||\boldsymbol{\beta}_0^{(m_k)}|| \leq (1 - H(n_i, p_0, R_{0i}^{2(m_k)}))||\boldsymbol{\beta}_0^{(m_k)}||$, and

$$sup_{\boldsymbol{\epsilon}_i}\{[1 - H(n_i, p_0, R_{0i}^{2(m_k)})]||\boldsymbol{\beta}_0^{(m_k)}||\} = \{[1 - inf_{\boldsymbol{\epsilon}_i} H(n_i, p_0, R_{0i}^{2(m_k)})]||\boldsymbol{\beta}_0^{(m_k)}||\}. \tag{A.29}$$

With Lemma A.5 indicating that $H(n_i, p_0, R_{0i}^{2(m_k)})$ is increasing with $R_{0i}^{2(m_k)}$ and the infimum of $R_{0i}^{2(m_k)}$ being achieved at $\boldsymbol{\epsilon}_i = \mathbf{0}$ with value

$$R_{inf}^{(m_k)} = \frac{||\boldsymbol{X}_{0i}\boldsymbol{\beta}_0^{(m_k)}||^2}{||\boldsymbol{X}_{0i}\boldsymbol{\beta}_0^{(m_k)}||^2 + ||\boldsymbol{X}_i\boldsymbol{\beta}_i||^2}, \tag{A.30}$$

we may find that

$$\lim_{\boldsymbol{\beta}_0^{(m_k)} \to \infty} \Big\{ [1 - inf_{\boldsymbol{\epsilon}_i} H(n_i, p_0, R_{0i}^{2(m_k)})]||\boldsymbol{\beta}_0^{(m_k)}|| \Big\} \tag{A.31}$$

$$= \lim_{\boldsymbol{\beta}_0^{(m_k)} \to \infty} \Big\{ \big[1 - H(n_i, p_0, R_{inf}^{2(m_k)})\big] ||\boldsymbol{\beta}_0^{(m_k)}|| \Big\} = 0, \tag{A.32}$$

and therefore we conclude that $(H(n_i, p_0, R_{0i}^{2(m_k)}) - 1)||\boldsymbol{\beta}_0^{(m_k)}|| \to 0$ uniformly in $\boldsymbol{\epsilon}_i$ on a convergent set. Hence,

$$\lim_{m_k \to \infty} E\big[1 - H(n_i, p_0, R_{0i}^{(m_k)})\big] ||\boldsymbol{\beta}_0^{(m_k)}|| = E\big[\lim_{m_k \to \infty} (1 - H(n_i, p_0, R_{0i}^{(m_k)}))||\boldsymbol{\beta}_0^{(m_k)}||\big] = 0. \tag{A.33}$$

Each element of $\Big|\big[H(n_i, p_0, R_{0i}^{2(m_k)}) - 1\big]\boldsymbol{\beta}_0^{(m_k)}\Big|$ being smaller than $|H(n_i, p_0, R_{0i}^{2(m_k)}) - 1|\|\boldsymbol{\beta}_0^{(m_k)}\|$ would additionally lead to $\lim_{m_k \to \infty} E\Big\{\big[1 - H(n_i, p_0, R_{0i}^{(m_k)})\big]\boldsymbol{\beta}_0^{(m_k)}\Big\} = \mathbf{0}$ and therefore

$$\lim_{m_k \to \infty} E\Big\{\big[1 - F(n_i, p_0, R_{0i}^{(m_k)})\big]\boldsymbol{\beta}_0^{(m_k)}\Big\} = \mathbf{0}. \tag{A.34}$$

For the second term on the right hand of (A.26), since $F(n_i, p_0, R_{0i}^{2(m_k)}) \to 1$ in probability and it is bounded, by the dominant control theorem (DCT), we may conclude

$$\lim_{m_k \to \infty} E\big[F(n_i, p_0, R_{0i}^{2(m_k)})(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i\big]$$

$$=E\big[\lim_{m_k\to\infty} F(n_i,p_0,R_{0i}^{2(m_k)})(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i\big] = E\big[(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i\big] = \mathbf{0}. \quad (A.35)$$

With (A.26), (A.34) and (A.35), we may conclude the following for the bias of $\boldsymbol{\beta}_0^{(m_k)}$

$$\lim_{m_k\to\infty} E\big(\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)}\big) = \mathbf{0}.$$

Second, we show the asymptotic results for the covariance matrix. Note that the covariance matrix of the Bayesian estimator $\boldsymbol{\beta}_{i,0}^{B(m_k)}$ is

$$E\big[(\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)})(\boldsymbol{\beta}_{i,0}^{B(m_k)} - \boldsymbol{\beta}_0^{(m_k)})'\big] \quad (A.36)$$

$$=E\Big\{\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big]^2\boldsymbol{\beta}_0^{(m_k)}\boldsymbol{\beta}_0^{(m_k)\prime}\Big\} \quad (A.37)$$

$$+E\Big\{\big[F(n_i,p_0,R_{0i}^{2(m_k)})\big]\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big]\boldsymbol{\beta}_0^{(m_k)}\boldsymbol{\epsilon}_i'\boldsymbol{X}_{0i}(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\Big\} \quad (A.38)$$

$$+E\Big\{\big[F(n_i,p_0,R_{0i}^{2(m_k)})\big]\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big](\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i\boldsymbol{\beta}_0^{(m_k)\prime}\Big\} \quad (A.39)$$

$$+E\Big\{F(n_i,p_0,R_{0i}^{2(m_k)})^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\boldsymbol{X}_{0i}'\boldsymbol{\epsilon}_i\boldsymbol{\epsilon}_i'\boldsymbol{X}_{0i}(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}\Big\}. \quad (A.40)$$

By Lemma A.4 and the proof of Lemma 3.2, we may find that (A.37) $\to \mathbf{0}$ since each element of $|[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1]^2\boldsymbol{\beta}_0^{(m_k)}\boldsymbol{\beta}_0^{(m_k)\prime}|$ is smaller than $\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big]^2\|\boldsymbol{\beta}_0^{(m_k)}\|^2$, where $|\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big]| \to 0$ and $\big[F(n_i,p_0,R_{0i}^{2(m_k)}) - 1\big]\|\boldsymbol{\beta}_0^{(m_k)}\|^2$ is bounded. Additionally, since $\big[1 - H(n_i,p_0,R_{0i}^{2(m_k)})\big]\|\boldsymbol{\beta}_0^{(m_k)}\| \to 0$ uniformly, it is uniformly bounded and therefore $\big[1-F(n_i,p_0,R_{0i}^{2(m_k)})\big]\|\boldsymbol{\beta}_0^{(m_k)}\|$ is uniformly bounded. Then by DCT, we may conclude that (A.38) or (A.39) $\to \mathbf{0}$. For (A.40), by DCT, we may have (A.40) $\to \sigma^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}$. Hence, we may conclude that (A.36) $\to \sigma^2(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}$. Every step above for M_i could be applied directly to M_c by replacing the corresponding quantities. □

