Detection of Latent Heteroscedasticity and Group-Based Regression Effects in Linear Models via Bayesian Model Selection

Thomas Anthony Metzger

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Christopher T. Franck, Chair
Marco A. R. Ferreira
Robert B. Gramacy
Leanna L. House

July 18, 2019
Blacksburg, Virginia

Keywords: heteroscedasticity, model selection, Bayesian

Detection of Latent Heteroscedasticity and Group-Based Regression Effects in Linear Models via Bayesian Model Selection

Thomas Anthony Metzger

(ABSTRACT)

Standard linear modeling approaches make potentially simplistic assumptions regarding the structure of categorical effects that may obfuscate more complex relationships governing data. For example, recent work focused on the two-way unreplicated layout has shown that hidden groupings among the levels of one categorical predictor frequently interact with the ungrouped factor. We extend the notion of a "latent grouping factor" to linear models in general. The proposed work allows researchers to determine whether an apparent grouping of the levels of a categorical predictor reveals a plausible hidden structure given the observed data. Specifically, we offer Bayesian model selection-based approaches to reveal latent group-based heteroscedasticity, regression effects, and/or interactions. Failure to account for such structures can produce misleading conclusions. Since the presence of latent group structures is frequently unknown *a priori* to the researcher, we use fractional Bayes factor methods and mixture g-priors to overcome lack of prior information. We provide an R package, `slgf`, that implements our methodology in practice, and demonstrate its usage in practice.

Detection of Latent Heteroscedasticity and Group-Based Regression Effects in Linear Models via Bayesian Model Selection

Thomas Anthony Metzger

(GENERAL AUDIENCE ABSTRACT)

Statistical models are a powerful tool for describing a broad range of phenomena in our world. However, many common statistical models may make assumptions that are overly simplistic and fail to account for key trends and patterns in data. Specifically, we search for hidden structures formed by partitioning a dataset into two groups. These two groups may have distinct variability, statistical effects, or other hidden effects that are missed by conventional approaches. We illustrate the ability of our method to detect these patterns through a variety of disciplines and data layouts, and provide software for researchers to implement this approach in practice.

Dedication

To the many, many educators throughout my life whose knowledge, enthusiasm, and patience have led me to this rich and rewarding field.

Acknowledgments

My time at Virginia Tech has been one of the most unexpected and rewarding experiences of my life. I am eternally grateful to the many family members, teachers, colleagues, friends, and students whose support made it possible.

Thank you to my family, for fostering my desire to learn. Your patience and enthusiasm for my education and new career have been unfaltering. I am happy to return to The Great State of Ohio, only six years late.

Thank you the faculty and staff of Hoban High School, who were the first to show me what mathematics and statistics could be. Thank you to my classmates, whose friendships that have remained over the years will always be appreciated.

Thank you to the faculty and staff of The Ohio State University, for your role in my development as a mathematician. I am proud to have the opportunity to return to work with the faculty and students I know well. Thank you also to my friends from Ohio State, who have always stood with me when I needed it.

Thank you to the faculty, staff, and students of Staunton High School, and to my friends there for their support. I will always be grateful for my time and experiences in Staunton that laid the foundation for who I am as an educator.

Thank you to Dr. Jeff Birch, for giving me the opportunity to join Virginia Tech and earn my Ph.D.

Thank you to the faculty and graduate students of the Virginia Tech Department of Statistics. It has been a privilege to learn and work with you for five years, and your friendship has

been an integral part of my success as a researcher and teacher. I look forward to remaining a part of the Virginia Tech community.

Thank you to Marco Ferreira, Bobby Gramacy, and Leanna House, for your suggestions on this work, and for your mentorship. I am proud to have completed this work with your guidance.

Finally, thank you to Chris Franck, a great statistician, researcher, teacher, and friend. Every accomplishment I might have in my career is thanks to you. I begin my new career with great optimism as a result of your guidance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this work we propose the Suspected Latent Grouping Factor (SLGF) methodology for Bayesian model selection. Our method is applicable to linear models with categorical predictors, an extremely versatile and widely used class of statistical models. We partition the levels of a user-chosen categorical predictor, which we call the SLGF, into two groups. These two groups may elicit a variety of effects on the observed data, including latent group-based regression effects, group-based interactions, group-based heteroscedasticity, and/or some combination of these structures. Most standard modeling approaches neglect to consider such structures or latent groups, but we demonstrate that they occur quite frequently in common datasets. Our methodology invokes several key components to detect these structures:

1. Combinatoric approach: because a practitioner will often lack prior information on the levels of the categorical predictor that should be grouped together (if they should be grouped together at all), we consider all possible latent grouping structures are considered. See Chapter 2 for further detail.

2. Bayesian model selection: we implement Bayes' Theorem, along with commonly used and effective priors on regression effects and error variances, to quantify the prob-

ability that specific structures underlie the data in question. This work provides a thorough description of the mathematical and computational approach necessary to compute these model probabilities under our flexible framework, for linear models with a categorical predictor. Thus we can implicitly conduct variable selection as well as select the most probable grouping structures simultaneously. See Chapter 2 for further detail.

3. Fractional Bayes factors; a major advantage of this proposed methodology it that it can compare homoscedastic and heteroscedastic models. An artifact of this feature is that a partial Bayes factor approach must be utilized in order to take full advantage of our objective Bayesian approach. We justify the use of a fractional Bayes factor and demonstrate its utility in our method. See Chapter 2 for further detail.

4. Simulation; we demonstrate the effectiveness of our methodology through a simulation study on several commonly used linear model structures. The study establishes that we can detect a variety of latent grouping structures underlying a dataset. See Chapter 2 for further detail.

5. User-friendly implementation; through the new R package `slgf`, we thoroughly demonstrate the process a practitioner would invoke to implement our method on a dataset. We demonstrate the functionality of `slgf` on several linear models that represent a wide range of disciplines, experimental designs, classical analysis methods, and latent grouping structures. This R package will be publically available on CRAN in the near future. See Chapter 3 for further detail.

It is our hope that this methodology becomes a staple in the analysis of common linear models for researchers in a variety of disciplines. Although that is an ambitious goal, we believe it is realistic for two reasons. First, our method is appropriate in a broad scope of layouts and fields; we illustrate our method's utility in across many datasets representing one-way analysis of variance (ANOVA), replicated and unreplicated two-way layouts, balanced

incomplete block designs, and analysis of covariance (ANCOVA), in the contexts of biology, agriculture, manufacturing, and engineering. Second, such a method will only be applied by researchers if it is straightforward and accessible to implement for non-statisticians. We believe that we accomplish this with the R package `slgf`.

The remainder of this dissertation is organized as follows. Chapter 2 provides a review of previous literature, lays the framework of the statistical model and combinatoric latent grouping approach, summarizes mathematical details on the computation of posterior model probabilities, describes the fractional Bayes factor approach and implementation, analyzes several representative data sets, and gives a simulation study demonstrating the effectiveness of our method. Chapter 3 describes the R package `slgf` that implements the methodology in practice, by reviewing the approach, providing several examples contrasting a classical approach with our proposed method, and demonstrating the functionality of the package. Chapter 4 summarizes the work and provides several avenues of future research.

Chapter 2

Suspected Latent Grouping Factor Methodology

2.1 Introduction

Linear models with categorical predictors are among the most frequently used statistical models, but oversimplification of the variance or regression effect structures can misrepresent key relationships within observed data. Figure 2.1 shows three relevant data sets. First, the left panel shows a simulated one-way ANOVA experiment, based on the statistics reported in Welch [1951]. The horizontal axis represents the levels of a treatment factor, and the vertical axis represents a continuous response. The center panel is an analysis of covariance (ANCOVA) that analyzes the breaking strength of a starch chip [Flurry, 1939]. The horizontal axis represents the chip's thickness in 10^{-4} inches, the vertical axis represents the breaking strength in grams, and the point shapes represent the plant from which the starch was derived. Finally, the right panel is an unreplicated two-way layout, measuring the genomic hybridization signal in dogs with lymphoma [Franck et al., 2013]. The horizontal axis represents whether the sample was taken from normal or tumor tissue, the vertical axis represents the intensity of the genomic hybridization signal, and the individual lines represent

six dogs studied. In each case, the levels of the categorical predictor appear to fall into one of two groups, although the group structure is unknown before data collection. The apparent group structure for the data in Figure 2.1 is represented by dark and light gray. Situations

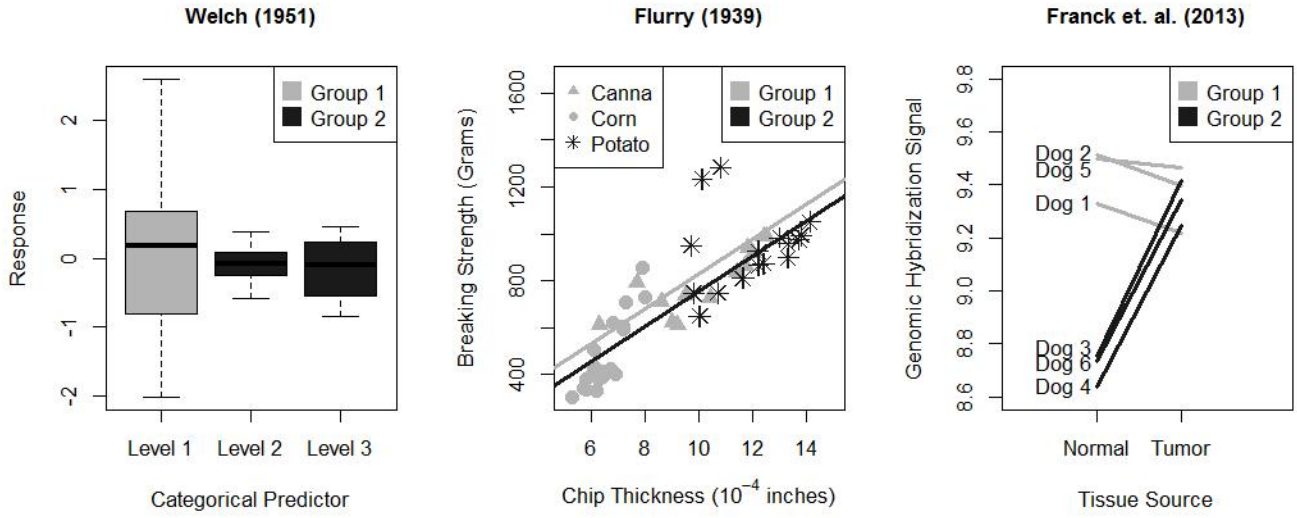

Figure 2.1: Data from Welch [1951] (left), Flurry [1939] (center), and Franck et al. [2013] (right). After a cursory examination of the data, a researcher might suspect that a latent grouping factor (emphasized by dark and light gray) underlies the levels of the categorical predictor.

similar to those illustrated in Figure 2.1 often arise in research. Perhaps after plotting the data or in reviewing previous related work, the researcher begins to suspect that there is a hidden grouping within the levels of the categorical predictor. We call this predictor the *suspected latent grouping factor*, or SLGF. This work aims to determine whether the latent groupings are plausible and how the group structure affects the response.

We must consider the SLGF with two key ideas in mind: first, that the SLGF might manifest itself through one of several *structures* in the data, namely, (i) group-specific regression effects, (ii) group-specific variances, and/or (iii) hidden interactions between groups and other model predictors. There are eight possible combinations of structures. Second, the specific assignment of levels to groups is unknown. Both of these aspects of the SLGF must be learned from the data in an unsupervised fashion: specifically, our proposed method

uses Bayesian model selection to assess the plausibility of the SLGF's impact on the data based on posterior probability. Formal detail can be found in Section 2.

Common linear model assumptions include homoscedasticity and a unique effect of the response at each level of categorical predictors; for a thorough review see Berry [1993]. Violations of these assumptions can have a wide variety of negative consequences on inference; see Rencher and Schaalje [2008], Deschamps [1991] and Scheffé [1959]. Model misspecification can lead to additional problems; see Rencher and Schaalje [2008], Rao [1971], and Deegan [1976].

Regarding the detection of heteroscedasticity, many analyses follow a two-stage approach. These include the methods proposed by Bartlett [1937], Levene [1960], Brown and Forsythe [1974], and Hartley [1950]. When heteroscedasticity is believed to be a function of a continuous predictor, many methods are available, including Breusch and Pagan [1979], Cook and Weisberg [1983], White [1980], Glejser [1969], Park [1966], Box and Hill [1974], Bickel [1978], Jobson and Fuller [1980], and Carroll and Ruppert [1982].

Regarding inference in the presence of heteroscedasticity, methods include those of Box and Cox [1964], Carroll and Ruppert [1988], Perthes [1855], Morrison [1983], Huber [1967], Eicker [1967], Cragg [1983], Hildreth and Houck [1968], Long and Ervin [2000], Polasek et al. [1998], Polasek and Pötzelberger [1994], Cuervo and Achcar [2009], Boscardin and Gelman [1994], MacKinnon and White [1985], Dumitrascu et al. [2015], and White [1980]; for a review see Hayes and Cai [2007] and Cribari-Neto and Zarkos [1999]. Many methods exist to detect heteroscedasticity or conduct inference in its presence. To our knowledge, ours is the first proposal of latent group-based heteroscedasticity alongside possibly unique regression effects and/or hidden interactions.

While an extension to more than two groups is natural, our choice of two groups is still a reasonable approach in many problems; Kharrati-Kopaei and Sadooghi-Alvandi [2007] and Franck [2018] study factor level groupings based on two groups in unreplicated two-way layouts, while Goldfeld and Quandt [1965] model heteroscedasticity as a function of two

groups, where groups are created by partitioning observations ordinally.

The main contribution of this work is to propose a method of probabilistically detecting the presence of hidden categorical level groupings, and to describe the model specifications that capture the effect of such groupings, through the use of Bayesian model selection. This work generalizes the notion of latent group-based effects from the two-way layout to linear models; see Kharrati-Kopaei and Sadooghi-Alvandi [2007], Franck et al. [2013], Franck and Osborne [2016], and Franck [2018]. There is no unified Bayesian model selection approach to account for these structures in the context of latent groupings of the levels of a categorical predictor in general. Although we illustrate our method in the contexts of the three specific settings shown in Figure 2.1, our proposal is flexible enough to be used in the context of any linear model with a categorical predictor.

The remaining structure of this paper is as follows. Section 2 describes the candidate models in the context of the SLGF, as well as the fractional Bayes factor and Bayesian model selection details. Section 3 describes our proposed Bayesian model specification in contexts of ANCOVA models and unreplicated two-way layouts. Section 4 describes a simulation study to assess the performance of our method. Section 5 applies our proposed method to the empirical data sets of Figure 2.1. Section 6 summarizes the proposed method and provides some additional comments. Additional mathematical details and simulation results for one-way ANOVA data are provided in Section 2.7.

2.2 Proposed Method

2.2.1 Specification of Linear Models with Categorical Predictors

We begin the development of our approach by elucidating the assignment of the levels of the SLGF into groups. As an illustration, consider the data analyzed by Franck et al. [2013], shown in the right panel of Figure 2.1: an unreplicated two-way layout with 6 rows

and 2 columns. We choose the dogs as the SLGF. Three examples of possible SLGF level assignments are shown in Figure 2.2: For a SLGF with levels $k = 1, \ldots, K+1$, let $\boldsymbol{k} =$

r_1	9.33	9.22
r_2	9.51	9.39
r_3	8.75	9.42
r_4	8.64	9.25
r_5	9.50	9.46
r_6	8.73	9.35
Example scheme 1		

r_1	9.33	9.22
r_2	9.51	9.40
r_3	8.75	9.42
r_4	8.64	9.25
r_5	9.50	9.46
r_6	8.73	9.34
Example scheme 2		

r_1	9.33	9.22
r_2	9.51	9.39
r_3	8.75	9.42
r_4	8.64	9.25
r_5	9.50	9.46
r_6	8.73	9.34
Example scheme 3		

Figure 2.2: Three possible grouping schemes of the data analyzed by Franck et al. [2013] (as shown in the rightmost panel of Figure 2.1) are shown here. Row membership is used to partition the data into two groups, shaded and unshaded. These grouping schemes are denoted $1, 2, 3 : 4, 5, 6$ (left), $1, 2, 4, 6 : 3, 5$ (center), and $1, 4, 5 : 2, 3, 6$ (right).

$(1, \ldots, K+1)^T$. We formally define a *grouping scheme*, denoted $d_s(\boldsymbol{k})$, as a partitioning that assigns the data into two groups based on each observation's corresponding level of the SLGF. Let $\mathcal{S}$ be the set of all possible schemes, and $s = 1, \ldots, S$ index the possible schemes. We refer to the example scheme 3, in the rightmost panel of Figure 2.2, to motivate the subsequent notational definitions; this example partitions levels $k = 1$, 4, and 5 separately from 2, 3, and 6. For a scheme $d_s(\boldsymbol{k})$, a colon separates the levels each group comprises; for example, the scheme of example 3 is denoted as $d_s(\boldsymbol{k}) = 1, 4, 5 : 2, 3, 6$. The most effective method to partition the levels of the SLGF into groups depends on the size and nature of the study in question. Many problems have K small enough that a combinatoric search over all possible grouping schemes is reasonable; see Franck et al. [2013], Franck and Osborne [2016], Franck [2018], and Kharrati-Kopaei and Sadooghi-Alvandi [2007]. We use the combinatoric search approach exclusively in this study.