A.2.6 Proof of Remark 3.7

For simplicity, consider a linear model $\boldsymbol{y}_i = \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i + \boldsymbol{\epsilon}_i$, where $\boldsymbol{\epsilon}_i \sim N_{n_i}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_i})$, $\tilde{\boldsymbol{\beta}}_i = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_i')'$, $\boldsymbol{\beta}_0 \in \mathbb{R}^{p_0}$, $\boldsymbol{\beta}_i \in \mathbb{R}^{p_i}$, $\boldsymbol{C}_i = diag(g_0 n_i(\boldsymbol{X}_{0i}'\boldsymbol{X}_{0i})^{-1}, g_i n_i(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1})$, σ^2 is unknown, and we use the following priors

$$
\begin{aligned}
\pi(\sigma^2) &\propto \frac{1}{\sigma^2},\\
\pi(g_i) &\propto g_i^{-\frac{3}{2}} \exp(-\frac{1}{2g_i}), i = 0, 1 \text{ or } 2,\\
\tilde{\boldsymbol{\beta}}_i \mid g_0, g_i, \sigma^2 &\sim N_{p_I}\left(\boldsymbol{0}, \sigma^2\boldsymbol{C}_i\right).
\end{aligned}
$$

Then, the joint posterior distribution for $(\tilde{\boldsymbol{\beta}}_i, \sigma^2, \boldsymbol{g}_i|\boldsymbol{y}_i)$ is

$$
\begin{aligned}
&f(\tilde{\boldsymbol{\beta}}_i, \sigma^2, \boldsymbol{g}_i|\boldsymbol{y}_i)\\
\propto&(\sigma^2)^{-\frac{n_i}{2}} exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)\}|\sigma^2\boldsymbol{C}_i|^{-\frac{1}{2}} exp\{-\frac{1}{2}\tilde{\boldsymbol{\beta}}_i'(\sigma^2\boldsymbol{C}_i)^{-1}\tilde{\boldsymbol{\beta}}_i\}\\
&\frac{1}{\sigma^2} g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) g_i^{-\frac{3}{2}} \exp(-\frac{1}{2g_i}).
\end{aligned}
$$

Then,

$$
\begin{aligned}
f(\tilde{\boldsymbol{\beta}}_i|\sigma^2, \boldsymbol{g}_i, \boldsymbol{y}_i) &\propto exp\Big\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i) - \frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i\Big\}\\
&\propto exp\Big\{-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\beta}}_i'\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i - 2\boldsymbol{y}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i) - \frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i\Big\}\\
&\propto exp\Big\{-\frac{1}{2\sigma^2}(\tilde{\boldsymbol{\beta}}_i'\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i + \tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i) + \frac{1}{\sigma^2}\boldsymbol{y}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i\Big\}\\
&\propto exp\Big\{-\frac{1}{2\sigma^2}[\tilde{\boldsymbol{\beta}}_i'(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})\tilde{\boldsymbol{\beta}}_i] + \frac{1}{\sigma^2}\boldsymbol{y}_i'\tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i\Big\}\\
&\propto exp\Big\{-\frac{1}{2}(\tilde{\boldsymbol{\beta}}_i - \mu_\beta)'\Sigma_\beta^{-1}(\tilde{\boldsymbol{\beta}}_i - \mu_\beta)\Big\},
\end{aligned}
$$

where $\mu_\beta = (\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i, \Sigma_\beta = \sigma^2(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}$.

$$
\begin{aligned}
& f(\sigma^2, \boldsymbol{g}_i|\boldsymbol{y}_i) \\
\propto & \int_{\tilde{\boldsymbol{\beta}}} (\sigma^2)^{-\frac{n_i}{2}} |\sigma^2\boldsymbol{C}_i|^{-\frac{1}{2}} (\sigma^2)^{-1} exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i) - \frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i\Big\} d\tilde{\boldsymbol{\beta}} \\
\propto & (\sigma^2)^{-\frac{n_i+p}{2}-1} \int_{\tilde{\boldsymbol{\beta}}} exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i) - \frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i\Big\} d\tilde{\boldsymbol{\beta}} \\
\propto & (\sigma^2)^{-\frac{n_i+p}{2}-1} exp\Big\{ - \frac{1}{2\sigma^2}\boldsymbol{y}_i'\boldsymbol{y}_i\Big\} \int_{\tilde{\boldsymbol{\beta}}} exp\Big\{ - \frac{1}{2}\big[(\tilde{\boldsymbol{\beta}}_i - \mu_\beta)'\Sigma_\beta^{-1}(\tilde{\boldsymbol{\beta}}_i - \mu_\beta) - \mu_\beta'\Sigma_\beta^{-1}\mu_\beta\big]\Big\} d\tilde{\boldsymbol{\beta}} \\
\propto & (\sigma^2)^{-\frac{n_i+p}{2}-1} |\Sigma_\beta|^{\frac{1}{2}} exp\Big\{ \frac{1}{2\sigma^2}\boldsymbol{y}_i'\boldsymbol{y}_i + \frac{1}{2}\mu_\beta'\Sigma_\beta^{-1}\mu_\beta\Big\} \\
\propto & (\sigma^2)^{-\frac{n_i+p}{2}-1} (\sigma^2)^{\frac{p}{2}} exp\Big\{ - \frac{1}{2\sigma^2}\boldsymbol{y}_i'\boldsymbol{y}_i + \frac{1}{2\sigma^2}\boldsymbol{y}_i'\tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\boldsymbol{y}_i\Big\} \\
\propto & (\sigma^2)^{-\frac{n_i}{2}-1} exp\Big\{ - \frac{1}{2\sigma^2}\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i\Big\},
\end{aligned}
$$

This indicates that,

$$
(\sigma^2, \boldsymbol{g}_i|\boldsymbol{y}_i) \sim IG\left(\frac{n_i}{2}, \frac{1}{2}\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i\right),
$$

and

$$
\begin{aligned}
& f(g_0|g_i, \boldsymbol{y}_i) \\
\propto & \int_{\sigma^2}\int_{\tilde{\boldsymbol{\beta}}} f(\tilde{\boldsymbol{\beta}}_i, \sigma^2, \boldsymbol{g}_i|\boldsymbol{y}_i) d\tilde{\boldsymbol{\beta}}_i d\sigma^2 \\
\propto & \int_{\sigma^2}\int_{\tilde{\boldsymbol{\beta}}} (\sigma^2)^{-\frac{n_i}{2}} exp\{-\frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)\} |\sigma^2\boldsymbol{C}_i|^{-\frac{1}{2}} \\
& exp\{-\frac{1}{2}\tilde{\boldsymbol{\beta}}_i'(\sigma^2\boldsymbol{C}_i)^{-1}\tilde{\boldsymbol{\beta}}_i\} \frac{1}{\sigma^2} g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) d\tilde{\boldsymbol{\beta}}_i d\sigma^2 \\
\propto & |\boldsymbol{C}_i|^{-\frac{1}{2}} g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) \int_{\sigma^2} (\sigma^2)^{-\frac{n_i+p}{2}-1} \\
& \int_{\tilde{\boldsymbol{\beta}}} exp\Big\{ - \frac{1}{2\sigma^2}(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i)'(\boldsymbol{y}_i - \tilde{\boldsymbol{X}}_i\tilde{\boldsymbol{\beta}}_i) - \frac{1}{2\sigma^2}\tilde{\boldsymbol{\beta}}_i'\boldsymbol{C}_i^{-1}\tilde{\boldsymbol{\beta}}_i\Big\}
\end{aligned}
$$

$$
\begin{aligned}
&\propto |\boldsymbol{C}_i|^{-\frac{1}{2}} g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) \int_{\sigma^2} (\sigma^2)^{-\frac{n_i+p}{2}-1} exp\Big\{ -\frac{1}{2\sigma^2}\boldsymbol{y}_i'\boldsymbol{y}_i + \frac{1}{2}\mu_\beta'\Sigma_\beta^{-1}\mu_\beta \Big\} \\
&\quad \int_{\tilde{\boldsymbol{\beta}}} exp\Big\{ -\frac{1}{2}\big[(\tilde{\boldsymbol{\beta}}_i - \mu_\beta)'\Sigma_\beta^{-1}(\tilde{\boldsymbol{\beta}}_i - \mu_\beta)\big]\Big\} d\tilde{\boldsymbol{\beta}} d\sigma^2 \\
&\propto g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) |\boldsymbol{C}_i|^{-\frac{1}{2}} \int_{\sigma^2} (\sigma^2)^{-\frac{n_i+p}{2}-1} |\Sigma_\beta|^{\frac{1}{2}} exp\Big\{ \frac{1}{2\sigma^2}\boldsymbol{y}_i'\boldsymbol{y}_i + \frac{1}{2}\mu_\beta'\Sigma_\beta^{-1}\mu_\beta \Big\} d\sigma^2 \\
&\propto g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) |\boldsymbol{C}_i|^{-\frac{1}{2}} |(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}|^{\frac{1}{2}} \int_{\sigma^2} (\sigma^2)^{-\frac{n_i}{2}-1} \\
&\quad exp\Big\{ -\frac{1}{2\sigma^2}\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i \Big\} d\sigma^2 \\
&\propto g_0^{-\frac{3}{2}} \exp(-\frac{1}{2g_0}) |\boldsymbol{C}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})|^{-\frac{1}{2}} \Big\{ \frac{1}{2}\boldsymbol{y}_i'\big[\boldsymbol{I}_{n_i} - \tilde{\boldsymbol{X}}_i(\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i + \boldsymbol{C}_i^{-1})^{-1}\tilde{\boldsymbol{X}}_i'\big]\boldsymbol{y}_i \Big\}^{-\frac{n_i}{2}}.
\end{aligned}
$$

A.2.7 Proof of Remark 3.8

Proof. Following definitions in Min and Sun (2016), consider the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and assume that there are m blocks design matrices $\boldsymbol{X} = (\boldsymbol{X}_1', \cdots, \boldsymbol{X}_m')'$ with the corresponding regression coefficients as $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_m')'$. Then, we consider the independent g-priors for $\boldsymbol{\beta}_j$

$$
p(\boldsymbol{\beta}_j|\sigma^2) \propto \exp(-\frac{1}{2g_i\sigma^2}\boldsymbol{\beta}_j'\boldsymbol{X}_j'\boldsymbol{X}_j\boldsymbol{\beta}_j). \tag{A.41}
$$

Denote $\boldsymbol{\gamma} \subseteq \{0, 1, \cdots, m\}$ and $\boldsymbol{\Gamma}$ as the collection of nonempty subset of $\{0, 1, \cdots, m\}$, where $\boldsymbol{\Gamma}$ serves as the index set. Under the commutativity condition of the projection matrices, $\boldsymbol{P}_i\boldsymbol{P}_j = \boldsymbol{P}_j\boldsymbol{P}_i, \forall i, j$,we could further define

$$
\begin{aligned}
\boldsymbol{P}_{\boldsymbol{\gamma}} &= \prod_{j\in\boldsymbol{\gamma}} \boldsymbol{P}_j, \\
\boldsymbol{A}_{\boldsymbol{\gamma}} &= \prod_{j\in\boldsymbol{\gamma}} \boldsymbol{P}_j \prod_{j'\in\{1,\cdots,m\}-\boldsymbol{\gamma}} (\boldsymbol{I}_n - \boldsymbol{P}_{j'}),
\end{aligned}
$$

$$p_{\boldsymbol{\gamma}} = rank(\boldsymbol{A}_{\boldsymbol{\gamma}}).$$

By Theorem 2 in Min and Sun (2016), the marginal density of $m(\boldsymbol{y}|\boldsymbol{g})$ is proportional to

$$m(\boldsymbol{y}|\boldsymbol{g}) \propto \frac{1}{(\boldsymbol{y}'\boldsymbol{R}\boldsymbol{y})^{\frac{n}{2}}} \prod_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{1}{(1+\sum_{j\in\boldsymbol{\gamma}} g_j)^{\frac{p_{\boldsymbol{\gamma}}}{2}}} \tag{A.42}$$

where

$$\boldsymbol{R} = \boldsymbol{I}_n + \sum_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} u_{\boldsymbol{\gamma}} \boldsymbol{P}_{\boldsymbol{\gamma}},\ u_{\boldsymbol{\gamma}} = (-1)^k \sum_{(j_1,\cdots,j_k)\in\boldsymbol{\gamma}} \left(\frac{g_{j_1}}{1+g_{j_1}} \frac{g_{j_2}}{1+g_{j_1}+g_{j_2}} \cdots \frac{g_k}{1+g_{j_1}+\cdots+g_{j_k}}\right).$$

where $k = |\boldsymbol{\gamma}|$ and $(j_1, \cdots, j_k)$ takes over all permutations of $\boldsymbol{\gamma}$.