Next we formalize the idea of model structures in the context of specific linear models. Recall from Section 2.1 that we must potentially accommodate eight model structures containing a mix of group-based regression effects, group-based variances, and group-based interactions. These structures must be tailored to both the researcher's suspicion and goals, as well as the data layout under consideration. Thus we propose the *model class*, which is

the set of models representing a particular structure. While structures reflect the presence or absence of group-based effects, classes prescribe specific corresponding models in the context of the data layout in question.

Now that the grouping schemes and model classes have been enumerated, consider a linear model with N centered observations $\boldsymbol{Y}$:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

with model matrix X parametrized to be full column rank, regression effects $\boldsymbol{\beta}$, and errors $\boldsymbol{\varepsilon} \overset{\text{iid}}{\sim} N(\mathbf{0}, \Sigma)$ with covariance matrix Σ. To account for the eight possible model structures previously described, we will partition $\boldsymbol{Y}$, $\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}$, and Σ. Partition $\boldsymbol{\beta}$ into four components: let α represent an intercept common to all models, let $\boldsymbol{\nu} = \{\nu_k\}_{k=1}^{K}$ represent the SLGF with $K+1$ levels, let $\boldsymbol{\tau} = \{\tau_j\}_{j=1}^{J}$ contain J other regression effects, categorical or continuous, and, let $\boldsymbol{\rho} = \{\rho_\ell\}_{\ell=1}^{L}$ represent L interactions with the SLGF. Then $\boldsymbol{\beta}_{(1+K+J+L)\times 1} = (\alpha,\ \boldsymbol{\nu}_{K\times 1},\ \boldsymbol{\tau}_{J\times 1},\ \boldsymbol{\rho}_{L\times 1})$.

Similarly, partition the model matrix $X_{N\times(1+K+J+L)} = (\mathbf{1}_{N\times 1}^T \,|W_{N\times K}\,|\,V_{N\times J}\,|\,U_{N\times L})$ into three matrices corresponding to the data related to α, $\boldsymbol{\nu}$, $\boldsymbol{\tau}$, and $\boldsymbol{\rho}$, respectively. Thus we can express (2.1) equivalently as

$$\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + U\boldsymbol{\rho} + \boldsymbol{\varepsilon}. \tag{2.2}$$

In cases where the effect structure for one of the terms in Equation (2.2) depends on a latent grouping scheme, denote that term with a tilde. For example, in structures with group-based interactions, model a group-based interaction $\tilde{\boldsymbol{\rho}}$ instead of usual the interaction $\boldsymbol{\rho}$. Similarly, heteroscedastic structures with group-based variances are modeled with error vector $\tilde{\boldsymbol{\varepsilon}}$ and corresponding covariance matrix $\tilde{\Sigma}$ instead of the homoscedastic counterparts $\boldsymbol{\varepsilon}$ and Σ.

When a model contains a group structure, we arrange the observations within $\boldsymbol{Y}$ to first contain the n_1 observations corresponding to s_1, followed by the n_2 observations corre-

sponding to s_2, denoted $\boldsymbol{Y} = (\boldsymbol{Y}_1, \boldsymbol{Y}_2)$. We similarly arrange the rows and columns of X to contain the corresponding observations, which helps concisely express our proposed forms of heteroscedasticity in Σ. For structures with distinct regression effects by group, arrange $\tilde{\boldsymbol{\nu}} = (\tilde{\boldsymbol{\nu}}_1, \tilde{\boldsymbol{\nu}}_2)$, and for classes with distinct variances by group, partition the error vector and covariance matrix such that $\tilde{\boldsymbol{\varepsilon}} \sim N(\mathbf{0}, \tilde{\Sigma})$ where $\tilde{\Sigma} = \left(\begin{array}{c|c} \sigma_1^2 I_{n_1 \times n_1} & 0_{n_1 \times n_2} \\ \hline 0_{n_2 \times n_1} & \sigma_2^2 I_{n_2 \times n_2} \end{array}\right)_{N \times N}$. Notice $N = n_1 + n_2$, for the effects corresponding to groups s_1 and s_2, respectively. With multiple schemes, classes, and structures under consideration, we next propose a Bayesian model selection approach to assess whether latent groups exist within the data, and if so, identify the appropriate grouping scheme $d_s(\boldsymbol{k}) \in \mathcal{S}$ (if present) and class $c \in \mathcal{C}$.

2.2.2 Bayesian Model Selection Details

Model Specification

Denote the set of all candidate models $\mathcal{M} = \{m_s^c\}$, indexed over all possible schemes $s = 1, ..., S$ and classes $c = 1, ..., C$, where $|\mathcal{M}| = M$; to ease the notational burden, we have denoted $d_s(\boldsymbol{k})$ as s in the subscript of m. Although each model matrix X and estimators for $\boldsymbol{\beta}$ and Σ depend on the grouping and class under consideration, for notational simplicity we do not index X, $\boldsymbol{\beta}$, or Σ by m, s, or c. Let the vector $\boldsymbol{\varphi}$ contain the single precision $\varphi := \frac{1}{\sigma^2}$ under homoscedastic models, and the corresponding subgroup precisions φ_1 and φ_2 under heteroscedastic models. Denote the precision matrix $\Phi = \Sigma^{-1}$ and $\boldsymbol{\theta} = \{\boldsymbol{\varphi}, \boldsymbol{\beta}\}$, the set of unknown model parameters; then we can express (2.1) conditionally as

$$\boldsymbol{Y} | m_s^c, \boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \Phi^{-1}) \tag{2.3}$$

for a given scheme $d_s(\boldsymbol{k}) \in \mathcal{S}$ and class $c \in \mathcal{C}$. Thus the likelihood function is given by

$$P(\boldsymbol{Y} | \boldsymbol{\theta}, m_s^c) = (2\pi)^{-\frac{N}{2}} |\Phi|^{\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(\boldsymbol{Y} - X\boldsymbol{\beta})^T \Phi (\boldsymbol{Y} - X\boldsymbol{\beta})\} \tag{2.4}$$

We consider two common prior specifications on the regression effects and precision(s). In both cases, we prefer noninformative priors on the precision(s) because prior information on precision is rarely available. First, we consider a noninformative approach, where we have

$$P(\boldsymbol{\beta}, \varphi, |m_s^c) \propto \varphi^{-1} \text{ (homoscedastic models), or} \tag{2.5}$$

$$P(\boldsymbol{\beta}, \boldsymbol{\varphi}, |m_s^c) \propto {\varphi_1}^{-1} \cdot {\varphi_2}^{-1} \text{ (heteroscedastic models)} \tag{2.6}$$

Next we consider the Zellner-Siow mixture g-prior [Zellner and Siow, 1980, Zellner, 1986, Liang et al., 2008], where

$$P(\alpha,\, \boldsymbol{\varphi}) \propto \varphi^{-1} \text{ (homoscedastic models)}, \tag{2.7}$$

$$P(\alpha,\, \boldsymbol{\varphi}) \propto \varphi_1^{-1} \cdot \varphi_2^{-1} \text{ (heteroscedastic models)}, \tag{2.8}$$

$$\boldsymbol{\beta}_{-\alpha} | \boldsymbol{\varphi}, m_s^c \sim N(\mathbf{0},\ g(X^T \Phi^{-1} X)^{-1}), \text{ and} \tag{2.9}$$

$$g \sim \text{IG}\left(\frac{1}{2},\ \frac{N}{2}\right) \tag{2.10}$$

where $\boldsymbol{\beta}_{-\alpha} := \boldsymbol{\beta} \setminus \{\alpha\}$. We use the Zellner-Siow mixture g-prior in relatively data-poor situations, such as unreplicated two-way layouts, to reduce the dimensionality of the parameter space with improper priors. This consequently lowers the minimal training sample size, a critical component of the fractional Bayes factor approach described in Section 2.2.2. We use the noninformative flat prior where data are more abundant; see Section 2.3.2 for more detail.

Model Priors for Classes and Schemes

We impose a uniform model prior by model class: $P(m^c) = \sum_{s=1}^{S} P(m_s^c) := \frac{1}{C}$. Depending on the data layout and model structures under consideration, various classes may contain different numbers of models; thus we subsequently divide each class's prior uniformly among

the models it contains. For example, in an ANOVA layout, one model class might represent the single mean model, where $P(m^c) = \frac{1}{C}$; note we do not index this model by s as there is no grouping structure present in this class. Alternatively, for a model class containing models with distinct regression effects by grouping scheme, the S individual models within the class would be given the prior $P(m_s^c) = \frac{1}{S \cdot C}$. By implementing Bayes' Theorem, posterior model probabilities approximated via the fractional Bayes factor (see Section 2.2.2) can then be easily computed for each individual model and class with the marginal probabilities of each model. For a given model $m_{c'}^{s'}$ we compute:

$$P(m_{s'}^{c'}|\boldsymbol{Y}) = \frac{P(\boldsymbol{Y}|m_{s'}^{c'})P(m_{s'}^{c'})}{\sum_{c=1}^{C}\sum_{s=1}^{S} P(\boldsymbol{Y}|m_s^c)P(m_s^c)} \tag{2.11}$$

We can then aggregate the overall probability of a given class c' as $P(m^{c'}|\boldsymbol{Y}) = \sum_{s \in \mathcal{S}} P(m_s^{c'}|\boldsymbol{Y})$, or for a given grouping scheme s' as $P(m_{s'}|\boldsymbol{Y}) = \sum_{c \in \mathcal{C}} P(m_{s'}^c|\boldsymbol{Y})$; hence the proposed method allows researchers to draw probabilistic conclusions about both structures and specific grouping schemes.

Fractional Bayes Factor Approach

We next motivate the need to use a fractional Bayes factor approach. Consider a comparison between arbitrary models m_1 and m_2, where m_1 and m_2 represent homoscedastic and heteroscedastic models, respectively, via $B^{12} = \frac{P(m_1|\boldsymbol{Y})P(m_2)}{P(m_2|\boldsymbol{Y})P(m_1)}$. For m_1 we use $P(\boldsymbol{\varphi}) = a \cdot \varphi^{-1}$, and for m_2 we use $P(\boldsymbol{\varphi}) = a' \cdot \varphi_1^{-1} \cdot \varphi_2^{-1}$ for arbitrary constants $a \neq a'$. So the Bayes factor to compare m_1 and m_2 is

$$B^{12} = \frac{a \cdot \int P(\boldsymbol{Y}|\boldsymbol{\beta}, \Phi, m_1)P(\boldsymbol{\beta})P(\boldsymbol{\varphi})d\boldsymbol{\beta}d\boldsymbol{\varphi} \cdot P(m_2)}{a' \cdot \int P(\boldsymbol{Y}|\boldsymbol{\beta}, \Phi, m_2)P(\boldsymbol{\beta})P(\boldsymbol{\varphi})d\boldsymbol{\beta}d\boldsymbol{\varphi} \cdot P(m_1)} \tag{2.12}$$

which is defined only up to the arbitrary constant $\frac{a}{a'}$ and thus inappropriate for use in model comparison. Note this problem arises from the use of noninformative priors on the precisions

when comparing homoscedastic and heteroscedastic models. Fractional Bayes factors were developed to elicit a cancellation of this constant, rendering a Bayes factor that is well-defined [O'Hagan, 1995].

For a description of fractional Bayes factors (FBFs), see O'Hagan [1995] and **?**. We use the FBF approach of O'Hagan rather than the intrinsic Bayes factor of Berger and Pericchi [1996], as the need to choose a training sample would be complicated by the potential scarcity of data induced by some clustering schemes.

To fully quantify the fractional marginal probability of a given model m_i, we must compute both $\int P(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$ and $\int P^b(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$ for some user-chosen fractional exponent b. O'Hagan [1995] provides several recommendations for b, including $b := \frac{m_0}{N}$, where m_0 is the minimal training sample size necessary for $P^b(\boldsymbol{Y}|m_i)$ to be proper. In this study we choose $b = \frac{m_0}{N}$, which elicits consistent model selection [O'Hagan, 1995]. In many cases, tractable expressions exist for both $\int P(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $\int P^b(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$. When these integrals are intractable, we use the Laplace approximation in the computation of both $\int P(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $\int P^b(\boldsymbol{Y}|\boldsymbol{\theta}_i, m_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$, as cubature-based approximation is computationally expensive and our study of the integrand surface indicates a suitable shape. See Section 2.7 for justification and further detail.

2.3 Applications

2.3.1 Application 1: ANCOVA Models

We first consider an ANCOVA scenario with continuous effect $\boldsymbol{\tau}$ and where the SLGF with $K+1$ levels corresponds to the single categorical predictor effect $\boldsymbol{\nu}$. We let $V_{N\times 1}$ contain the observed continuous covariate and $W_{N\times K}$ be the appropriate categorical effect design matrix. We consider models with and without the interaction effect $\boldsymbol{\rho}$, which governs whether the

linear trends share a common slope. Thus we begin with the model given by

$$\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + U\boldsymbol{\rho} + \boldsymbol{\varepsilon} \tag{2.13}$$

yielding the likelihood function

$$P(\boldsymbol{Y}|m_s^c, \alpha, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\rho}, \boldsymbol{\varphi}) = (2\pi)^{-\frac{N}{2}}|\Phi|^{\frac{1}{2}}\cdot\exp\{-\frac{1}{2}(\boldsymbol{Y}-\mathbf{1}^T\alpha-W\boldsymbol{\nu}-V\boldsymbol{\tau}-U\boldsymbol{\rho})^T\Phi(\boldsymbol{Y}-\mathbf{1}^T\alpha-W\boldsymbol{\nu}-V\boldsymbol{\tau}-U\boldsymbol{\rho})\} \tag{2.14}$$

To consider the model without an interaction, let $\boldsymbol{\rho} := \mathbf{0}$. For models with a group effect, let $\boldsymbol{\nu} := \tilde{\boldsymbol{\nu}}$; for models with a group-by-continuous predictor interaction, let $\boldsymbol{\rho} := \tilde{\boldsymbol{\rho}}$; and finally for heteroscedastic models, let $\boldsymbol{\varepsilon} := \tilde{\boldsymbol{\varepsilon}}$. As an illustration, we consider eight distinct model classes.

1. Class I (m^{I}): the "null" model, with no categorical or continuous covariate effects and homoscedastic error variance, contains 1 model with no grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

2. Class II (m^{II}): the "simple linear regression (SLR)" model, with a continuous covariate effect only and homoscedastic error variance, contains 1 model with no grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + V\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

3. Class III (m^{III}): the "ANCOVA" model with categorical and continuous covariate effects and homoscedastic error variance, contains 1 model with no grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

4. Class IV (m_s^{IV}): the "group-contracted ANCOVA" model with group and continuous covariate effects and homoscedastic error variance, contains $2^{K-1} - 1$ schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\tilde{\boldsymbol{\nu}} + V\boldsymbol{\tau} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

5. Class V (m^{V}): the "interaction ANCOVA" model with categorical and continuous covariate effects, level-based interaction, and homoscedastic error variance, contains 1 model with no grouping scheme; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + U\boldsymbol{\rho} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

6. Class VI (m_s^{VI}): the "group-interaction" model, with group and continuous covariate effects, group-based interaction, and homoscedastic error variance, contains $2^{K-1}-1$ schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\tilde{\boldsymbol{\nu}} + V\boldsymbol{\tau} + U\tilde{\boldsymbol{\rho}} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0},\, \sigma^2 I)$

7. Class VII (m_s^{VII}): the "heteroscedastic group-contracted" model (heteroscedastic Class IV), with group and continuous covariate effects and heteroscedastic error variance, contains $2^{K-1}-1$ schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\tilde{\boldsymbol{\nu}} + V\boldsymbol{\tau} + \tilde{\boldsymbol{\varepsilon}},\ \tilde{\boldsymbol{\varepsilon}} \sim N\left(\mathbf{0},\, \tilde{\Sigma} = \left[\begin{array}{c|c} \sigma_1^2 I_{n_1\times n_1} & 0_{n_1\times n_2} \\ \hline 0_{n_2\times n_1} & \sigma_2^2 I_{n_2\times n_2} \end{array}\right]_{N\times N}\right)$

8. Class VIII (m_s^{VIII}): the "heteroscedastic group-interaction" model (heteroscedastic Class VI), with group and continuous covariate effects, group-based interaction, and heteroscedastic error variance, contains $2^{K-1}-1$ schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\tilde{\boldsymbol{\nu}} + V\boldsymbol{\tau} + U\tilde{\boldsymbol{\rho}} + \tilde{\boldsymbol{\varepsilon}},\ \tilde{\boldsymbol{\varepsilon}} \sim N\left(\mathbf{0},\, \tilde{\Sigma} = \left[\begin{array}{c|c} \sigma_1^2 I_{n_1\times n_1} & 0_{n_1\times n_2} \\ \hline 0_{n_2\times n_1} & \sigma_2^2 I_{n_2\times n_2} \end{array}\right]_{N\times N}\right)$

Thus the structure of no group-based effects is represented by Classes I, II, III, and V; group-based regression effects are represented by Classes IV, VI, VII, and VIII; group-based variances are represented by Classes VII and VIII.