In our case, for M_1, we have two blocks, $m = 2$ with $\boldsymbol{\Gamma} = \{\{0\},\{1\},\{0,1\}\}$, $\boldsymbol{P}_0 = \boldsymbol{P}_{X_{01}}$ and $\boldsymbol{P}_1 = \boldsymbol{P}_{X_1}$. Then, the marginal density $m(\boldsymbol{y}_1|g_0, g_1)$ is proportional to

$$\begin{aligned} m(\boldsymbol{y}_1|g_0,g_1) \propto & (1+g_0)^{-\frac{p_0}{2}}(1+g_1)^{-\frac{p_1}{2}}(1+g_0+g_1)^{-\frac{p_2}{2}}(\boldsymbol{y}_1'(\boldsymbol{I}_{n_1} - \frac{g_0}{1+g_0}\boldsymbol{P}_0 - \frac{g_1}{1+g_1}\boldsymbol{P}_1 \\ & + \left(\frac{g_0 g_1}{(1+g_0)(1+g_0+g_1)} + \frac{g_0 g_1}{(1+g_1)(1+g_0+g_1)}\right)\boldsymbol{P}_0\boldsymbol{P}_1)\boldsymbol{y}_1)^{-\frac{n_1}{2}}, \end{aligned}$$

where p_0 =rank$(\boldsymbol{P}_0(\boldsymbol{I}_{n_1} - \boldsymbol{P}_1))$, p_1 =rank$(\boldsymbol{P}_1(\boldsymbol{I}_{n_1} - \boldsymbol{P}_0))$ and p_2 =rank$(\boldsymbol{P}_0\boldsymbol{P}_1)$. If we further assume the orthognality of $\boldsymbol{X}_0'\boldsymbol{X}_1 = \boldsymbol{0}$, then $\boldsymbol{P}_0\boldsymbol{P}_1 = \boldsymbol{P}_1\boldsymbol{P}_0 = \boldsymbol{0}$ with p_0 =rank$(\boldsymbol{P}_0)$ =rank$(\boldsymbol{X}_0)$, p_1 =rank$(\boldsymbol{P}_1)$ =rank$(\boldsymbol{X}_1)$ and $p_2 = 0$. The marignal density is

$$m(\boldsymbol{y}_1|g_0,g_1) \propto (1+g_0)^{-\frac{p_0}{2}}(1+g_1)^{-\frac{p_1}{2}}(\boldsymbol{y}_1'(\boldsymbol{I}_{n_1} - \frac{g_0}{1+g_0}\boldsymbol{P}_0 - \frac{g_1}{1+g_1}\boldsymbol{P}_1)\boldsymbol{y}_1)^{-\frac{n_1}{2}}.$$

□

A.3 Full Conditional Distributions in Chapter 4

For convenience, consider a general reparametrized model $y_{ij} \sim Bin(n_{ij}, p_{ij}), v_{ij} = log(p_{ij}/(1-p_{ij})), \boldsymbol{v} = \mu \mathbf{1}_n + \boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0 + \boldsymbol{X}_\gamma \boldsymbol{\gamma} + \boldsymbol{X}_z \boldsymbol{z} + \boldsymbol{\epsilon}$, where $\boldsymbol{v}$ is a vector of $v_{ij}, i = 1, \cdots, I, j = 1, \cdots, J, n = IJ$, and $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$. $\boldsymbol{\beta}_0$ is fixed, $\boldsymbol{X}_{\beta_0}$ is continuous, $\boldsymbol{\beta}_\gamma$ is fixed, $\boldsymbol{X}_\gamma$ is categorical, $\boldsymbol{z}$ corresponds to random effects. The priors are specified as: $\mu \propto 1, \sigma^2 \propto 1/\sigma^2, \boldsymbol{\beta}_0 \sim N_{p_0}(\mathbf{0}, g_0\sigma^2(\boldsymbol{X}_{\beta_0}'\boldsymbol{X}_{\beta_0})^{-1}), \boldsymbol{\gamma} \sim N_K(\mathbf{0}, \boldsymbol{G}\sigma^2)$, where $\boldsymbol{G} = diag(g_1, \cdots, g_K), \boldsymbol{z} \sim N_I(\mathbf{0}, \delta(\boldsymbol{I} - \rho \boldsymbol{C})^{-1}), g_0 \sim IG(1/2, n/2), g_j \sim IG(1/2, 1/2), \delta \sim IG(1/2, 1/2), \rho \sim Unif(\rho_{min}, \rho_{max})$. Then, denote $\boldsymbol{X} = (\mathbf{1}_n, \boldsymbol{X}_{\beta_0}, \boldsymbol{X}_\gamma, \boldsymbol{X}_z)$, $\boldsymbol{\beta} = (\mu, \boldsymbol{\beta}_0', \boldsymbol{\gamma}', \boldsymbol{z}')'$. Then, $\bar{\boldsymbol{X}}_\mu = (\boldsymbol{X}_{\beta_0}, \boldsymbol{X}_\gamma, \boldsymbol{X}_z), \bar{\boldsymbol{X}}_{\beta_0} = (\mathbf{1}_n, \boldsymbol{X}_\gamma, \boldsymbol{X}_z), \bar{\boldsymbol{X}}_\gamma = (\mathbf{1}_{IJ}, \boldsymbol{X}_{\beta_0}, \boldsymbol{X}_z)$, $\bar{\boldsymbol{X}}_z = (\mathbf{1}_n, \boldsymbol{X}_{\beta_0}, \boldsymbol{X}_\gamma), \bar{\boldsymbol{\mu}} = (\boldsymbol{\beta}_0', \boldsymbol{\gamma}', \boldsymbol{z}'), \bar{\boldsymbol{\beta}}_0 = (\mu, \boldsymbol{\gamma}', \boldsymbol{z}')', \bar{\boldsymbol{\gamma}} = (\mu, \boldsymbol{\beta}_0', \boldsymbol{z}')', \bar{\boldsymbol{z}} = (\mu, \boldsymbol{\beta}_0', \boldsymbol{\gamma}')'$. Let $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}_n$ and the joint posterior density of $(\mu, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, \boldsymbol{z}, \sigma^2, \delta, \rho, g_0, g_1, g_2)$ given $\boldsymbol{y}$:

$$
\begin{aligned}
&f(\boldsymbol{v}, \mu, \boldsymbol{\beta}_0, \boldsymbol{\gamma}, \boldsymbol{z}, \sigma^2, \delta, \rho, g_0, g_1, g_2 | \boldsymbol{y}) \\
\propto& f(\boldsymbol{y}|\boldsymbol{v}) f(\boldsymbol{v}) f(\mu) f(\boldsymbol{\beta}_0|\sigma^2, g_0) f(\boldsymbol{\gamma}|\sigma^2, g_1, g_2) f(\boldsymbol{z}|\delta, \rho) f(\sigma^2) f(\delta) f(g_0) f(g_1) f(g_2),
\end{aligned}
$$

Let $\cdot$ denote all remaining parameters.