We use the FBF approach outlined in Section 2.2.2. Typically, the most complex model considered in an ANCOVA study will have a small minimal training sample size relative to the overall sample size. Thus we choose noninformative priors on the regression effects and precision(s) as described in Equations (2.5) and (2.6), so the marginal density of the data conditional on the model is

$$P(\boldsymbol{Y}|m_s^c) = \iint P(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\varphi}, m_s^c) P(\boldsymbol{\beta}) P(\boldsymbol{\varphi}) d\boldsymbol{\beta} d\boldsymbol{\varphi} \tag{2.15}$$

This marginal density is analytically integrable over all of the homoscedastic classes defined previously. In heteroscedastic cases (classes VII and VIII), we use a Laplace approximation over the log-variances to approximate $P(\boldsymbol{Y}|m_s^{\text{VII}})$ and $P(\boldsymbol{Y}|m_s^{\text{VIII}})$. We demonstrate the performance of this approach through a simulation study in Section 2.4.1 and through empirical data sets in Section 2.5.1.

2.3.2 Application 2: Unreplicated Two-Way Layouts

Next we consider an unreplicated two-way layout with R rows, C columns, and $N = R \times C$ observations. Because of the unreplicated nature of such a design, the full set of standard interaction effects cannot be incorporated due to insufficient degrees of freedom. Treat the row effects as the SLGF $\boldsymbol{\nu}$, and let $\boldsymbol{\tau}$ contain the column effects; note by transposing the data table we could treat the column effects as the SLGF as well.

Our choice of model classes is based on the idea of hidden additivity, where interactions are treated as a group-by-column effect [Franck et al., 2013]. The usual "additive" model $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + \boldsymbol{\varepsilon}$ accounts for only row and column main effects. The group-based model, partitioned by levels of the row effect, does not include column effects but does consider group-by-column interactions, denoted by $\tilde{\boldsymbol{\rho}}$. We require at least 2 levels of k (rows) in both groups to ensure there are enough degrees of freedom to estimate this group-by-column interaction. We thus begin with the model

$$\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + U\tilde{\boldsymbol{\rho}} + \boldsymbol{\varepsilon} \tag{2.16}$$

where we let $\tilde{\boldsymbol{\rho}} := \mathbf{0}$ in the additive model, $\boldsymbol{\tau} := \mathbf{0}$ in cases with scheme-based regression effects, and $\boldsymbol{\varepsilon} := \tilde{\boldsymbol{\varepsilon}}$ in cases with group-based heteroscedasticity. So the full likelihood function is given by

$$P(\boldsymbol{Y}|m_s^c, \alpha, \boldsymbol{\nu}, \boldsymbol{\tau}, \tilde{\boldsymbol{\rho}}, \boldsymbol{\varphi}) = (2\pi)^{-\frac{N}{2}}|\Phi|^{\frac{1}{2}}\exp\{-\frac{1}{2}(Y-\mathbf{1}\alpha-W\boldsymbol{\nu}-V\boldsymbol{\tau}-U\tilde{\boldsymbol{\rho}})^T\Phi(Y-\mathbf{1}\alpha-W\boldsymbol{\nu}-V\boldsymbol{\tau}-U\tilde{\boldsymbol{\rho}})\} \tag{2.17}$$

In this layout we consider four model classes:

1. Class I (m^{I}): the "additive" model, where column effects are equivalent across rows and error variance is constant across all observations, contains 1 model with no grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

2. Class II (m_s^{II}): the "group-by-column interaction" model, where levels are divided into

two latent groups based on grouping scheme $d_s(\boldsymbol{k})$ with distinct means and equivalent variance, contains $2^{K-1} - K - 1$ grouping schemes [Franck et al., 2013]; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + U\tilde{\boldsymbol{\rho}} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$

3. Class III (m_s^{III}): the "heteroscedastic additive" model, where levels are divided into two latent groups based on grouping scheme $d_s(\boldsymbol{k})$ with equivalent means and group-based variances, contains $2^{K-1} - K - 1$ grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + \tilde{\boldsymbol{\varepsilon}},\ \tilde{\boldsymbol{\varepsilon}} \sim N\left(\mathbf{0}, \tilde{\Sigma} = \left[\begin{array}{c|c} \sigma_1^2 I_{n_1\times n_1} & 0_{n_1\times n_2} \\ \hline 0_{n_2\times n_1} & \sigma_2^2 I_{n_2\times n_2} \end{array}\right]_{N\times N}\right)$

4. Class IV (m_s^{IV}): the "heteroscedastic group-by-column interaction" model where levels are divided into two latent groups based on grouping scheme $d_s(\boldsymbol{k})$ with distinct means and group-based variances, contains $2^{K-1} - K - 1$ grouping schemes; $\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + U\tilde{\boldsymbol{\rho}} + \tilde{\boldsymbol{\varepsilon}},\ \tilde{\boldsymbol{\varepsilon}} \sim N\left(\mathbf{0}, \tilde{\Sigma} = \left[\begin{array}{c|c} \sigma_1^2 I_{n_1\times n_1} & 0_{n_1\times n_2} \\ \hline 0_{n_2\times n_1} & \sigma_2^2 I_{n_2\times n_2} \end{array}\right]_{N\times N}\right)$

Thus there are $M = (3 \times 2^{K-1}) - 3K - 2$ models considered.

Bayesian Model Specification: Unreplicated Two-Way Layouts

The two-way unreplicated layout is typically modeled with a high ratio of parameters to data points. Thus we must take care in choosing priors that allow us to successfully incorporate the FBF approach. With noninformative priors on the regression effects, the minimal training sample size needed to estimate $R-1$ regression effects, $2 \cdot (C-1)$ group-by-column interactions, and 2 error variances would be prohibitively large in relation to the sample size N; indeed, $b = \frac{m_0}{N} < \frac{1}{2}$ only when $R \geq 7$. In this work, we propose using the Zellner-Siow mixture g-prior on regression coefficients to reduce the dimensionality of the improper prior. Our use of this automatic prior on the regression effects lowers the minimal training sample size to $m_0 = 3$, and thus the fractional exponent $b = \frac{m_0}{N}$, to a value that allows us to

successfully implement the FBF. Thus we let

$$P(\alpha) \propto 1, \tag{2.18}$$

$$P(\boldsymbol{\beta}|\boldsymbol{\varphi},\, g) = N(\mathbf{0}, g(X^T\Phi X)^{-1}), \tag{2.19}$$

$$P(\boldsymbol{\varphi}) \propto \varphi^{-1} \text{ (for models with homoscedasticity)}, \tag{2.20}$$

$$P(\boldsymbol{\varphi}) \propto \varphi_1^{-1}\varphi_2^{-1} \text{ (for models with heteroscedasticity), and} \tag{2.21}$$

$$P(g) = \text{IG}\left(\frac{1}{2},\ \frac{N}{2}\right) \tag{2.22}$$

The marginal density of the data conditional on the model is

$$P(\boldsymbol{Y}|m_s^c) = \iiiint P(\boldsymbol{Y}|\alpha, \boldsymbol{\beta}, \boldsymbol{\varphi}, m_s^c)P(\alpha)P(\boldsymbol{\beta}|\boldsymbol{\varphi}, g)P(\boldsymbol{\varphi})P(g)d\alpha d\boldsymbol{\beta} d\boldsymbol{\varphi} dg \tag{2.23}$$

It is well-known in the homoscedastic case that the Zellner-Siow mixture g-prior advantageously elicits a Cauchy distribution on the regression effects [Liang et al., 2008]. We show the Cauchy result also holds in the heteroscedastic case; that is:

$\boldsymbol{\beta} \sim \text{MVCauchy}_p\left(\text{location} = 0,\ \text{scale} = \left(\frac{X^T\Phi X}{n}\right)^{-1}\right)$; see Section 2.7 for proof.

In classes with homoscedasticity, this intergral is intractable over g and thus a Laplace approximation is conducted over a single dimension; in heteroscedastic cases, a three-dimensional Laplace approximation is used to integrate g, $\lambda_1 = \log(\varphi_1)$, and $\lambda_2 = \log(\varphi_2)$.

2.4 Simulation Studies

2.4.1 Simulation Study: ANCOVA Models

In order to simulate ANCOVA data, we generated independent draws x uniformly over the interval $(0, 10)$. Outcomes $\boldsymbol{Y}$ were then simulated according to each of the eight classes

Class	Parameters
I (Null)	$\sigma^2 = 1$
II (SLR)	$\tau = 0.5$, $\sigma^2 = 1$
III (ANCOVA)	$\alpha = 2.0$, $\nu = (4.0, 6.0, 8.0)$, $\tau = 0.5$, $\sigma^2 = 1$
IV (Group-Contracted ANCOVA)	$\alpha = 0.0$, $\tilde{\nu} = (3.0)$, $\tau = 0.5$, $\sigma^2 = 1$
V (Interaction ANCOVA)	$\alpha = 0.5$, $\nu = (1.0, 1.5, 2.0)$, $\rho = (0.25, 0.5, 0.75, 1.0)$, $\tau = 0.5$, $\sigma^2 = 1$
VI (Group-Interaction ANCOVA)	$\alpha = 0.0$, $\tilde{\nu} = (0.8)$, $\tilde{\rho} = (0, 1)$, $\tau = 1$, $\sigma^2 = 1$
VII (Heteroscedastic Group-Contracted ANCOVA)	$\alpha = 0.0$, $\tilde{\nu} = (3.0)$, $\tau = 0.5$, $\sigma_1^2 = 1$, $\sigma_2^2 = 5$
VIII (Heteroscedastic Group-Interaction ANCOVA)	$\alpha = 0.0$, $\tilde{\nu} = (3.0)$, $\tau = 0.5$, $\tilde{\rho} = (0, 1)$, $\sigma_1^2 = 1$, $\sigma_2^2 = 5$

Table 2.1: Settings for the eight model classes in the ANCOVA simulation study where $\alpha = 0$, $K = 4$, and $N = 360$.

described in Section 2.3.1 with continuous, categorical, interaction, and/or group-based effects, as well as errors with group-based heteroscedasticity, as appropriate. In this study, we simulated $n_k = 90$ observations at each level of the SLGF; another simulation study with $n_k = 10$ is provided in Section 2.7. We note that the settings in Table 2.1 are not calibrated to be equivalent in the total effect between classes, where such a calibration would be non-trivial. For example, the Class V model contains eight non-null parameters compared to the Class I model's single parameter; thus we avoid comparing overall performance between classes in this study.

Figure 2.8 shows that overall our method accurately attributes posterior model probability to the correct model class for the eight classes described in Section 2.3.1. For Class I (null model) data, the posterior probability of the true class is high relative to the other classes, indicating that our method neither misattributes noise to a group-based structure, nor appears to systematically favor one of the other model classes when it does select the wrong model. Class II (SLR model) data performs similarly, but Classes IV (group-contracted ANCOVA) and VII (group-interaction) capture some posterior probability as well. The Class III (ANCOVA) data setting is favored by the correct class, with Class V (interaction ANCOVA) being the second choice. Class IV (group-contracted ANCOVA) data appears to be the most difficult to detect based on the parameters chosen for this study; although its

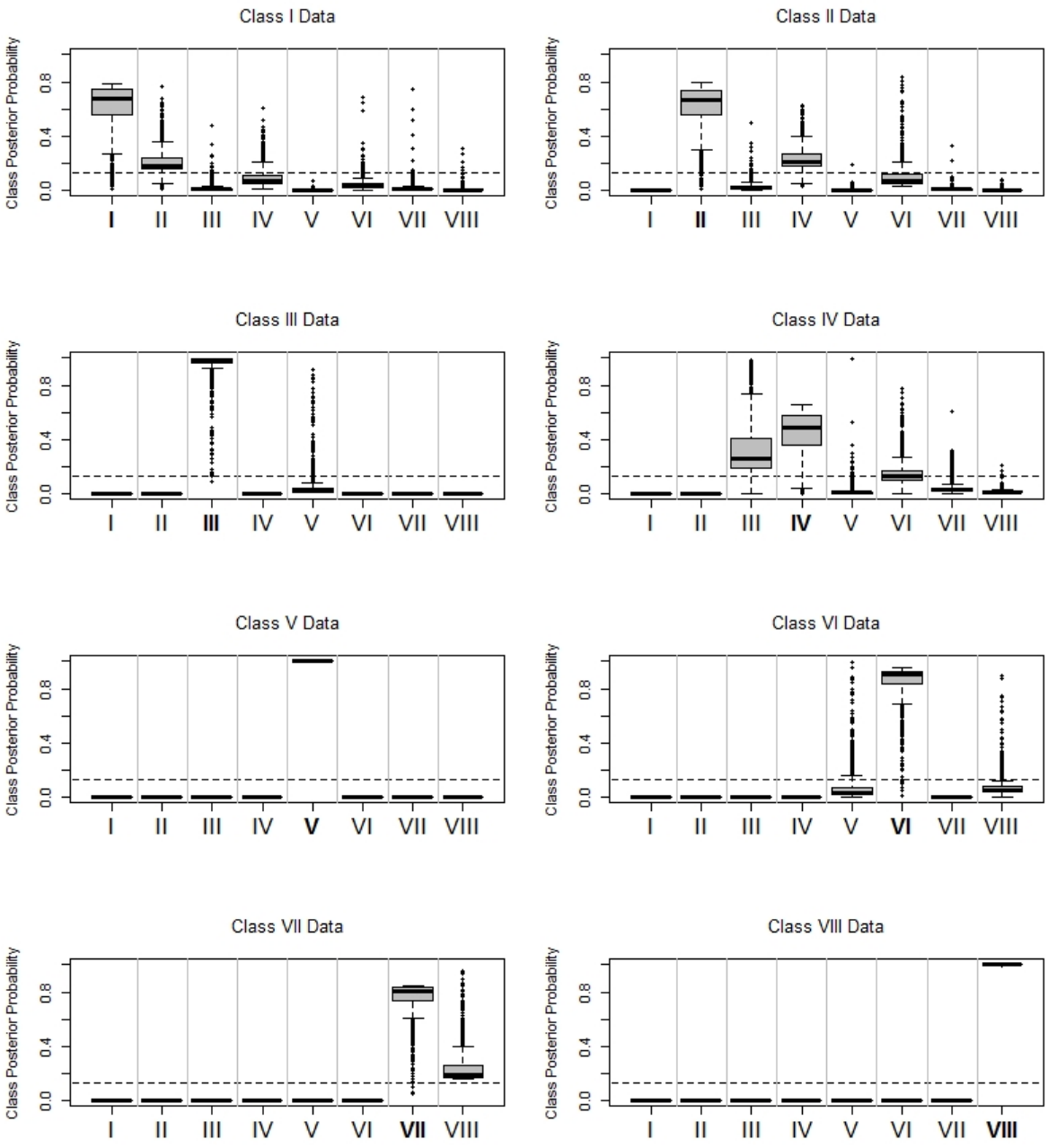

Figure 2.3: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo data sets with $K = 4$ levels of the categorical predictor, each with 90 observations, for a total of $N = 360$ observations. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

posterior probabilities are generally higher than the other classes', it is often mistaken for Class III data, meaning in this case a group-based categorical effect is often mistaken for the full categorical effect. Note that these misattributed classes differ by only two parameters in the regression structure. For Class V (interaction ANCOVA) data, the correct model class is favored. Class VI (group-interaction) data is also correctly favored the majority of the time; when misclassified, it is usually chosen as Class V (interaction ANCOVA) or VIII (heteroscedastic ANCOVA). Classes VII (heteroscedastic group-contracted) and VIII (heteroscedastic group-interaction) perform well; when wrong, Class VII tends to favor a spurious interaction effect.

These results indicate that, in general, our method tends to identify the correct class. When no group-based structure is present, our method tends to not erroneously fit spurious effects or variances. Thus group-based regression effects, interactions, and/or variances are detected with high accuracy when present. We see similar performance with the additional study provided in Section 2.7 where $N = 40$.

2.4.2 Simulation Study: Unreplicated Two-Way Layouts

Row and column effects along with error variance(s) (provided in Table 2.2) were simulated to generate unreplicated two-way layouts. We consider layouts of size 10×5 where $N = 50$. Another setting with a smaller effect size, and a study on layouts of size 5×5, are given in Section 2.7.

Figure 2.4 shows that our method also tends to favor the true class in the two-way unreplicated layout, with high parameter to data ratio and the mixture g-prior. For additive data from Class I, the true class is generally favored; the most probable alternative is typically the heteroscedastic additive Class III. Under large effect size, Class II (group-by-column interaction) data is correctly identified. Class III (heteroscedastic additive) and Class IV

Class	Parameters
I (Additive Model)	$\alpha = 1$, $\boldsymbol{\nu} \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma^2 = 1$
II (Group-by-Column Interaction)	$\alpha = 1$, $\boldsymbol{\nu} \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.8, 2.6, 3.4, 4.2\}$, $\boldsymbol{\tau}_2 \in \{4.2, 3.4, 2.6, 1.8, 1.0\}$, $\sigma^2 = 1$
III (Heteroscedastic Additive)	$\alpha = 1$, $\boldsymbol{\nu} \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.10$
IV (Heteroscedastic Group-by-Column Interaction)	$\alpha = 1$, $\boldsymbol{\nu} \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.8, 2.6, 3.4, 4.2\}$, $\boldsymbol{\tau}_2 \in \{4.2, 3.4, 2.6, 1.8, 1.0\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.10$

Table 2.2: Settings for the four model classes in the two-way unreplicated layout simulation study with 10×5 layouts with larger effect size.

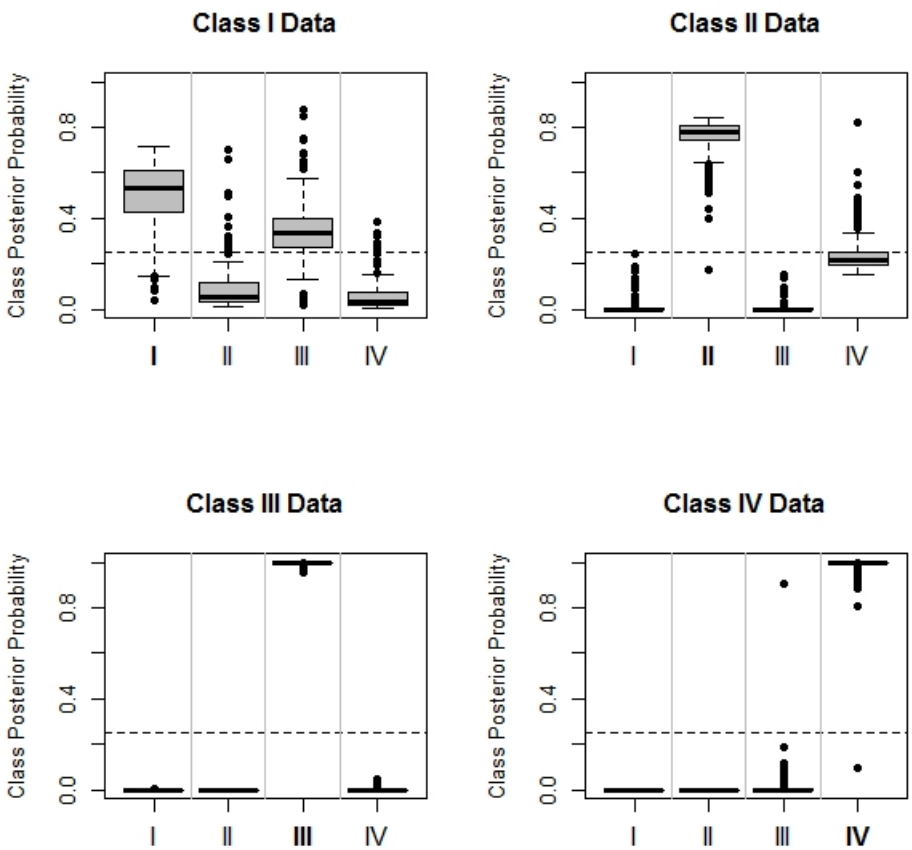

Figure 2.4: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo 10×5 layouts. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

(heteroscedastic group-by-column interaction) data are both favored correctly as well regarding both effect sizes.

2.5 Case Studies

2.5.1 Case Study: ANCOVA

We revisit the Flurry [1939] data, shown in the center panel of Figure 2.1 and in Figure 2.5. The breaking strength of a chip coated with a film derived from one of three plant materials is studied as a function of a continuous predictor (the thickness of the film) and a categorical predictor (the plant type from which the film was developed). The plot seems to indicate some degree of heteroscedasticity between canna and corn, versus potato; results are shown in Figure 2.5. Two models receive 93% of the posterior probability: the heteroscedastic group-contracted model, with $P(m_{1,2:3}^{\mathrm{VII}}|\boldsymbol{Y}) \approx 0.51$, and the heteroscedastic group-interaction model, with $P(m_{1,2:3}^{\mathrm{VIII}}|\boldsymbol{Y}) \approx 0.42$.

2.5.2 Case Study: Two-Way Unreplicated Layouts

Franck et al. [2013] and Franck and Osborne [2016] examine a two-way unreplicated layout describing the copy number variation for genes in six dogs with lymphoma. Samples were taken from healthy and diseased tissue within each dog. It is clear in the plot that dogs behave differently by group: dogs 1, 2, and 5 appear to behave distinctly from dogs 3, 4, and 6. With six subjects by row, there are $1 + 3 \cdot (2^{6-1} - 6 - 1) = 76$ candidate models, including the null model and three classes each containing 25 models. Thus we show only the six most probable models, which account for 98.1% of the posterior probability; these results are shown in Figure 2.6. We conclude with high probability that subjects 1, 2, and 5 behave distinctly from 3, 4, and 6; this scheme over all classes has probability $P(m_{1,2,5:3,4,6}|\boldsymbol{Y}) \approx .9680$. More specifically, we also conclude homoscedastic behavior along with this grouping scheme, the

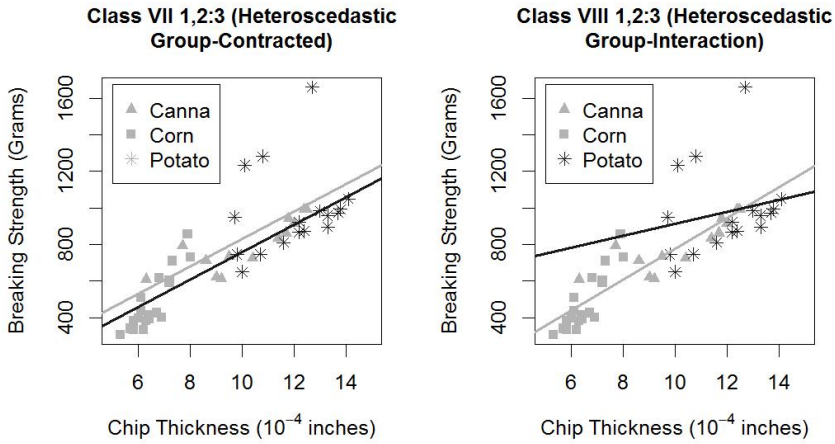

Figure 2.5: The most probable models from the Flurry [1939] data set are plotted above: the heteroscedastic group-contracted model (left), with $P(m^{\text{VII}}_{1,2:3}|\boldsymbol{Y}) \approx 0.51$, and the heteroscedastic group-interaction model (right), with $P(m^{\text{VIII}}_{1,2:3}|\boldsymbol{Y}) \approx 0.42$. Overall the grouping scheme where canna and corn are grouped together accounts for about 93.8% of the posterior model probability, while heteroscedastic models account for about 93.0% of the posterior model probability.

group-by-column interaction model, with $P(m_{1,2,5:3,4,6}|\boldsymbol{Y}) \approx .9040$.

2.6 Discussion

Our proposed method is a flexible and intuitive approach to accommodate linear models with latent group-based effects underlying the data. This method generalizes the homoscedastic, two-way unreplicated layout work of Franck [2018] to a heteroscedastic approach to any linear model with a categorical predictor. By partitioning the data according to the levels of a categorical predictor, we can detect latent structures within the data that might influence the regression effects, error variance, or interaction effects. Often, these group-based structures have straightforward and intuitive interpretations in the context of the data, making the approach particularly useful to domain experts. The use of fractional Bayes factors allows us to compare homoscedastic and heteroscedastic models with minimal prior influence. Regarding priors on the regression effects, we explored the performance of noninformative and

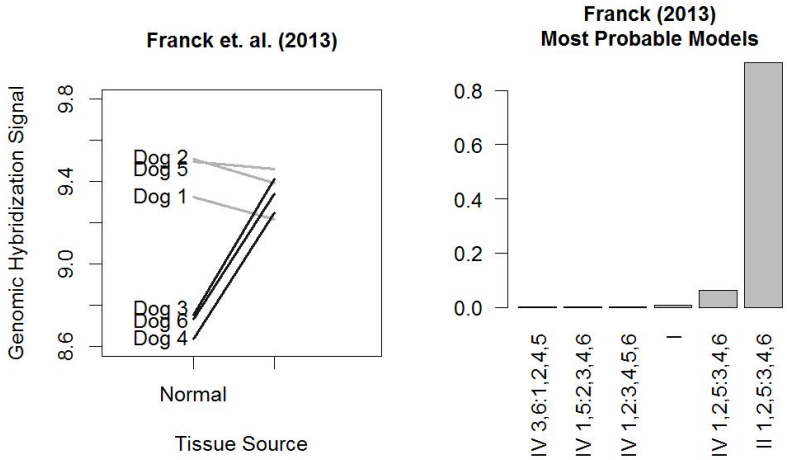

Figure 2.6: A two-way unreplicated layout with apparent group-based regression effects [Franck et al., 2013], where $P(m^{\mathrm{II}}_{1,2,5:3,4,6}|\boldsymbol{Y}) \approx 0.904$ and $P(m^{\mathrm{IV}}_{1,2,5:3,4,6}|\boldsymbol{Y}) \approx 0.064$. Overall the grouping scheme $1,2,5:3,4,6$ accounts for approximately 96.8% of the posterior model probability.

mixture g-priors in the contexts of ANCOVA and two-way unreplicated layouts, but many other choices on priors and data layouts can be considered. We considered cases in which the number of levels of the SLGF K was relatively small, so that a combinatoric search over all possible grouping schemes was computationally feasible. In cases where K is large, a Markov-chain model composition (MC3) could be used to search the model space.

In some cases, our method leads to a contraction of the number of effects modeled. For instance, in the ANCOVA examples illustrated in Section 2.3.1, estimating a group effect $\tilde{\boldsymbol{\nu}}$ as opposed to a categorical effect $\boldsymbol{\nu}$ will reduce the degrees of freedom used to estimate the effect from K to 2. Alternatively, in the two-way unreplicated layout examples, modeling a group-by-column interaction rather than column effects will expand the degrees of freedom used from $C-1$ to $2 \cdot (C-1)$.

While we have illustrated our method using two latent groups, an extension to three or more groups is straightforward conceptually but increases the complexity of the models, the number of model classes, and the number of models to consider. Fortunately, the two-group assumption has been shown to lead to useful inferences in several previous works, including

Kharrati-Kopaei and Sadooghi-Alvandi [2007], Franck et al. [2013], Franck and Osborne [2016], and Franck [2018].

2.7 Supplement

2.7.1 Marginal Model Probability Calculations

To prevent numeric underflow we consider model probabilites on the logarithmic scale in all cases. Consider the set of all log-marginal probabilities $\mathcal{L} = \{\log P(Y|m_s^c)\}$. Let $\ell^\star = \max \mathcal{L}$ and $m^\star = \arg\max_m \mathcal{L}$. Transform $\mathcal{L}^\star = \mathcal{L} - \ell^\star$ to obtain the set $\exp[\mathcal{L}^\star] = \{\exp[\log P(Y|m_s^c) - \log P(Y|m^\star)]\} = \{\frac{P(Y|m_s^c)}{P(Y|m^\star)}\} = \{B^{m_s^c,\, m^\star}\}$, representing Bayes factors for each model relative to the most probable model. Note the untransformed model probabilities are given by $P = \{\frac{P(Y|m_s^c)P(m_s^c)}{\sum P(Y|m_s^c)P(m_s^c)}\} = \{\frac{P(Y|m_s^c)P(m_s^c)/P(Y|m^\star)}{\sum P(Y|m_s^c)P(m_s^c)/P(Y|m^\star)}\} = \{\frac{B^{m_s^c,\, m^\star}P(m_s^c)}{\sum B^{m_s^c,\, m^\star}P(m_s^c)}\}$; thus we can easily use these transformed log-marginal probabilities to obtain posterior model probabilities. Our Laplace approximations generally provided values comparable to quadrature and cubature-based approximations at less computational expense.

2.7.2 Derivation of Model Probabilities: Noninformative Regression Effect Priors

Let $H_X := X(X^TX)^{-1}X^T$.

Homoscedastic Error Variance, No Group-Based Regression Effects

$$
\begin{aligned}
P(\boldsymbol{Y}|m) &= \int_{\Theta} \left[(2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(\boldsymbol{Y} - X\boldsymbol{\beta})^T \Sigma^{-1} (\boldsymbol{Y} - X\boldsymbol{\beta})\} \right]^b P(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} |\Sigma|^{-\frac{b}{2}} \exp\{-b \cdot \frac{1}{2}[\boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} - 2\boldsymbol{\beta}^T X^T \Sigma^{-1} \boldsymbol{Y} + \boldsymbol{\beta}^T X^T \Sigma^{-1} X \boldsymbol{\beta}]\} \Sigma^{-1} d\boldsymbol{\beta} d\gamma \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} \exp\{-b \cdot \frac{1}{2}[\boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} - 2\boldsymbol{\beta}^T X^T \Sigma^{-1} \boldsymbol{Y} + \boldsymbol{\beta}^T X^T \Sigma^{-1} X \boldsymbol{\beta}]\} d\boldsymbol{\beta} d\gamma \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - 2\boldsymbol{\beta}^T X^T \boldsymbol{Y} + \boldsymbol{Y}^T H \boldsymbol{Y}]\} \times \\
&\qquad \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{Y}^T \boldsymbol{Y} - \boldsymbol{Y}^T H \boldsymbol{Y}]\} d\boldsymbol{\beta} d\gamma \\
&= \int_0^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} (2\pi)^{+\frac{p}{2}} |b^{-1}\gamma^{-1}(X^T X)^{-1}|^{+\frac{1}{2}} \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{Y}^T (I - H) \boldsymbol{Y}]\} d\gamma \\
&= (2\pi)^{-\frac{Nb-P}{2}} b^{-\frac{P}{2}} |X^T X|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) \left(\frac{b \cdot \mathrm{SSResid}^{\mathrm{I}}}{2}\right)^{-\frac{Nb-P}{2}} \\
&= \pi^{-\frac{Nb-P}{2}} b^{-\frac{Nb}{2}} |X^T X|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) (\mathrm{SSResid}^{\mathrm{I}})^{-\frac{Nb-P}{2}}. \ \square
\end{aligned}
$$

$$
\begin{aligned}
q^b(\boldsymbol{Y}|m) &= \frac{\int_{\Theta} P(Y|m^{\mathrm{I}}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\Theta} P^b(Y|m^{\mathrm{I}}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\
&= \frac{\pi^{-\frac{N-P}{2}} |X^T X|^{-\frac{1}{2}} \Gamma\left(\frac{N-P}{2}\right) (\mathrm{SSResid}^{\mathrm{I}})^{-\frac{N-P}{2}}.}{\pi^{-\frac{Nb-P}{2}} b^{-\frac{Nb}{2}} |X^T X|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) (\mathrm{SSResid}^{\mathrm{I}})^{-\frac{Nb-P}{2}}} \\
&= \pi^{-\frac{N(1-b)}{2}} b^{\frac{Nb}{2}} (\mathrm{SSResid}^{\mathrm{I}})^{-\frac{N(1-b)}{2}} \frac{\Gamma\left(\frac{N-P}{2}\right)}{\Gamma\left(\frac{Nb-P}{2}\right)}. \ \square
\end{aligned}
$$