1. The full conditional distribution for $v_{ij}|\cdot$ is

$$
f(v_{ij}|\cdot) \propto exp\left\{ \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}v_{ij} - n_{ij}log(1+e^{v_{ij}}) - \frac{1}{2}(\boldsymbol{v} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v} - \boldsymbol{X}\boldsymbol{\beta}) \right\}.
$$

2. The full conditional distribution for $\mu|\cdot$

$$
\begin{aligned}
f(\mu|\cdot) &\propto exp\left\{-\frac{1}{2}(\boldsymbol{v}-\mathbf{1}_n\mu-\bar{\boldsymbol{X}}_\mu\bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\mathbf{1}_n\mu-\bar{\boldsymbol{X}}_\mu\bar{\boldsymbol{\mu}})\right\} \\
&\propto exp\{-\frac{1}{2}[\frac{n}{\sigma^2}\mu^2-2(\boldsymbol{v}-\bar{\boldsymbol{X}}_\mu\bar{\boldsymbol{\mu}})'\boldsymbol{\Sigma}^{-1}\mathbf{1}_n\mu]\},
\end{aligned}
$$

which implies that $(\mu|\cdot)\sim N\left(u,\sigma_\mu^2\right)$, where $u=\sigma_u^2\mathbf{1}_n'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{X}}_\mu\bar{\boldsymbol{\mu}}), \sigma_\mu^2=\sigma^2/n.$

3. The full conditional distribution for $\boldsymbol{\beta}_0|\cdot$ is

$$
\begin{aligned}
f(\boldsymbol{\beta}_0|\cdot) \propto & exp\{-\frac{1}{2}(\boldsymbol{v}-\bar{\boldsymbol{X}}_{\beta_0}\bar{\boldsymbol{\beta}}_0-\boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{X}}_{\beta_0}\bar{\boldsymbol{\beta}}_0-\boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0)\} \\
& exp\{-\frac{1}{2g_0\sigma^2}\boldsymbol{\beta}_0'(\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0})\boldsymbol{\beta}_0\},
\end{aligned}
$$

which implies that $(\boldsymbol{\beta}_0|\cdot)\sim N_{p_0}\left(\boldsymbol{u}_{\beta_0},\boldsymbol{\Sigma}_{\beta_0}\right)$, where

$$
\begin{aligned}
\boldsymbol{u}_{\beta_0} =& \boldsymbol{\Sigma}_{\beta_0}\boldsymbol{X}_{\beta_0}{}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{X}}_{\beta_0}\bar{\boldsymbol{\beta}}_0) \\
\boldsymbol{\Sigma}_{\beta_0} =& \left[\boldsymbol{X}_{\beta_0}{}'(\boldsymbol{\Sigma}^{-1}+(g\sigma^2)^{-1}\boldsymbol{I}_{p_0})\boldsymbol{X}_{\beta_0}\right]^{-1}=\frac{g_0\sigma^2}{g_0+1}\left(\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0}\right)^{-1}.
\end{aligned}
$$

4. The full conditional distribution for $\boldsymbol{\gamma}|\cdot$ is

$$
\begin{aligned}
f(\boldsymbol{\gamma}|\cdot) &\propto exp\{-\frac{1}{2}(\boldsymbol{v}-\bar{\boldsymbol{X}}_\gamma\bar{\boldsymbol{\gamma}}-\boldsymbol{X}_\gamma\boldsymbol{\gamma})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{X}}_\gamma\bar{\boldsymbol{\gamma}}-\boldsymbol{X}_\gamma\boldsymbol{\gamma})\}\cdot exp\{-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\gamma}\} \\
&\propto exp\{-\frac{1}{2}[\boldsymbol{\gamma}'(\boldsymbol{X}_\gamma'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_\gamma+\Lambda^{-1})\boldsymbol{\gamma}-2(\boldsymbol{v}-\bar{\boldsymbol{X}}_\gamma\bar{\boldsymbol{\gamma}})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_\gamma\boldsymbol{\gamma}]\}
\end{aligned}
$$

where $\boldsymbol{\Lambda}=diag(g_1\sigma^2,g_2\sigma^2)$. This indicates $\boldsymbol{\gamma}\sim N_K\left(\boldsymbol{u}_\gamma,\boldsymbol{\Sigma}_\gamma\right)$, where

$$
\boldsymbol{u}_\gamma=\boldsymbol{\Sigma}_\gamma\boldsymbol{X}_\gamma'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\bar{\boldsymbol{X}}_\gamma\bar{\boldsymbol{\gamma}}),
$$

$$\boldsymbol{\Sigma}_\gamma = diag(\frac{g_1\sigma^2}{1+g_1}, \frac{g_2\sigma^2}{1+g_2}).$$

5. The full conditional distribution for $\boldsymbol{z}|\cdot$ is

$$\begin{aligned}
& f(\boldsymbol{z}|\cdot) \\
\propto & exp\{-\frac{1}{2}(\boldsymbol{v}-\bar{\boldsymbol{X}}_z\bar{\boldsymbol{z}}-\boldsymbol{X}_z\boldsymbol{z})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{v}-\bar{\boldsymbol{X}}_z\bar{\boldsymbol{z}}-\boldsymbol{X}_z\boldsymbol{z})\}exp\{-\frac{1}{2}\boldsymbol{z}'[\delta(\boldsymbol{I}-\rho\boldsymbol{C})^{-1}]^{-1}\boldsymbol{z}\} \\
\propto & exp\Big\{-\frac{1}{2}\big[\big(\boldsymbol{z}'[\boldsymbol{X}_z'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_z+\delta^{-1}(\boldsymbol{I}-\rho\boldsymbol{C})]\boldsymbol{z}\big)-2(\boldsymbol{v}-\bar{\boldsymbol{X}}_z\bar{\boldsymbol{z}})'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_z\boldsymbol{z}\big]\Big\},
\end{aligned}$$

which indicates that $(\boldsymbol{z}|\cdot) \sim N_{p_z}(\boldsymbol{u}_z, \boldsymbol{\Sigma}_z)$, where

$$\begin{aligned}
\boldsymbol{u}_z &= \boldsymbol{\Sigma}_z\boldsymbol{X}_z'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\bar{\boldsymbol{X}}_z\bar{\boldsymbol{z}}), \\
\boldsymbol{\Sigma}_z &= [\boldsymbol{X}_z'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_z+\delta^{-1}(\boldsymbol{I}-\rho\boldsymbol{C})]^{-1}.
\end{aligned}$$

6. The full conditional distribution for $\sigma^2|\cdot$

$$\begin{aligned}
f(\sigma^2|\cdot) \propto & |\sigma^2\boldsymbol{I}_n|^{-\frac{1}{2}}exp\{-\frac{1}{2\sigma^2}(\boldsymbol{v}-\boldsymbol{X\beta})'(\boldsymbol{v}-\boldsymbol{X\beta})\}|g_0\sigma^2(\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0})^{-1}|^{-\frac{1}{2}}(\sigma^2)^{-1} \\
& exp\{-\frac{1}{2g_0\sigma^2}\boldsymbol{\beta}_0'\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0\}(g_1\sigma^2)^{-\frac{1}{2}}(g_2\sigma^2)^{-\frac{1}{2}}exp\{-\frac{1}{2g_1\sigma^2}\gamma_1^2-\frac{1}{2g_2\sigma^2}\gamma_2^2\} \\
\propto & (\sigma^2)^{-\frac{n+p_0}{2}-2}exp\left\{-\frac{1}{\sigma^2}[\frac{(\boldsymbol{v}-\boldsymbol{X\beta})'(\boldsymbol{v}-\boldsymbol{X\beta})}{2}+\frac{{\gamma_1}^2}{2g_1}\right. \\
& \left.+\frac{{\gamma_2}^2}{2g_2}+\frac{\boldsymbol{\beta}_0'\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0}{2g_0}]\right\}
\end{aligned}$$

7. The full conditional distribution for $\delta|\cdot$ is

$$f(\delta|\cdot) \propto |\delta(\boldsymbol{I}-\rho\boldsymbol{C})^{-1}|^{-\frac{1}{2}}exp\{-\frac{1}{2\delta}\boldsymbol{z}'(\boldsymbol{I}-\rho\boldsymbol{C})\boldsymbol{z}\}\frac{1}{\delta^{a_0+1}}exp\{-\frac{b_0}{\delta}\}$$

$$\propto \delta^{-(\frac{p_I}{2}+a_0)-1} exp\{-\frac{1}{\delta}[\frac{\boldsymbol{z}'(\boldsymbol{I}-\rho\boldsymbol{C})\boldsymbol{z}}{2}+b_0]\},$$

which implies that

$$(\delta|\cdot) \sim IG\left(\frac{p_c}{2}+a_0, \frac{\boldsymbol{z}'(\boldsymbol{I}-\rho\boldsymbol{C})\boldsymbol{z}}{2}+b_0\right).$$

8. The full conditional distribution for $\rho|\cdot$ is

$$\begin{aligned} & f(\rho|\cdot) \\ & \propto |\delta(\boldsymbol{I}-\rho\boldsymbol{C})^{-1}|^{-\frac{1}{2}} exp\{-\frac{1}{2\delta}\boldsymbol{z}'(\boldsymbol{I}-\rho\boldsymbol{C})\boldsymbol{z}\}\frac{1}{\rho_{max}-\rho_{min}}\mathbf{1}_\rho(\rho_{min}<\rho<\rho_{max}) \\ & \propto |\delta(\boldsymbol{I}-\rho\boldsymbol{C})^{-1}|^{-\frac{1}{2}} exp\left\{\frac{\rho\boldsymbol{z}'\boldsymbol{C}\boldsymbol{z}}{2\delta}\right\}\mathbf{1}_\rho(\rho_{min}<\rho<\rho_{max}). \end{aligned}$$

9. The full conditional distribution for $g_0|\cdot$ is

$$\begin{aligned} f(g_0|\cdot) & \propto |g_0\sigma^2(\boldsymbol{X}_{\beta_0}'\boldsymbol{X}_{\beta_0})^{-1}|^{-\frac{1}{2}} exp\left\{-\frac{1}{2g_0\sigma^2}\boldsymbol{\beta}_0'(\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0})\boldsymbol{\beta}_0\right\} g_0^{-\frac{3}{2}} exp\left\{-\frac{n}{2g_0}\right\} \\ & \propto g_0^{-\frac{p_0+1}{2}-1} exp\left\{-\frac{1}{2g_0}[\frac{\boldsymbol{\beta}_0'\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0}\boldsymbol{\beta}_0}{\sigma^2}+n]\right\}, \end{aligned}$$

which implies that

$$(g_0|\cdot) \sim IG\left((p_0+1)/2, [\frac{\boldsymbol{\beta}_0'(\boldsymbol{X}_{\beta_0}{}'\boldsymbol{X}_{\beta_0})\boldsymbol{\beta}_0}{\sigma^2}+n]/2\right).$$

10. The full conditional distribution for $g_i|\cdot$ is

$$f(g_1, g_2|\cdot) \propto |\Lambda|^{-\frac{1}{2}} exp\{-\frac{1}{2}\boldsymbol{\gamma}'\Lambda^{-1}\boldsymbol{\gamma}\} g_1^{-\frac{3}{2}} exp\{-\frac{1}{2g_1}\} g_2^{-\frac{3}{2}} exp\{-\frac{1}{2g_2}\}$$

$$\propto g_1^{-\frac{1}{2}} exp\{-\frac{\gamma_1^2}{2g_1\sigma^2}\} g_1^{-\frac{3}{2}} exp\{-\frac{1}{2g_1}\} \cdot g_2^{-\frac{1}{2}} exp\{-\frac{\gamma_2^2}{2g_2\sigma^2}\} g_2^{-\frac{3}{2}} exp\{-\frac{1}{2g_2}\}$$

$$\propto g_1^{-1-1} exp\left\{-\frac{1}{2g_1}[\frac{\gamma_1^2}{\sigma^2}+1]\right\} \cdot g_2^{-1-1} exp\left\{-\frac{1}{2g_2}[\frac{\gamma_2^2}{\sigma^2}+1]\right\},$$

which implies that $f(g_1|\cdot)$ and $f(g_2|\cdot)$ are conditional independent IG distributions with

$$(g_i|\cdot) \sim IG\left(1, [\frac{(\gamma_i)^2}{\sigma^2}+1]/2\right), i = 1, 2.$$

Bibliography

Agliari, A. and Parisetti, C. C. (1988). A-g Reference Informative Prior: A Note on Zellner's g Prior. *The Statistician*, 37(3):271.

Bailey, W. N. (1935). *Generalized hypergeometric series*. The University Press (Cambridge tracts in mathematics and mathematical physics, no. 32).