Homoscedastic Error Variance, Group-Based Regression Effects

$$
\begin{aligned}
P(\boldsymbol{Y}|m_s) &= \int_{\Theta} \left[(2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(\boldsymbol{Y} - X\boldsymbol{\beta})^T \Sigma^{-1} (\boldsymbol{Y} - X\boldsymbol{\beta})\} \right]^b P(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} \exp\{-b \cdot \frac{1}{2}[\boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} - 2\boldsymbol{\beta}^T X^T \Sigma^{-1} \boldsymbol{Y} + \boldsymbol{\beta}^T X^T \Sigma^{-1} X \boldsymbol{\beta}^T]\} d\boldsymbol{\beta}_1^T d\boldsymbol{\beta}_2^T d\gamma \\
&= \int_0^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{\beta}^T X^T X \boldsymbol{\beta} - 2\boldsymbol{\beta}^T X^T \boldsymbol{Y} + \boldsymbol{Y}^T H \boldsymbol{Y}]\} \times \\
&\qquad \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{Y}^T \boldsymbol{Y} - \boldsymbol{Y}^T H \boldsymbol{Y}]\} d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_2 d\gamma \\
&= \int_0^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{Nb}{2}-1} (2\pi)^{+\frac{P}{2}} |b^{-1}\gamma^{-1}(X_1 X_1)^{-1}|^{+\frac{1}{2}} |b^{-1}\gamma^{-1}(X_2 X_2)^{-1}|^{+\frac{1}{2}} \times \\
&\qquad \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{Y}_1^T (I - H_1) \boldsymbol{Y}_1 + \boldsymbol{Y}_2^T (I - H_2) \boldsymbol{Y}_2]\} d\gamma \\
&= (2\pi)^{-\frac{Nb-P}{2}} b^{-\frac{P}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) \left(\frac{b \cdot \left[\mathrm{SSResid}_1^{\mathrm{II}} + \mathrm{SSResid}_2^{\mathrm{II}}\right]}{2}\right)^{-\frac{Nb-P}{2}} \\
&= \pi^{-\frac{Nb-P}{2}} b^{-\frac{Nb}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) (\mathrm{SSResid}_1^{\mathrm{II}} + \mathrm{SSResid}_2^{\mathrm{II}})^{-\frac{Nb-P}{2}}. \ \square
\end{aligned}
$$

$$
\begin{aligned}
q^b(\boldsymbol{Y}|m_s) &= \frac{\int_{\Theta} P(Y|m_s^{\mathrm{II}}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int_{\Theta} P^b(Y|m_s^{\mathrm{II}}, \boldsymbol{\theta}) P(\boldsymbol{\theta}) d\boldsymbol{\theta}} \\
&= \frac{\pi^{-\frac{N-P}{2}} b^{-\frac{N}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{N-P}{2}\right) (\mathrm{SSResid}_1^{\mathrm{II}} + \mathrm{SSResid}_2^{\mathrm{II}})^{-\frac{N-P}{2}}}{\pi^{-\frac{Nb-P}{2}} b^{-\frac{Nb}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) (\mathrm{SSResid}_1^{\mathrm{II}} + \mathrm{SSResid}_2^{\mathrm{II}})^{-\frac{Nb-P}{2}}} \\
&= \pi^{-\frac{N(1-b)}{2}} b^{\frac{Nb}{2}} (\mathrm{SSResid}_1^{\mathrm{II}} + \mathrm{SSResid}_2^{\mathrm{II}})^{-\frac{N(1-b)}{2}} \frac{\Gamma\left(\frac{N-P}{2}\right)}{\Gamma\left(\frac{Nb-P}{2}\right)}. \ \square
\end{aligned}
$$

Heteroscedastic Error Variance

A Laplace approximation is used to evaluate $\int P(\boldsymbol{Y}|\boldsymbol{\varphi}, m) P(\boldsymbol{\varphi}) d\boldsymbol{\varphi}$, parametrized with respect to the log-variances $\lambda_1 = \log \sigma_1^2$ and $\lambda_2 = \log \sigma_2^2$. Denote Λ as the log-variance matrix,

J_Λ as the transformation Jacobian, $(\lambda_1^\star,\, \lambda_2^\star)$ as the mode of the log-marginal distribution, and $\nabla^\star$ as the Hessian evaluated at this mode. A subscript of b refers to the same quantities calculated with respect to the fractional exponentiated likelihood. The joint modes $(\lambda_1^\star,\, \lambda_2^\star)$ and $(\lambda_{b_1}^\star,\, \lambda_{b_2}^\star)$ are computed using the function `optim` in R; similarly, the Hessians $\nabla^\star$ and $\nabla_b^\star$ are evaluated at these values using the function `hessian` in the package `numderiv` [Gilbert and Varadhan, 2016, R Core Team, 2017].

We first integrate over the regression effects vector $\boldsymbol{\beta}$. Let $H_\Phi := \Phi X(X^T\Phi X)^{-1}X^T\Phi$.

$$
\begin{aligned}
P^b(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_1,\varphi_2) &= \int\limits_{\Theta} \left[(2\pi)^{-\frac{N}{2}}|\Phi|^{\frac{1}{2}}\cdot\exp\{-\frac{1}{2}(\boldsymbol{Y}-X\boldsymbol{\beta})^T\Phi(\boldsymbol{Y}-X\boldsymbol{\beta})\}\right]^b P(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \int\limits_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}}\varphi_1{}^{\frac{n_1 b}{2}-1}\varphi_2{}^{\frac{n_2 b}{2}-1}\exp\{-b\cdot\frac{1}{2}[\boldsymbol{Y}^T\Phi\boldsymbol{Y}-2\boldsymbol{\beta}^TX^T\Phi\boldsymbol{Y}+\boldsymbol{\beta}^TX^T\Phi X\boldsymbol{\beta}^T]\}d\boldsymbol{\beta} \\
&= \int\limits_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}}\varphi_1{}^{\frac{n_1 b}{2}-1}\varphi_2{}^{\frac{n_2 b}{2}-1}\times \\
&\qquad \exp\{-b\cdot\frac{1}{2}[\boldsymbol{\beta}^TX^T\Phi X\boldsymbol{\beta}-2\boldsymbol{\beta}^TX^T\boldsymbol{Y}+\boldsymbol{Y}^TH_\Phi\boldsymbol{Y}]\}\times \\
&\qquad \exp\{-b\cdot\frac{1}{2}[\boldsymbol{Y}^T\Phi\boldsymbol{Y}-\boldsymbol{Y}^TH_\Phi\boldsymbol{Y}]\}d\boldsymbol{\beta} \\
&= (2\pi)^{-\frac{Nb-P}{2}}\varphi_1{}^{\frac{n_1 b}{2}-1}\varphi_2{}^{\frac{n_2 b}{2}-1}b^{-\frac{P}{2}}|X^T\Phi X|^{-\frac{1}{2}}\exp\{-b\cdot\frac{1}{2}[\boldsymbol{Y}^T\Phi\boldsymbol{Y}-\boldsymbol{Y}^TH_\Phi\boldsymbol{Y}]\}
\end{aligned}
$$

We reparametrize the precisions of $P^b(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_1,\varphi_2)$ and $P(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_1,\varphi_2)$ to log-variances λ_1 and λ_2 to elicit a shape more conducive to the Laplace approximation. For $\lambda_1 = \ln\varphi_1^{-1}$ and $\lambda_2 = \ln\varphi_2^{-1}$, $P(\boldsymbol{Y}|m_s^{\mathrm{III}},\lambda_1,\lambda_2) = P(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_1,\varphi_2)\cdot|J_\Lambda|$ where $\frac{\partial\varphi_1}{\partial\lambda_1} = -\exp\{-\lambda_1\}$ and $\frac{\partial\varphi_2}{\partial\lambda_2} = -\exp\{-\lambda_2\}$, so $|J_\Lambda| = \exp\{-(\lambda_1+\lambda_2)\}$. Thus the Laplace approximation for the fractional marginal model probability is given by

$$
q^b(\boldsymbol{Y}|m_s^{\mathrm{III}}) \approx \frac{(2\pi)|-\nabla^\star|^{-\frac{1}{2}}\cdot P(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_{b_1}^\star,\varphi_{b_2}^\star)|J_\Lambda^\star|}{(2\pi)|-\nabla_b^\star|^{-\frac{1}{2}}\cdot P^b(\boldsymbol{Y}|m_s^{\mathrm{III}},\varphi_{b_1}^\star,\varphi_{b_2}^\star)|J_{b_\Lambda}^\star|}.\ \square
$$

Empirical comparisons with cubature-based approximations, in addition to the shapes of the marginal densities $P(\boldsymbol{Y}|m_s^{\mathrm{III}},\lambda_1,\lambda_2)$ and $P^b(\boldsymbol{Y}|m_s^{\mathrm{III}},\lambda_1,\lambda_2)$, indicate that this Laplace

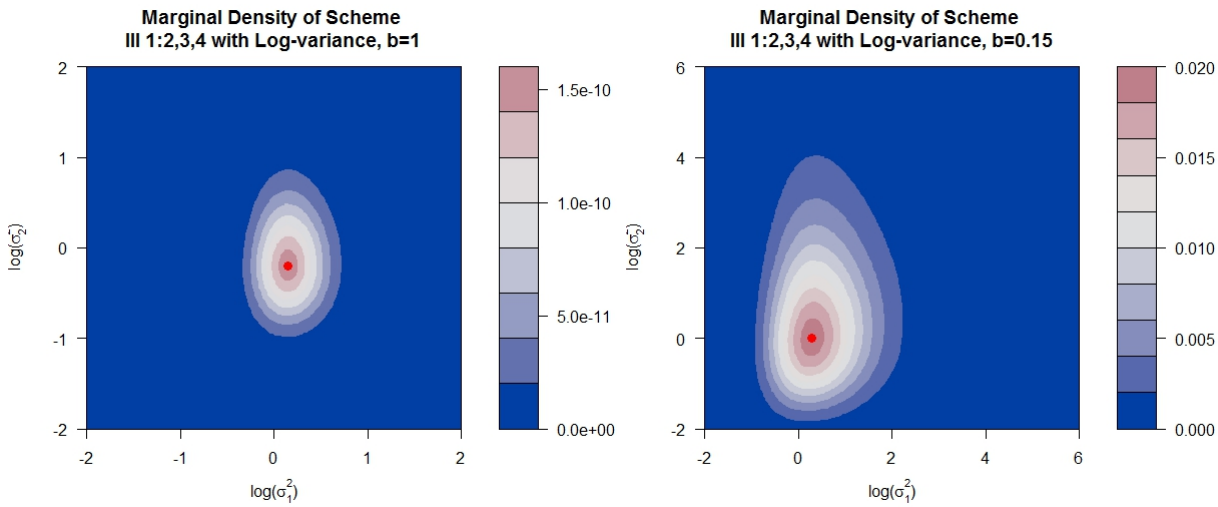

Figure 2.7: The marginal densities $P(\boldsymbol{Y}|m^{\mathrm{II}}_{1:2,3,4}, \lambda_1, \lambda_2)$ and $P^b(\boldsymbol{Y}|m^{\mathrm{III}}_{1:2,3,4}, \lambda_1, \lambda_2)$ for one particular grouping of an ANOVA layout with $K = 4$ levels of $n_k = 10$ observations each show an overall shape conducive to the Laplace approximation in both the raw and exponentiated likelihood cases.

approximation is satisfactory.

Special Case: Heteroscedastic Error Variance with no Regression Effects Spanning Two Variances

$$
\begin{aligned}
P(\boldsymbol{Y}|m_s) &= \int_{\Theta} \left[(2\pi)^{-\frac{N}{2}} |\Sigma|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2}(\boldsymbol{Y} - X\boldsymbol{\beta})^T \Sigma^{-1} (\boldsymbol{Y} - X\boldsymbol{\beta})\} \right]^b P(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_0^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} |\Sigma|^{-\frac{b}{2}} \exp\{-b \cdot \frac{1}{2}[\boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} - 2\boldsymbol{\beta}^T X^T \Sigma^{-1} \boldsymbol{Y} + \boldsymbol{\beta}^T X^T \Sigma^{-1} X \boldsymbol{\beta}]\} \\
&\qquad d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_2 d\gamma_1 d\gamma_2 \\
&= \int_0^{\infty}\int_0^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{n_1 b}{2}-1} \gamma^{\frac{n_2 b}{2}-1} \exp\{-b \cdot \frac{1}{2}[\boldsymbol{Y}^T \Sigma^{-1} \boldsymbol{Y} - 2\boldsymbol{\beta}^T X^T \Sigma^{-1} \boldsymbol{Y} + \boldsymbol{\beta}^T X^T \Sigma^{-1} X \boldsymbol{\beta}]\} \\
&\qquad d\boldsymbol{\beta}_1 d\boldsymbol{\beta}_2 d\gamma_1 d\gamma_2 \\
&= \int_0^{\infty}\int_0^{\infty} (2\pi)^{-\frac{Nb}{2}} \gamma^{\frac{n_1 b}{2}-1} \gamma^{\frac{n_2 b}{2}-1} (2\pi)^{+\frac{P}{2}} |b^{-1}\gamma^{-1}(X_1^T X_1)^{-1}|^{+\frac{1}{2}} |b^{-1}\gamma^{-1}(X_2^T X_2)^{-1}|^{+\frac{1}{2}} \times \\
&\qquad \exp\{-b \cdot \frac{\gamma}{2}[\boldsymbol{Y}_1^T (I - H_1)\boldsymbol{Y}_1 + \boldsymbol{Y}_2^T (I - H_2)\boldsymbol{Y}_2]\} d\gamma_1 d\gamma_2 \\
&= (2\pi)^{-\frac{Nb-P}{2}} b^{-\frac{P}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{Nb-P}{2}\right) \left(\frac{b \cdot \left[\mathrm{SSResid}_1^{\mathrm{IV}} + \mathrm{SSResid}_2^{\mathrm{IV}}\right]}{2}\right)^{-\frac{Nb-P}{2}} \\
&= (2\pi)^{-\frac{Nb-P}{2}} b^{-\frac{P}{2}} |X_1^T X_1|^{-\frac{1}{2}} |X_2^T X_2|^{-\frac{1}{2}} \Gamma\left(\frac{n_1 b - p_1}{2}\right) \Gamma\left(\frac{n_2 b - p_2}{2}\right) \times \\
&\qquad \left(\frac{b \cdot \mathrm{SSResid}_1^{\mathrm{IV}}}{2}\right)^{-\frac{n_1 b - p_1}{2}} \left(\frac{b \cdot \mathrm{SSResid}_2^{\mathrm{IV}}}{2}\right)^{-\frac{n_2 b - p_2}{2}} .
\end{aligned}
$$

$$
\begin{aligned}
q^b(\boldsymbol{Y}|m_s) &= \frac{\int\limits_{\Theta} P(Y|m_s,\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int\limits_{\Theta} P^b(Y|m_s^{\mathrm{IV}},\boldsymbol{\theta})P(\boldsymbol{\theta})d\boldsymbol{\theta}} \\
&= \frac{(2\pi)^{-\frac{N-P}{2}}|X_1X_1|^{-\frac{1}{2}}|X_2X_2|^{-\frac{1}{2}}\Gamma\left(\frac{n_1-p_1}{2}\right)\Gamma\left(\frac{n_2-p_2}{2}\right)}{(2\pi)^{-\frac{Nb-P}{2}}b^{-\frac{P}{2}}|X_1X_1|^{-\frac{1}{2}}|X_2X_2|^{-\frac{1}{2}}\Gamma\left(\frac{n_1b-p_1}{2}\right)\Gamma\left(\frac{n_2b-p_2}{2}\right)}\times \\
&\qquad \frac{\left(\frac{\mathrm{SSResid}_1^{\mathrm{IV}}}{2}\right)^{-\frac{n_1-p_1}{2}}\left(\frac{\mathrm{SSResid}_2^{\mathrm{IV}}}{2}\right)^{-\frac{n_2-p_2}{2}}}{\left(\frac{b\cdot\mathrm{SSResid}_1^{\mathrm{IV}}}{2}\right)^{-\frac{n_1b-p_1}{2}}\left(\frac{b\cdot\mathrm{SSResid}_2^{\mathrm{IV}}}{2}\right)^{-\frac{n_2b-p_2}{2}}} \\
&= \pi^{-\frac{N(1-b)}{2}}b^{\frac{Nb}{2}}(\mathrm{SSResid}_1^{\mathrm{IV}})^{-\frac{n_1(1-b)}{2}}(\mathrm{SSResid}_2^{\mathrm{IV}})^{-\frac{n_2(1-b)}{2}}\frac{\Gamma\left(\frac{n_1-p_1}{2}\right)}{\Gamma\left(\frac{n_1b-p_1}{2}\right)}\frac{\Gamma\left(\frac{n_2-p_2}{2}\right)}{\Gamma\left(\frac{n_2b-p_2}{2}\right)}. \ \square
\end{aligned}
$$