Bayarri, M. J., Berger, J. O., Forte, A., and García-Donato, G. (2012). Criteria for Bayesian model choice with application to variable selection. *The Annals of Statistics*, 40(3):1550 – 1577.

Bayarri, M. J., Berger, J. O., Jang, W., Ray, S., Pericchi, L. R., and Visser, I. (2019). Prior-based Bayesian information criterion. *Statistical Theory and Related Fields*, 3(1):2–13.

Berger, J., Bayarri, M. J., and Pericchi, L. R. (2014). The Effective Sample Size. *Econometric Reviews*, 33(1-4):197–217.

Berger, J., Jang, W., Ray, S., Pericchi, L. R., Rejoinder, I., Berger, J., and Fields, R. (2019). Rejoinder by James Berger, Woncheol Jang, Surajit Ray, Luis R. Pericchi and Ingmar Visser. *Statistical Theory and Related Fields*, 3(1):37–39.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*. Springer-Verlag, New York.

Berger, J. O., Strawderman, W., and Tang, D. (2005). Posterior propriety and admissibility of hyperpriors in normal hierarchical models. *Annals of Statistics*, 33(2):606–646.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43(1):1–20.

Bodnar, O., Link, A., Arendacká, B., Possolo, A., and Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine*, 36(2):378–399.

Brockwell, S. E. and Gordon, I. R. (2001). A comparison of statistical methods for meta-analysis. *Statistics in Medicine*, 20(6):825–840.

Bujkiewicz, S., Thompson, J. R., Riley, R. D., and Abrams, K. R. (2016). Bayesian meta-analytical methods to incorporate multiple surrogate endpoints in drug development process. *Statistics in Medicine*, 35(7):1063–1089.

Bujkiewicz, S., Thompson, J. R., Sutton, A. J., Cooper, N. J., Harrison, M. J., Symmons, D. P., and Abrams, K. R. (2013). Multivariate meta-analysis of mixed outcomes: A Bayesian approach. *Statistics in Medicine*, 32(22):3926–3943.

Burke, D. L., Ensor, J., and Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine*, 36(5):335–351.

Chen, D.-G., Liu, D., Min, X., and Zhang, H. (2020). Relative efficiency of using summary versus individual data in random-effects meta-analysis. *Biometrics*, 76(4):1319–1329.

Chung, Y., Rabe-Hesketh, S., and Choi, I.-H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine*, 32(23):4071–4089.

Clyde, M. and George, E. I. (2000). Flexible empirical Bayes estimation for wavelets. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 62(4):681–698.

Cooper, H. and Patall, E. A. (2009). The Relative Benefits of Meta-Analysis Conducted With Individual Participant Data Versus Aggregated Data. *Psychological Methods*, 14(2):165–176.

Cortez, P. and Silva, A. M. G. (2008). Using data mining to predict secondary school student performance.

Cui, W. and George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138(4):888–900.

De Oliveira, V. (2012). Bayesian analysis of conditional autoregressive models. *Annals of the Institute of Statistical Mathematics*, 64(1):107–133.

Dean, C., Ugarte, M., and Militino, A. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57(1):197–202.

Du, J. (2018). *Bayesian Hierarchical Modeling of Colorectal and Breast Cancer Data in Missouri*. PhD thesis, University of Missouri–Columbia.

Egozcue, M., Garcia, L., and Wong, W. (2009). On some covariance inequalities for monotonic and non-monotonic functions. *Journal of Inequalities in Pure and Applied Mathematics*, 10(3):1–7.

Findley, D. F. (1991). Counterexamples to parsimony and BIC. *Annals of the Institute of Statistical Mathematics*, 43(3):505–514.

Fragoso, T. M., Bertoli, W., and Louzada, F. (2018). Bayesian model averaging: A systematic review and conceptual classification. *International Statistical Review*, 86(1):1–28.

Garcia-Estevez, L. and Moreno-Bueno, G. (2019). Updating the role of obesity and cholesterol in breast cancer. *Breast Cancer Research*, 21(1):35.

George, E. I. (2000). The Variable Selection Problem. *Journal of the American Statistical Association*, 95(452):1304–1308.

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). Adaptive rejection metropolis sampling within gibbs sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4):455–472.

Goel, P. K. and Zellner, A. (1986). *Bayesian inference and decision techniques : essays in honor of Bruno de Finetti.* Studies in Bayesian econometrics and statistics: v. 6. North-Holland Pub. Co.

Hanks, E. M., Schliep, E. M., Hooten, M. B., and Hoeting, J. A. (2015). Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. *Environmetrics*, 26(4):243–254.

Hansen, M. H. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774.

Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *American Statistician*, 64(4):325–334.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–401.

Hong, H., Wang, C., and Rosner, G. L. (2021). Meta-analysis of rare adverse events in randomized clinical trials: Bayesian and frequentist methods. *Clinical Trials*, 18(1):3–16. PMID: 33258698.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 75(1):139–159.

Hurtado Rúa, S. M., Mazumdar, M., and Strawderman, R. L. (2015). The choice of prior distribution for a covariance matrix in multivariate meta-analysis: a simulation study. *Statistics in Medicine*, 34(30):4083–4104.

Jackson, C. H., Best, N. G., and Richardson, S. (2009). Bayesian graphical models for regression on multiple data sets with different variables. *Biostatistics (Oxford, England)*, 10(2):335–351.

Jackson, D., Riley, R., and White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20):2481–2498.

Jahan, F., Duncan, E. W., Cramb, S. M., Baade, P. D., and Mengersen, K. L. (2020). Multivariate Bayesian meta-analysis: joint modelling of multiple cancer types using summary statistics. *International Journal of Health Geographics*, 19(1):1–19.

Kass, R. E. and Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, 90(430):773–795.

Kass, R. E. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*, 90(431):928–934.

Kim, H., Sun, D., and Tsutakawa, R. K. (2001). A bivariate bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association*, 96(456):1506–1521.

Lantz, B. (2013). *Machine Learning with R*. Packt Publishing.

Lee, D. and Mitchell, R. (2012). Boundary detection in disease mapping studies. *Biostatistics*, 13(3):415–426.

Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, pages 179–191. Springer.

Li, H., Lim, D., Chen, M. H., Ibrahim, J. G., Kim, S., Shah, A. K., and Lin, J. (2021). Bayesian network meta-regression hierarchical models using heavy-tailed multivariate random effects with covariate-dependent variances. *Statistics in Medicine*, 40(15):3582–3603.

Li, Y. and Clyde, M. A. (2018). Mixtures of g-priors in generalized linear models. *Journal of the American Statistical Association*, 113(524):1828–1845.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, 103(481):410–423.