2.7.3 Derivation of Model Probabilities: Mixture g Regression Effect Priors

$$
\begin{aligned}
P(\boldsymbol{Y}, g|m_s) &= \int_0^\infty\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty P^b(\boldsymbol{Y}|\alpha,\beta,\varphi,m_s)P(\alpha,\varphi)P(\beta|\varphi,g)P(g)d\alpha d\beta d\varphi dg \\
&= \int_0^\infty\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty (2\pi)^{-\frac{Nb}{2}}\varphi^{\frac{Nb}{2}}\exp\{-\frac{\varphi b}{2}(\boldsymbol{Y}-\mathbf{1}^T\alpha - X\boldsymbol{\beta})^T(\boldsymbol{Y}-\mathbf{1}^T\alpha - X\boldsymbol{\beta})\}\times \\
&(2\pi)^{-\frac{P}{2}}|g\varphi^{-1}(X^TX)^{-1}|^{-\frac{1}{2}}\exp\{-\frac{\varphi}{2}\frac{\boldsymbol{\beta}^TX^TX\boldsymbol{\beta}}{g}\}\varphi^{-1}P(g)d\alpha d\boldsymbol{\beta}d\varphi dg \\
&= \int_0^\infty\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty (2\pi)^{-\frac{Nb+P}{2}}\varphi^{\frac{Nb}{2}-1}\times \\
&\exp\{-\frac{\varphi}{2}(\alpha^2b\mathbf{1}^T\mathbf{1} - 2\alpha b(\mathbf{1}^T\boldsymbol{Y} - \mathbf{1}^TX\boldsymbol{\beta}) + b(\boldsymbol{Y}-X\boldsymbol{\beta})H_{\mathbf{1}}(\boldsymbol{Y}-X\boldsymbol{\beta}))\}\times \\
&|g\varphi^{-1}(X^TX)^{-1}|^{-\frac{1}{2}}\exp\{-\frac{\varphi}{2}(\boldsymbol{\beta}^T\left(\frac{X^TX}{g}+bX^TX\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^TbX^T\boldsymbol{Y})\}\times \\
&\exp\{-\frac{\varphi}{2}(b\boldsymbol{Y}^T\boldsymbol{Y} - b(\boldsymbol{Y}-X\boldsymbol{\beta})H_{\mathbf{1}}(\boldsymbol{Y}-X\boldsymbol{\beta}))\}P(g)d\alpha d\boldsymbol{\beta}d\varphi dg \\
&= \int_0^\infty\int_0^\infty\int_{-\infty}^\infty (2\pi)^{-\frac{Nb+P-1}{2}}\varphi^{\frac{Nb+P-1}{2}-1}b^{-\frac{1}{2}}|X^TX|^{-\frac{1}{2}}N^{-\frac{1}{2}}g^{-\frac{P}{2}}\times \\
&\exp\{-\frac{\varphi}{2}(\boldsymbol{\beta}^T\left(\frac{X^TX}{g}+bX^TX\right)\boldsymbol{\beta} - 2\boldsymbol{\beta}^TbX^T\boldsymbol{Y} + \frac{b^2g}{1+bg}\boldsymbol{Y}^TH_X\boldsymbol{Y})\}\times \\
&\exp\{-\frac{\varphi}{2}(b\boldsymbol{Y}^T\boldsymbol{Y} - b\boldsymbol{Y}^TH_{\mathbf{1}}\boldsymbol{Y} - \frac{b^2g}{1+bg}\boldsymbol{Y}^TH_X\boldsymbol{Y})\}P(g)d\boldsymbol{\beta}d\varphi dg \\
&= \int_0^\infty\int_0^\infty (2\pi)^{-\frac{Nb-1}{2}}\varphi^{\frac{Nb-1}{2}-1}b^{-\frac{1}{2}}N^{-\frac{1}{2}}(1+bg)^{-\frac{P}{2}}\times \\
&\exp\{-\frac{\varphi}{2}(b\boldsymbol{Y}^T\boldsymbol{Y} - b\boldsymbol{Y}^TH_{\mathbf{1}}\boldsymbol{Y} - \frac{b^2g}{1+bg}\boldsymbol{Y}^TH_X\boldsymbol{Y})\}d\varphi dg \\
&= \int_0^\infty \frac{\Gamma\left(\frac{Nb-1}{2}\right)}{\sqrt{\pi}^{Nb-1}\sqrt{N}}b^{-\frac{Nb}{2}}(1+bg)^{-\frac{Nb-P-1}{2}}\mathrm{SST}^{-\frac{Nb-1}{2}}[1+bg(1-R^2)]^{-\frac{Nb-1}{2}}P(g)dg. \ \square
\end{aligned}
$$

After integrating over α, $\boldsymbol{\beta}$, and $\boldsymbol{\varphi}$, we obtain

$$P(\boldsymbol{Y}, g|m_s^c) = \int_0^\infty \frac{\Gamma(\frac{Nb-1}{2})b^{-\frac{Nb}{2}}}{\sqrt{\pi}^{Nb-1}\sqrt{N}} \frac{\mathrm{SST}^{-\frac{Nb-1}{2}}}{(1+bg)^{\frac{Nb-P-1}{2}}}[1+bg(1-R^2)]^{-\frac{Nb-1}{2}} g^{-1.5}\exp\{-\frac{N}{2g}\}dg \quad (2.24)$$

To execute the Laplace approximation over g, we must obtain the mode $g^\star = \underset{g}{\operatorname{argmax}}\, P(\boldsymbol{Y}, g|m_s^c)$, the root of the equation

$$-Qb^2(P+3)g^3 + (b(Nb-P-4)-2Q)g^2 + (Nb(2-R^2)-3)g + N := 0 \qquad (2.25)$$

where $Q = 1 - R^2$. We also require the Hessian evaluated at the mode,

$$\begin{aligned} H^\star &= \frac{\partial^2}{\partial g^2}[\log((1+bg)^{\frac{Nb-P-1}{2}}(1+Qbg)^{-\frac{Nb-1}{2}} g^{-\frac{3}{2}}\exp\{-\frac{N}{2g}\})]\Big|^{g=g^\star} \\ &= \frac{1}{2}\left[\frac{(Nb-1)b^2Q^2}{(1+Qbg^\star)^2} - \frac{(Nb-P-1)b^2}{(1+bg^\star)^2} + \frac{3}{(g^\star)^2} - \frac{2N}{(g^\star)^3}\right]. \end{aligned}$$

These expressions are appropriate to use in homoscedastic classes with either global or distinct regression effects; we simply compute the corresponding R^2 and use the appropriate P based on the model under consideration.

In classes with heteroscedasticity, the integral is intractable over both $\boldsymbol{\varphi}$ and g; thus we must employ a three-dimensional Laplace approximation to evaluate this integral. For computational ease and to improve the accuracy of the approximation, we again parametrize with respect to the log-variance; let Λ represent the log-variance matrix. Denote $H_\Lambda = \Lambda\mathbf{1}(\mathbf{1}^T\Lambda\mathbf{1})^{-1}\mathbf{1}^T\Lambda$; then integrating out the global intercept and regression effects yields an expression for $P^b(\boldsymbol{Y}, \lambda_1, \lambda_2, g|m_s^c)$:

$$= (2\pi)^{-\frac{Nb+P-1}{2}}\lambda_1^{\frac{n_1 b}{2}-1}\lambda_2^{\frac{n_2 b}{2}-1} g^{-\frac{P}{2}} b^{-\frac{P+1}{2}} |X^T\Lambda X|^{\frac{1}{2}} |\mathbf{1}^T\Lambda\mathbf{1}|^{-\frac{1}{2}} \left|\left(\frac{1+bg}{bg}\right)X^T\Lambda X - X^T H_\Lambda X\right|^{-\frac{1}{2}} \times$$
$$J_\Lambda \cdot \exp\{-\frac{b}{2}[\boldsymbol{Y}\Lambda\boldsymbol{Y} - \boldsymbol{Y}^T H_\Lambda \boldsymbol{Y} - \boldsymbol{Y}^T(\Lambda - H_\Lambda)^T X(\frac{1+bg}{bg}X^T\Lambda X - X^T H_\Lambda X)X^T(\Lambda - H_\Lambda)\boldsymbol{Y}^T]\}$$

The joint mode $(\lambda_1^\star,\ \lambda_2^\star,\ g^\star)$ is computed using the function `optim` in R; similarly, the Hessian

is computed at this value using the function `hessian` in the package `numderiv`.

Heteroscedastic Zellner-Siow Cauchy Result

Let $\boldsymbol{\beta}|\Phi,\, g \sim N(\mathbf{0},\, g(X^T\Phi X)^{-1})$ and $g \sim \text{IG}(\frac{1}{2},\, \frac{N}{2})$.

Then $P(\boldsymbol{\beta}|g,\, \Phi) = (2\pi)^{-\frac{P}{2}}|g(X^T\Phi X)^{-1}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\beta}\frac{X^T\Phi X}{g}\boldsymbol{\beta}\}$ and $P(g) = \frac{\left(\frac{N}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}g^{-\frac{3}{2}}\exp\{-\frac{N}{2g}\}$.

$$
\begin{aligned}
P(\boldsymbol{\beta}|\boldsymbol{\varphi}) &= \int P(\boldsymbol{\beta}|\Phi,\, g)P(g)dg \\
&= \int (2\pi)^{-\frac{P}{2}}|g(X^T\Phi X)^{-1}|^{-\frac{1}{2}}\exp\{-\frac{1}{2}\boldsymbol{\beta}\frac{X^T\Phi X}{g}\boldsymbol{\beta}\} \times \frac{\left(\frac{N}{2}\right)^{\frac{1}{2}}}{\Gamma\left(\frac{1}{2}\right)}g^{-\frac{3}{2}}\exp\{-\frac{N}{2g}\}dg \\
&= \int (2\pi)^{-\frac{P}{2}}|X^T\Phi X|^{\frac{1}{2}}(\frac{N}{2})^{\frac{1}{2}}\pi^{-\frac{1}{2}}g^{-\frac{P+3}{2}}\exp\{-\frac{1}{2}\boldsymbol{\beta}^T\frac{X^T\Phi X}{g}\boldsymbol{\beta} - \frac{1}{2}\cdot\frac{N}{g}\}dg \\
&= \int (2\pi)^{-\frac{P}{2}}|X^T\Phi X|^{\frac{1}{2}}N^{\frac{1}{2}}2^{-\frac{1}{2}}\pi^{-\frac{1}{2}} \times g^{-\frac{P+1}{2}-1}\exp\{-\frac{\boldsymbol{\beta}^T X^T\Phi X\boldsymbol{\beta}+N/2}{g}\}dg \\
&= 2^{-\frac{P+1}{2}}\pi^{-\frac{P+1}{2}}|X^T\Phi X|^{\frac{1}{2}}N^{\frac{1}{2}}[(\boldsymbol{\beta}^T X^T\Phi X\boldsymbol{\beta}+N)^{\frac{P+1}{2}}2^{-\frac{P+1}{2}}\Gamma\left(\frac{P+1}{2}\right)]^{-1} \\
&= \pi^{-\frac{P+1}{2}}|X^T\Phi X|^{\frac{1}{2}}N^{\frac{1}{2}}(\boldsymbol{\beta}^T X^T\Phi X\boldsymbol{\beta}+N)^{-\frac{P+1}{2}}\Gamma\left(\frac{P+1}{2}\right)^{-1} \\
&= \pi^{-\frac{P+1}{2}}|X^T\Phi X|^{\frac{1}{2}}N^{\frac{1}{2}}(\frac{\boldsymbol{\beta}^T X^T\Phi X\boldsymbol{\beta}+N}{N})^{-\frac{P+1}{2}}N^{-\frac{P+1}{2}}\Gamma\left(\frac{P+1}{2}\right)^{-1} \\
&= \pi^{-\frac{P+1}{2}}|X^T\Phi X|^{\frac{1}{2}}N^{-\frac{P}{2}}(1+\boldsymbol{\beta}^T\frac{X^T\Phi X}{N}\boldsymbol{\beta})^{-\frac{P+1}{2}}\Gamma\left(\frac{P+1}{2}\right)^{-1} \\
&= \pi^{-\frac{p+1}{2}}\Gamma\left(\frac{P+1}{2}\right)^{-1} \times |\frac{X^T\Phi X}{N}|^{\frac{1}{2}}(1+\boldsymbol{\beta}^T\frac{X^T\Phi X}{N}\boldsymbol{\beta})^{-\frac{P+1}{2}} \\
&\sim \text{MVCauchy}_P(\text{location} = 0,\, \text{scale} = \left(\frac{X^T\Phi X}{N}\right)^{-1}).\ \square
\end{aligned}
$$

2.7.4 Supplemental ANCOVA Simulation Study

We let $n_k := 10$ for each level of the SLGF. The parameter settings are provided in Table 2.1.

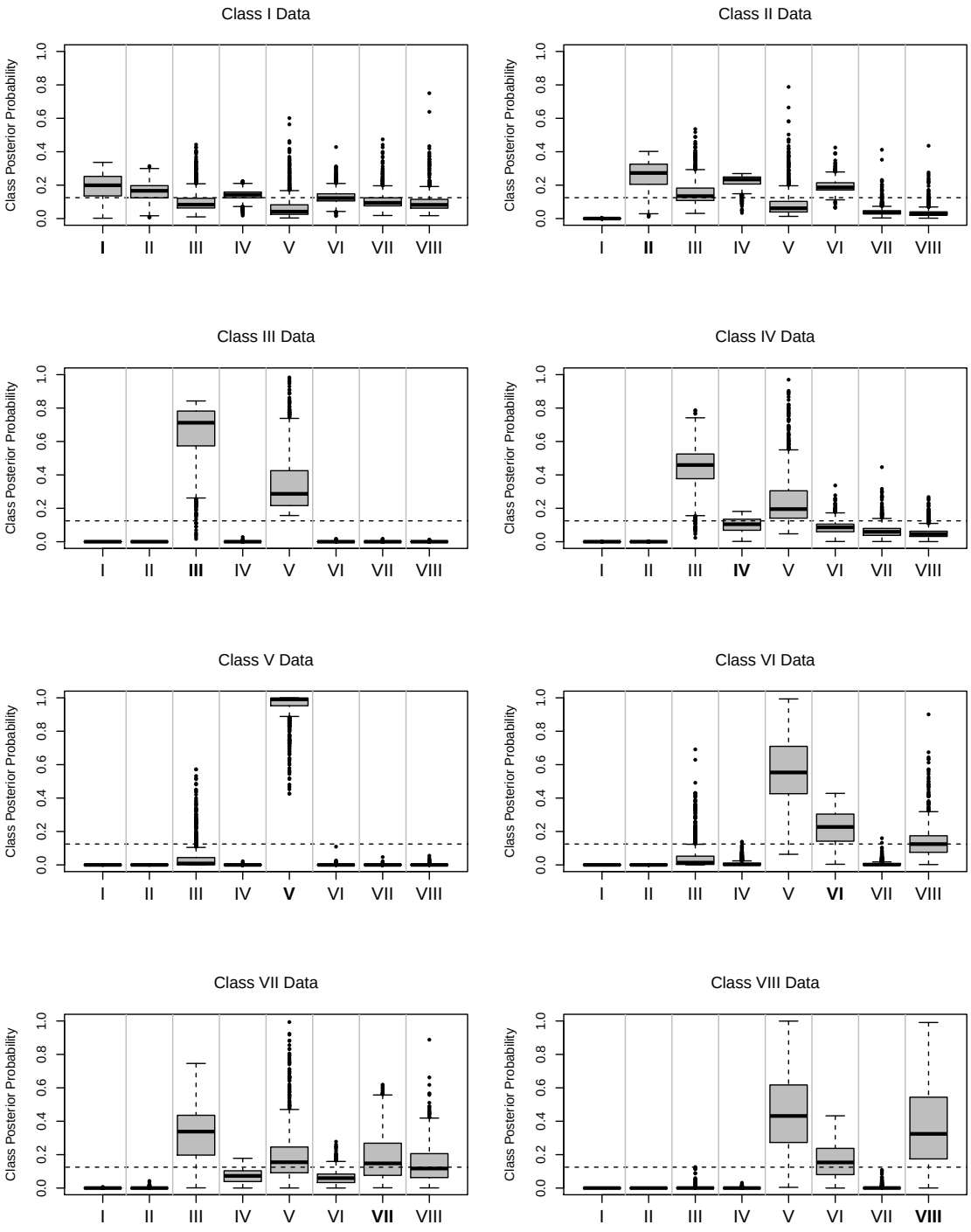

Figure 2.8: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo data sets with $K = 4$ levels of the categorical predictor, each with 10 observations, for a total of $N = 40$ observations. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

Class	Parameters
I (Additive Model)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5, 7, 8, 9, 10\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma^2 = 1$
II (Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5, 7, 8, 9, 10\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$, $\boldsymbol{\tau}_2 \in \{3.0, 2.5, 2.0, 1.5, 1.0\}$, $\sigma^2 = 1$
III (Heteroscedastic Additive)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5, 7, 8, 9, 10\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$
IV (Heteroscedastic Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5, 7, 8, 9, 10\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$, $\boldsymbol{\tau}_2 \in \{3.0, 2.5, 2.0, 1.5, 1.0\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$

Table 2.3: Settings for the four model classes in the 10×5 two-way layout simulation study with smaller effect size.

Class	Parameters
I (Additive Model)	$\boldsymbol{\nu} \in \{1, 1.5, 2, 2.5, 3\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma^2 = 1$
II (Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.8, 2.6, 3.4, 4.2\}$, $\boldsymbol{\tau}_2 \in \{4.2, 3.4, 2.6, 1.8, 1.0\}$, $\sigma^2 = 1$
III (Heteroscedastic Additive)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$
IV (Heteroscedastic Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.8, 2.6, 3.4, 4.2\}$, $\boldsymbol{\tau}_2 \in \{4.2, 3.4, 2.6, 1.8, 1.0\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$

Table 2.4: Settings for the four model classes in the two-way layout simulation study with 5×5 layouts with larger effect size.

2.7.5 Supplemental Two-way Layout Simulation Study

We provide three additional simulation studies in the twoway layout scenario: 10×5 layouts with a smaller effect size than the study provided in Section 2.4.2, as well as 5×5 studies with larger and smaller effect sizes.