Lin, L. and Chu, H. (2018). Bayesian multivariate meta-analysis of multiple factors. *Research Synthesis Methods*, 9(2):261–272.

MacNab, Y. C. (2022). Bayesian disease mapping: Past, present, and future. *Spatial Statistics*, page 100593.

Maruyama, Y. and George, E. I. (2011). Fully Bayes factors with a generalized g-prior. *The Annals of Statistics*, 39(5):2740 – 2765.

Mercer, L., Wakefield, J., Chen, C., and Lumley, T. (2014). A comparison of spatial smoothing methods for small area estimation with sampling weights. *Spatial statistics*, 8:69–85.

Min, X. and Sun, D. (2016). Bayesian model selection for a linear model with grouped covariates. *Annals of the Institute of Statistical Mathematics*, 68(4):877–903.

Moreno, E., Vázquez-Polo, F. J., and Negrín, M. A. (2018). Bayesian meta-analysis: The role of the between-sample heterogeneity. *Statistical Methods in Medical Research*, 27(12):3643–3657.

Paciorek, C. J. (2010). The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators. *Statistical Science*, 25(1):107 – 125.

Page, G. L., Liu, Y., He, Z., and Sun, D. (2017). Estimation and prediction in the presence of spatial confounding for spatial linear models. *Scandinavian Journal of Statistics*, 44(3):780–797.

Pfeffermann, D. et al. (2013). New important developments in small area estimation. *Statistical Science*, 28(1):40–68.

Riley, R. D., Abrams, K. R., Lambert, P. C., Sutton, A. J., and Thompson, J. R. (2007). An evaluation of bivariate random-effects meta-analysis for the joint synthesis of two correlated outcomes. *Statistics in Medicine*, 26(1):78–97.

Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L., and Boutitie, F. (2008). Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Statistics in Medicine*, 27(11):1870–1893.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default bayes factors for anova designs. *Journal of Mathematical Psychology*, 56(5):356 – 374.

Schmaltz, C. L. (2012). *Marginally modeling misaligned regions and handling masked failure causes with imprecision*. PhD thesis, University of Missouri–Columbia.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Shao, J. (2003). *Mathematical Statistics*. Springer Texts in Statistics. Springer New York, New York, NY.

Siegel, L., Rudser, K., Sutcliffe, S., Markland, A., Brubaker, L., Gahagan, S., Stapleton, A. E., and Chu, H. (2020). A Bayesian multivariate meta-analysis of prevalence data. *Statistics in Medicine*, 39(23):3105–3119.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(1):1–52.

Simpson, D., Rue, H., Riebler, A., Martins, T. G., Sørbye, S. H., et al. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28.

Som, A., Hans, C. M., and MacEachern, S. N. (2015). Block Hyper-g Priors in Bayesian Regression.

Som, A., Hans, C. M., and MacEachern, S. N. (2016). A conditional Lindley paradox in Bayesian linear models. *Biometrika*, 103(4):993–999.

Sparks, D. K., Khare, K., and Ghosh, M. (2015). Necessary and Sufficient Conditions for High-Dimensional Posterior Consistency under g-Priors. *Bayesian Analysis*, 10(3):627 – 664.

Stone, M. (1979). Comments on Model Selection Criteria of Akaike and Schwarz. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):276–278.

Sun, D., Tsutakawa, R., and Speckman, P. (2004). Posterior Distribution of Hierarchical Models Using CAR(1) Distributions. *Biometrika*, 86.

Thomsen, I. and Holmøy, A. M. K. (1998). Combining Data from Surveys and Administrative Record Systems. The Norwegian Experience. *International Statistical Review / Revue Internationale de Statistique*, 66(2):201–221.

Tierney, L. and B.Kadane, J. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):82–86.

Verde, P. E., Ohmann, C., Morbach, S., and Icks, A. (2016). Bayesian evidence synthesis for exploring generalizability of treatment effects: A case study of combining randomized and non-randomized results in diabetes. *Statistics in Medicine*, 35(10):1654–1675.

Wang, M. (2017). Mixtures of g -priors for analysis of variance models with a diverging number of parameters. *Bayesian Anal.*, 12(2):511–532.

Weakliem, D. L. (1999). A critique of the Bayesian information criterion for model selection. *Sociological Methods & Research*, 27(3):359–397.

Wei, Y. and Higgins, J. P. T. (2013). Bayesian multivariate meta-analysis with multiple outcomes. *Statistics in medicine*, 32(17):2911–34.

Wei, Y., Huang, Y., Yang, W., Huang, Q., Chen, Y., Zeng, K., Chen, J., and Chen, J. (2021). The significances and clinical implications of cholesterol components in human breast cancer. *Science Progress*, 104(3):00368504211028395. PMID: 34510991.

Williams, D., Rast, P., and Bürkner, P.-C. (2018). Bayesian Meta-Analysis with Weakly Informative Prior Distributions.

Woodard, R., Sun, D., He, Z., and Sheriff, S. L. (1999). Estimating hunting success rates via Bayesian generalized linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(4):456–472.

Wu, H.-H., Ferreira, M. A. R., and Gompper, M. E. (2016). Consistency of hyper-g-prior-based bayesian variable selection for generalized linear models. *Brazilian Journal of Probability and Statistics*, 30(4):691–709.

Yuan, Z. and Yang, Y. (2005). Combining linear regression models: When and how? *Journal of the American Statistical Association*, 100(472):1202–1214.

Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association*, 57(298):348–368.

Zellner, A. and Siow, A. (1980). Posterior odds ratios for selected regression hypotheses. *Trabajos de Estadistica Y de Investigacion Operativa*, 31(1):585–603.

Zhang, H., Huang, X., Gan, J., Karmaus, W., and Sabo-Attwood, T. (2016). A two-component G-Prior for variable selection. *Bayesian Analysis*, 11(2):353–380.

Zhang, J., Carlin, B. P., Neaton, J. D., Soon, G. G., Nie, L., Kane, R., Virnig, B. A., and Chu, H. (2014). Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clinical Trials*, 11(2):246–262.

Zhang, Z., Jordan, M., and Yeung, D.-Y. (2009). Posterior consistency of the silverman g-prior in bayesian model choice. In Koller, D., Schuurmans, D., Bengio, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc.

VITA

Qiao Wang was born in Shanxi, China, on August 1, 1992. She earned her B.S. from the Central China Normal University (2015) in Mathematics and Statistics. She got her M.A. degree in Statistics at the University of Missouri, Columbia, Missouri (2017). Later on, she pursued a doctoral degree in Statistics under the supervision of Drs. Dongchu Sun and Chong(Zhuoqiong) He. After graduation, she decides to work as a postdoc associate with Dr. Hwanhee Hong at the Duke University.