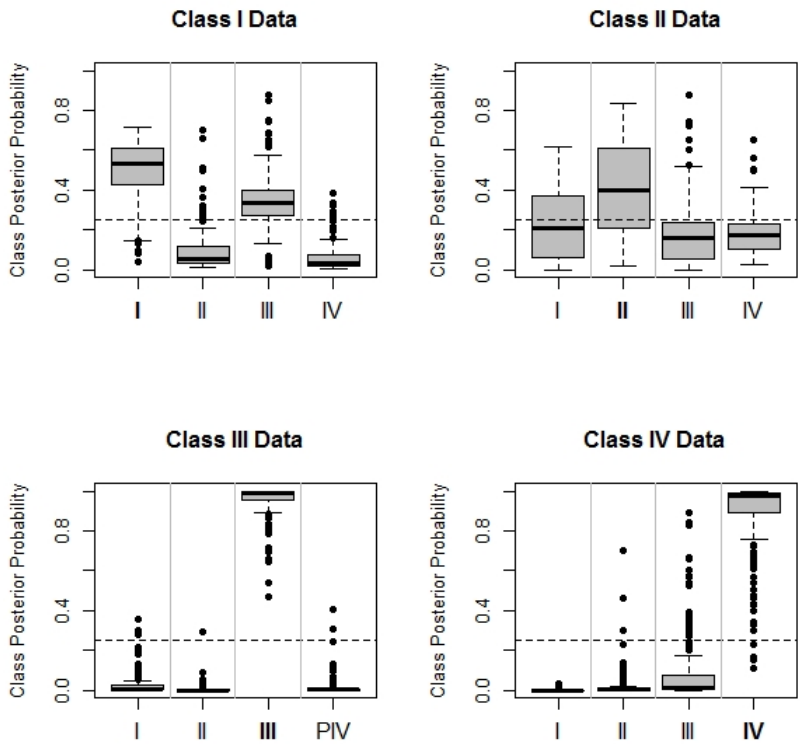

Figure 2.9: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo 10×5 layouts with smaller effect size. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

Class	Parameters
I (Additive Model)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma^2 = 1$
II (Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$, $\boldsymbol{\tau}_2 \in \{3.0, 2.5, 2.0, 1.5, 1.0\}$, $\sigma^2 = 1$
III (Heteroscedastic Additive)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau} \in \{1, 2, 3, 4, 5\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$
IV (Heteroscedastic Group-by-Column Interaction)	$\boldsymbol{\nu} \in \{1, 2, 3, 4, 5\}$, $\boldsymbol{\tau}_1 \in \{1.0, 1.5, 2.0, 2.5, 3.0\}$, $\boldsymbol{\tau}_2 \in \{3.0, 2.5, 2.0, 1.5, 1.0\}$, $\sigma_1^2 = 1.0$, $\sigma_2^2 = 0.25$

Table 2.5: Settings for the four model classes in the two-way layout simulation study with 5×5 layouts with smaller effect size.

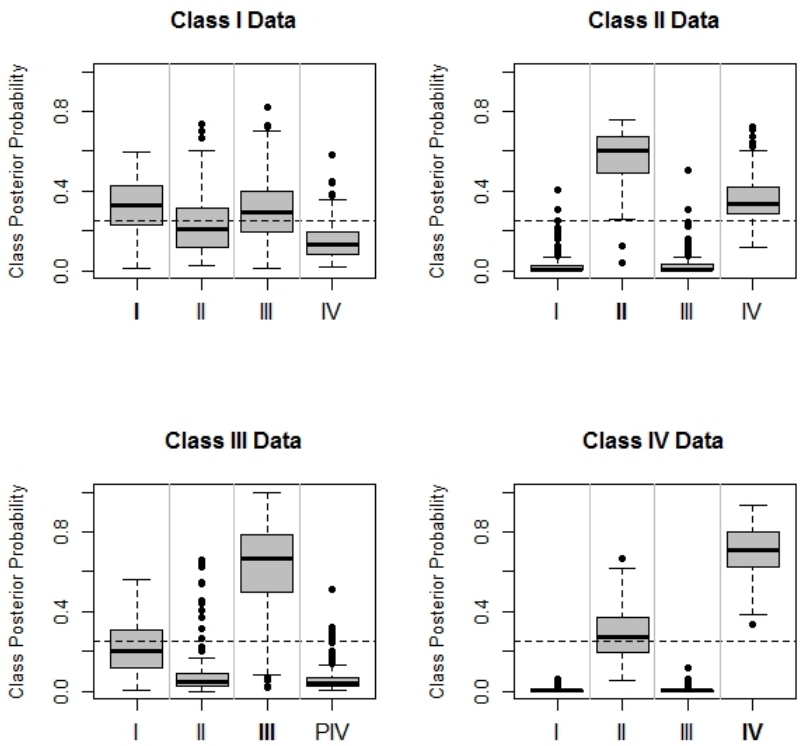

Figure 2.10: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo 5×5 layouts with larger effect size. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

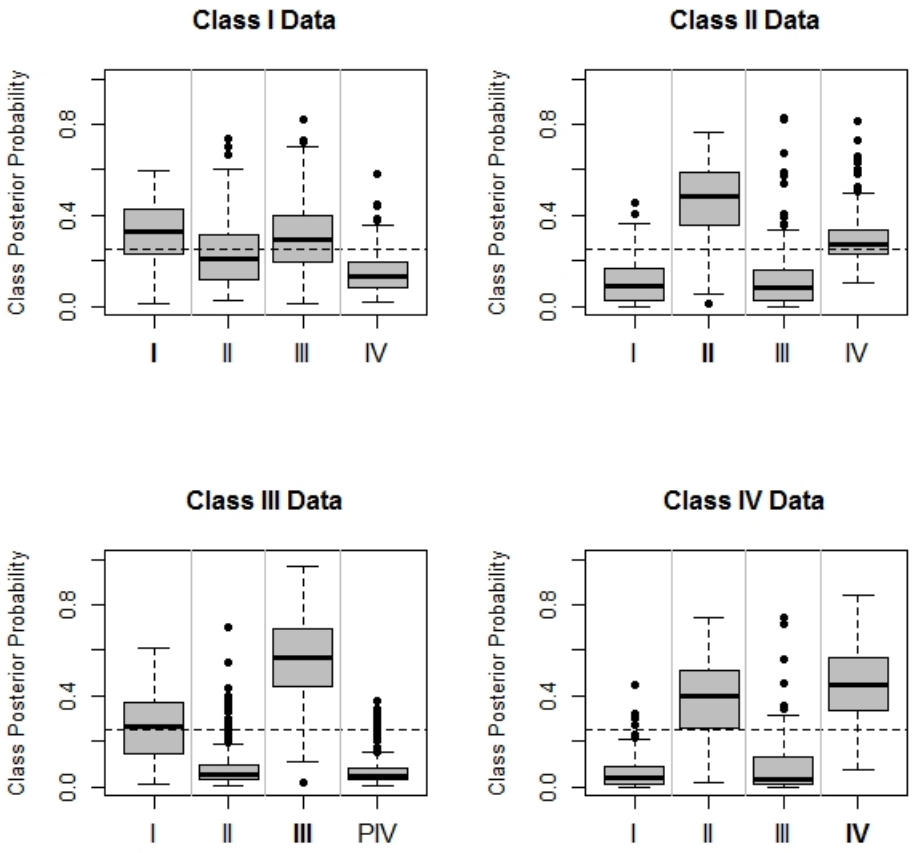

Figure 2.11: Posterior probabilities (y-axis) by class based on 1000 Monte Carlo 5×5 layouts with smaller effect size. The true model class is emphasized in bold on the x-axis. The dashed line indicates the prior by model class.

Chapter 3

Bayesian Model Selection with Group-Based Regression Effects and Heteroscedasticity using the R Package slgf

Linear models are flexible and among the most frequently implemented statistical methods. However, their performance can suffer when assumptions are violated. We focus on the scenario in which the levels of a categorical predictor exhibit two latent groups, leading to group-based regression effects and/or heteroscedastic error variance. First we review the SLGF (suspected latent grouping factor) method. Next, using both observational and experimental data, we illustrate the usage of the R package `slgf` in the context of several common linear model layouts: one-way analysis of variance (ANOVA), analysis of covariance (ANCOVA), a two-way replicated layout, a two-way unreplicated layout, and a balanced incomplete block design (BIBD). We have selected data that reveal the shortcomings of classical analyses to emphasize the advantage our method can provide when a latent grouping structure underlies the data.

3.1 Introduction

Linear models with categorical predictors are pervasive in the social and natural sciences. Many well-established classical and Bayesian methods exist to analyze such models, but conventional approaches may often rely on assumptions that are restrictive and fail to account for more subtle latent structures within the data. Several examples of such datasets can be found in Franck and Osborne [2016], Franck [2018], Kharrati-Kopaei and Sadooghi-Alvandi [2007], and Metzger and Franck [2019]. Consider Figure 3.1, which illustrate three of the five datasets considered herein. In each case, upon inspecting the plot, the researcher might suspect that there is a latent grouping structure (indicated by color-coding) where groups are formed by partitioning the levels of a factor; we call this the *suspected latent grouping factor*, or SLGF. The left panel represents a one-way analysis of variance (ANOVA) study where a continuous measurement of olfactory acuity (y-axis) is modeled as a function of age, where age is represented in five categories (x-axis) [O'Brien and Heft, 1995]. We suspect that the levels 1, 2, and 3 of the slgf, age, have distinct means and error variances from levels 4 and 5. The center panel shows a BIBD, where abrasion from a road test was measured on four tires (blocks). Four treatments (compounds used to manufacture the tires) were analyzed, but due to manufacturing limitations each tire was made from sections of three compounds each. Compounds A and B appear to behave similarly to one another, as do C and D. Finally, the right panel shows the example described by Meek and Ozgur [1991], where the torque required to tighten a locknut was measured as a function of a plating process and a threading technique. The plating processes analyzed included treatments with cadmium and wax, heat treating, and phosphate and oil, denoted CW, HT, and PO, respectively. The threading techniques studied include bolt and mandrel, the types of fixture on which each locknut was affixed to conduct the test. We suspect that observations measured on a bolt may have a higher error variance than those from the mandrel.

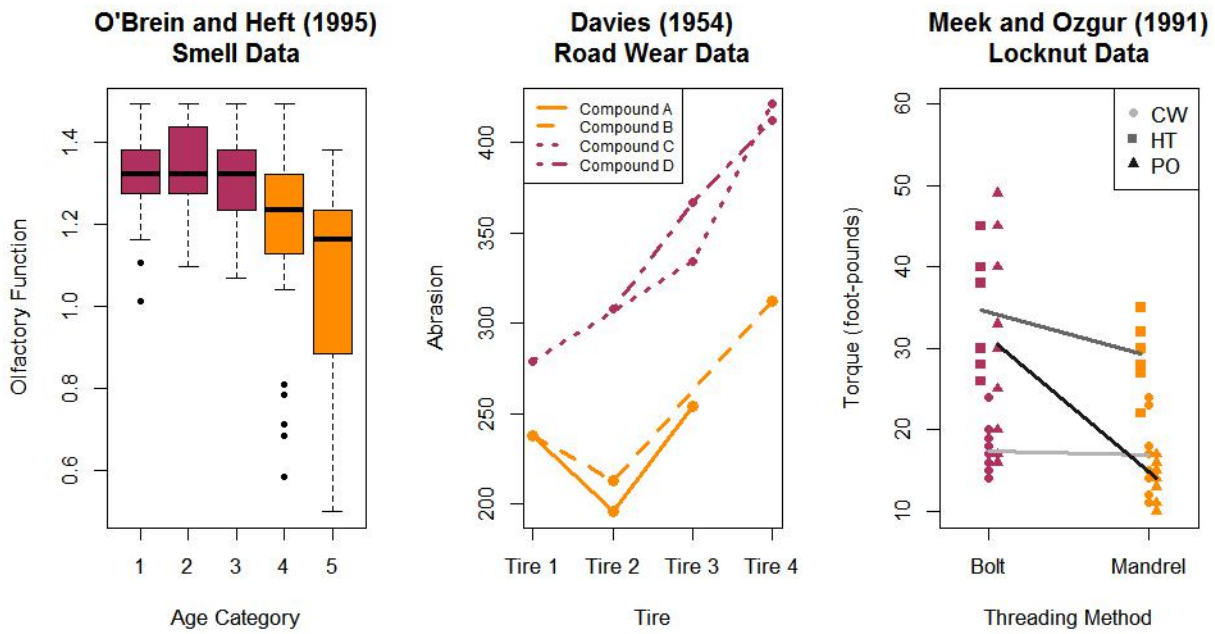

Figure 3.1: Data from O'Brien and Heft [1995] (left), Davies [1954] (center), and Meek and Ozgur [1991] (right). These data sets appear to have a latent grouping structure, color-coded for emphasis. The left panel appears to show two groups with both distinct means and error variances; the center panel shows two groups with distinct regression effects; and the right panel shows group-based error variances, and group-based interactions may also be present.

3.1.1 Model Classes and Grouping Schemes

We first describe the model classes associated with the SLGF methodology; for a thorough review see Metzger and Franck [2019]. Note that the color-coding in Figure 3.1 is based on subjective appearance. In the absence of prior information regarding grouping structures, we consider all possible ways to partition the levels of the SLGF We denote $\mathcal{S}$ as the set of all such possible *grouping schemes*. The function `lm.slgf` automatically considers all possible grouping schemes based on a user-chosen SLGF, where the user can also specify the minimum number of levels of the SLGF that may comprise an individual group. Notationally, we denote the grouping schemes of Figure 3.1 as 1,2,3:4,5 and A,B:C,D and bolt:mandrel respectively. The number of possible grouping schemes grows exponentially with the number of levels of the SLGF, so `slgf` is recommended for use with categorical predictors with 10 or fewer levels.

Additionally, our suspicions in Figure 3.1 from left to right of group-based means and variances, group-based regression effects, and group-based variances are also nothing but subjective guesses at this point. We must also account for all such user-specified model structures in addition to the combinatorically possible grouping schemes. We thus introduce the concept of a *model class*, which summarizes the model effects along with its error variance structure (either homoscedastic or heteroscedatic). We illustrate this idea with the `roadwear` data set in the center panel of Figure 3.1:

```
> head(roadwear)
  abrasion compound tire
1      238        A    1
2      238        B    1
3      279        C    1
4      196        A    2
5      213        B    2
6      308        D    2
```

The column `abrasion` denotes the continuous response, `compound` the treatment factor, and `tire` the blocking factor. We consider `compound` as the SLGF. Thus we might reasonably consider many models; three non-exhaustive examples are provided here:

1. `abrasion~compound+tire`, where the `compound` and `tire` effects each have three degrees of freedom. This simple, standard model does not rely on the SLGF methodology, and its inclusion as a candidate helps assess the plausibility that a grouping structure actually underlies the data.

2. `abrasion~group+tire`, where the `tire` effect again has three degrees of freedom, but the `group` effect contracts four levels of `compound` into two groups so the `group` effect has only one degree of freedom, thus reducing the dimensionality of the model.

3. `abrasion~group`, where the only effect estimated is a one parameter `group` effect.

These models are just three possible parametrizations, and `slgf` allows the user to tailor the set of candidate models to subject-specific considerations. When the user includes the string `group` in a model specification, the SLGF is automatically considered in the context of a group-based effect. With each such specification, the user must also specify whether there is also latent group-based heteroscedasticity; when a 1 is indicated in the corresponding argument of `het`, group-based variances will automatically be considered. It is possible to consider group-based variances without group-based regression effects if the group term does not appear in the corresponding `usermodel`. Such a model specification, along with either a single error variance or group-based error variances, comprise a model class; see Section 3.4 for more detail.

Note that some model classes may contain one model only. For example, when the first example given above is considered in the model space, there is only possible candidate model. On the other hand, example 2 contains a `group` effect, and there are six possible grouping schemes by `compound`, so this model class contains six models. If example 1 is considered with group-based variances, there are still three possible grouping schemes for the variances but

only one possible regression effect specification, so this class would also contain six models. See Metzger and Franck [2019] for a more thorough description of grouping variables and schemes.

3.1.2 Classical Approach

Let us narrow our focus to the leftmost panel, the olfactory data set. A researcher's first inclination for analysis might consist of the following steps:

1. Choose a signifcance level α;
2. For the age-level means $\{\mu_i\}_{i=1}^5$, conduct an F-test to compare the null hypothesis, $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ vs. H_A : at least one mean is different;
3. If we reject the null hypothesis, conduct a post-hoc analysis to compare the distinct means of each age category.

In R, this analysis might is usually conducted by fitting a null model with a single mean effect, an alternative model with an age category effect with four degrees of freedom, and comparing the two via the `anova` function: We note that the p-value of the ANOVA F-test is approximately 1.3947×10^{-11}, and thus conclude that there are distinct means for any reasonable significance level α. We then might perform a post-hoc analysis, such as the Tukey honest significant difference test [Tukey, 1949]:

```
> TukeyHSD(aov(smell.alt), "as.factor(agecat)")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = smell.alt)
```

```
$`as.factor(agecat)`
           diff         lwr          upr     p adj
2-1  0.02824415 -0.08663923  0.143127531 0.9610275
3-1 -0.01075188 -0.14506207  0.123558306 0.9994692
4-1 -0.11580171 -0.22577846 -0.005824966 0.0335696
5-1 -0.25727569 -0.36786495 -0.146686433 0.0000000
3-2 -0.03899603 -0.17462764  0.096635579 0.9325096
4-2 -0.14404587 -0.25563257 -0.032459160 0.0043323
5-2 -0.28551984 -0.39771027 -0.173329416 0.0000000
4-3 -0.10504983 -0.23655127  0.026451604 0.1838271
5-3 -0.24652381 -0.37853793 -0.114509694 0.0000070
5-4 -0.14147398 -0.24863454 -0.034313415 0.0032756
```

Note this method computes $\frac{\binom{5}{2}}{2} = 10$ p-values for pairwise difference comparisons, concluding, for instance, that the means of levels 1 and 2, 1 and 3, and 2 and 3 are statistically equivalent. On the other hand, levels 1 and 5, among many others, have statistically distinct means. The comparison between levels 4 and 1 has a p-value significant at $\alpha = 0.05$, but not $\alpha = 0.01$. And perhaps most importantly, this method does not account for the apparent group-based heteroscedasticity.

Our proposed method and the `slgf` package address these concerns in a user-friendly manner. The user can specify the model classes they wish to consider using familiar R syntax, and posterior model probabilities are automatically computed for each model class and grouping scheme. This can summarize both the effects that should be included in the model, and the latent grouping structure underlying the SLGF, if it is indeed present.

The remainder of this chapter is organized as follows: Section 2 provides the mathematical background of the SLGF approach; Section 3 outlines the functions that comprise the `slgf` package; Section 4 provides detailed examples demonstrating the functionality of `slgf`; and Section 5 provides discussion and future avenues to improve the package.

3.2 Mathematical Details

3.2.1 Model Specification

We next briefly review a model framework flexible enough to account for all possible model classes, employing the notation of Metzger and Franck [2019]. We begin with the model

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{3.1}$$

where $\boldsymbol{Y}$ is an $N \times 1$ vector of continuous observations; X is an $N \times P$ design matrix; $\boldsymbol{\beta}$ is a $P \times 1$ vector of regression effects; and $\boldsymbol{\varepsilon}$ is an $N \times 1$ residual vector where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \Sigma)$. We must augment our notation to account for the full SLGF with K degrees of freedom or a 2-degree of freedom group effect; interactions with the SLGF or group effect; other effects of interest unrelated to the SLGF; and potential group-based heteroscedasticity. Thus we let $X = (\mathbf{1}^T|W|V|U)$ and $\boldsymbol{\beta} = (\alpha, \boldsymbol{\nu}, \boldsymbol{\tau}, \boldsymbol{\rho})$ to obtain

$$\boldsymbol{Y} = \mathbf{1}^T\alpha + W\boldsymbol{\nu} + V\boldsymbol{\tau} + U\boldsymbol{\rho} + \boldsymbol{\varepsilon} \tag{3.2}$$

where $\mathbf{1}^T$ is an $N \times 1$ vector of 1s; α is a scalar intercept common to all models; $\boldsymbol{\nu}$ is the K-dimensional vector of SLGF effects; W is the corresponding $K \times N$ design matrix; $\boldsymbol{\tau}$ is the J-dimensional vector of additional effects; V is the corresponding $J \times N$ design matrix; $\boldsymbol{\rho}$ is the L-dimensional vector of SLGF-interaction effects; and U is the corresponding $L \times N$ design matrix.

Not all linear models will incorporate each term; for example, the ANOVA layout in the left panel of Figure 1 contains only a single categorical predictor so $\tau := \mathbf{0}$ and $\rho := \mathbf{0}$. When the SLGF is modeled as a group effect, denote this effect as $\tilde{\boldsymbol{\nu}}$; when there is an interaction involving the group effect, denote it as $\tilde{\boldsymbol{\rho}}$.

In heteroscedastic contexts, let σ_1^2 and σ_2^2 represent the error variances of groups 1 and

2, respectively. Let $\tilde{\Sigma}$ denote the covariance matrix where the ith diagonal element is σ_1^2 if y_i belongs to group 1, or σ_2^2 if y_i belongs to group 2.

3.2.2 Parameter Priors

For the usual case where prior information is unavailable, we prefer noninformative priors on the regression effects and error variance(s); for the regression effects, this is a flat prior. For homoscedastic models,

$$P(\boldsymbol{\beta}, \sigma^2|m) \propto \frac{1}{\sigma^2} \tag{3.3}$$

and for a heteroscedastic models

$$P(\boldsymbol{\beta}, \sigma_1^2, \sigma_2^2|m) \propto \frac{1}{\sigma_1^2 \cdot \sigma_2^2} \tag{3.4}$$

However, in contexts with limited data, such as the two-way unreplicated layout and BIBD illustrated in Sections 4.3 and 4.4, respectively, we employ the Zellner-Siow mixture g-prior [Zellner and Siow, 1980, Zellner, 1986, Liang et al., 2008], which reduces the minimal training sample size necessary for the computation of the fractional Bayes factor (see Section 3.2.3 for further detail). For homoscedastic models,

$$P(\alpha, \sigma^2|m) \propto \frac{1}{\sigma^2} \text{ and } \boldsymbol{\beta}_{-\alpha}|\Sigma, g, m \sim N(\mathbf{0},\, g(X^T\Sigma X)^{-1}); \tag{3.5}$$

for heteroscedastic models,

$$P(\alpha, \sigma_1^2, \sigma_2^2|m) \propto \frac{1}{\sigma_1^2 \cdot \sigma_2^2} \text{ and } \boldsymbol{\beta}_{-\alpha}|\tilde{\Sigma}, g, m \sim N(\mathbf{0},\, g(X^T\tilde{\Sigma}X)^{-1}); \tag{3.6}$$

and in both cases,

$$g \sim \mathrm{IG}\big(\frac{1}{2},\, \frac{N}{2}\big). \tag{3.7}$$

3.2.3 Fractional Bayes Factors and Posterior Model Probabilities

Note that if we were to form a standard Bayes factor comparing a homoscedastic and heteroscedastic model, the constants associated with the noninformative priors would not cancel one another and the Bayes factor would be defined only up to an unspecified constant. Thus we invoke a fractional Bayes factor approach to compute well-defined posterior model probabilities for each model; for a thorough review see O'Hagan [1995], and see Metzger and Franck [2019] for a discussion of fractional Bayes factors in the context of this problem.

Let $\mathcal{M}$ represent the full set of models under consideration, representing all classes and grouping schemes of interest. Denote $\boldsymbol{\theta}$ as the full set of unknown parameters associated with a model $m \in \mathcal{M}$ and $\pi(\boldsymbol{\theta})$ as the joint prior on these parameters. A fractional Bayes is a ratio of two fractional marginal model probabilites, where a fractional marginal likelihood is defined as

$$q^b(\boldsymbol{Y}|\boldsymbol{\theta}) = \frac{\int P(\boldsymbol{Y}|\boldsymbol{\theta}, m)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int P(\boldsymbol{Y}|\boldsymbol{\theta}, m)^b\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \tag{3.8}$$

for a some fractional exponent b. Although O'Hagan [1995] provides several recommendations for choice of b, `slgf` exclusively implements $b = \frac{m_0}{N}$ where m_0 is the minimal training sample size required for $P(\boldsymbol{Y}|m)$ to be a proper distribution. Thus we must compute the integrals $\int P(\boldsymbol{Y}|\boldsymbol{\theta}, m)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ and $\int P(\boldsymbol{Y}|\boldsymbol{\theta}, m)^b\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$ for all $m \in \mathcal{M}$. In the case of noninformative regression priors, $\boldsymbol{\beta}$ is integrated analytically, and σ^2 or σ_1^2, σ_2^2 are integrated using a Laplace approximation after a log-variance transformation. In the Zellner-Siow mixture g-prior case, α and $\boldsymbol{\beta}_{-\alpha}$ are integrated analytically, and σ^2 or σ_1^2, σ_2^2 and g are again integrated using a Laplace approximation with a log-variance transformation. Let $\tilde{\boldsymbol{\theta}}$ represent the set of unknown parameters after the regression effects $\boldsymbol{\beta}$ have been integrated out. Then for dimensionality $d = 2$ in the noninformative prior case and $d = 3$ in the Zellner-Siow mixture g-prior case,

$$\log\left(\int P(\boldsymbol{Y}|\tilde{\boldsymbol{\theta}}, m)\pi(\tilde{\boldsymbol{\theta}})d\tilde{\boldsymbol{\theta}}\right) \approx \frac{d}{2}\log(2\pi) - \frac{1}{2}\log|-H^\star| + \log(P(\boldsymbol{Y}|\tilde{\boldsymbol{\theta}}^\star, m)) \tag{3.9}$$

where $\tilde{\boldsymbol{\theta}}^\star$ is the mode of $P(\boldsymbol{Y}|\tilde{\boldsymbol{\theta}}, m)\pi(\tilde{\boldsymbol{\theta}})$ and $H^\star$ is the Hessian matrix evaluated at $\tilde{\boldsymbol{\theta}}^\star$. We compute these values with the functions `optim` and `numDeriv::hessian`, respectively. We make a similar computation for $\int P(\boldsymbol{Y}|\tilde{\boldsymbol{\theta}}, m)^b \pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}}$ to compute the fractional marginal model probability $q^b(\boldsymbol{Y}|\boldsymbol{\theta})$ for all $m \in \mathcal{M}$, well defined for both homoscedastic and heteroscedastic models. Once log-fractional marginal likelihoods have been computed for all models, we subtract the maximum from this set so that the set of log-fractional marginal likelihoods has been rescaled to have a maximum of 0. Each value is exponentiated to obtain a set of fractional marginal likelihoods with maximum 1. This adjustment helps to avoid numerical underflow when computing posterior model probabilities.

3.2.4 Model Priors

With this adjusted set of fractional marginal likelihoods, we next consider the priors for the model space. The package `slgf` imposes a uniform prior by model class, and for classes containing multiple models, the prior on each class is uniformly divided among the models it contains. We finally compute posterior model probabilities for each model:

$$P(m'|\boldsymbol{Y}) = \frac{P(\boldsymbol{Y}|m')P(m')}{\sum\limits_{\mathcal{M}} P(\boldsymbol{Y}|m)P(m)} \tag{3.10}$$

3.2.5 Parameter Estimation

We compute estimators for the regression parameters, variance(s), and g (when the Zellner-Siow mixture g-prior is used). Estimation of the regression parameters is straightforward:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}} P(\boldsymbol{Y}|X, \boldsymbol{\beta}, \Sigma) \tag{3.11}$$

so that $\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{Y}$. Estimation of $\hat{\Sigma}$ and g is more intensive. Recall we transform $\lambda = \log \sigma^2$ for one or both variances included in a model; let $\hat{\boldsymbol{\lambda}} := \{\hat{\lambda}\}$ in homoscedastic

models or $\{\hat{\lambda}_1, \hat{\lambda}_2\}$ in heteroscedastic models. Then

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \int P(\boldsymbol{Y}|X, \boldsymbol{\beta}, \Sigma) P(\boldsymbol{\beta}) P(\Sigma) d\boldsymbol{\beta} \tag{3.12}$$

or,

$$\{\hat{\boldsymbol{\lambda}}, \hat{g}\} = \arg\max_{\boldsymbol{\lambda}, g} \int P(\boldsymbol{Y}|X, \boldsymbol{\beta}, \Sigma, g) P(\alpha) P(\boldsymbol{\beta}_{-\alpha}|\Sigma) P(g) d\boldsymbol{\beta} \tag{3.13}$$

Then, $\hat{\boldsymbol{\sigma}}^2 = \exp\{\hat{\boldsymbol{\lambda}}\}$ for $\hat{\boldsymbol{\sigma}}^2 = \{\hat{\sigma}^2\}$ or $\hat{\boldsymbol{\sigma}}^2 = \{\hat{\sigma}_1^2, \hat{\sigma}_2^2\}$.

3.3 Functionality

The function `lm.slgf` requires several inputs to compute and output posterior model probabilities for all models and model classes of interest. The user begins with a dataframe containing a continuous response, at least one categorical predictor, and any other covariates of interest. To prevent confusion with the SLGF methodology, this dataframe cannot contain column names with the character string "group". The user must first identify a suspected latent grouping factor, usually by plotting the data and noting a latent structure within the levels of a categorical predictor as illustrated in Section 3.1. The user indicates, via the arguments `response` and `lgf`, character strings corresponding to the response and the suspected latent grouping factor variable names, respectively.

Next the user determines the model classes they wish to evaluate. We note the distinction between these model classes and the R class of a variable. The argument `usermodels` is a list where each element contains a string of R class `formula` or `character`. The user also specifies which classes should also be considered in a heteroscedastic context via the argument `het`, which provides an indicator 1 or 0 corresponding to each model class specified in `usermodels`. Together the arguments `usermodels` and `het` create the full set of model classes considered.

Next the user chooses a prior to place on the regression effects. As described in Section 3.2.2, `prior="flat"` (the default) implements the noninformative prior and `prior="zs"`

imposes the Zellner-Siow mixture g-prior.

Finally the user must specify the minimum number of levels of the SLGF that can comprise a group, via the argument `min.levels`, which defaults to 1. Because the number of possible grouping schemes increases exponentially with K, the user can reduce the number of candidate models, and hence speed up the computation, by increasing `min.levels`; however, because we partition into two groups, note `min.levels` may not exceed $\frac{K+1}{2}$. Additionally, when considering data with limited degrees of freedom, increasing `min.levels` can also ensure estimability of the specified `usermodels`; see Section 3.4.3 for more detail.

3.4 Illustrations

We next demonstrate the implementation of this methodology with the R package `slgf`.

3.4.1 ANOVA

First we revisit the dataset (denoted `smell`) analyzed by O'Brien and Heft [1995], who measured olfactory acuity `olf` on a continuous scale as a function of age `agecat`, where age groups were divided into five categorical levels. We note that levels 4 and 5 of the categorical predictor, age category, appear to have higher variance than levels 1, 2, and 3, but most standard analyses assume homoscedasticity. We consider the usual analysis comparing the null model, with a single mean, versus the alternative model, with 4 degrees of freedom for the mean effects, with homoscedastic error variance. Note we obtain maximum likelihood estimates for the single variance of $\hat{\sigma}^2_{\text{Null}} = 0.04334$ and $\hat{\sigma}^2_{\text{Alt}} = 0.03211$ for the null and alternative models, respectively. Instead consider `agecat` as the suspected latent grouping factor. The apparent latent grouping structure we observe is denoted 1,2,3:4,5 (or equivalently, 4,5:1,2,3). The means may also differ by level, but this is more difficult to distinguish by the plot. Thus we first consider the following model and heteroscedasticity

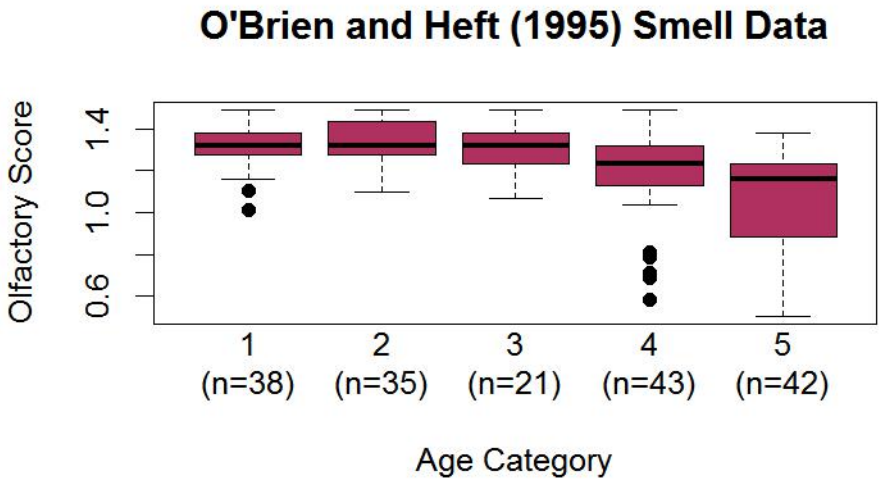

Figure 3.2: O'Brien and Heft [1995] studied olfactory function by age (y-axis), where age was divided into five categories (x-axis). The data suggests potential heteroscedasticity, with latent grouping structure 1,2,3:4,5.

structures: `smell.models <- list("olf∼1", "olf∼agecat", "olf∼group")` `smell.het <- c(0, 0, 1)` This elicits four model classes: a homoscedastic global mean, homoscedastic age level means, homoscedastic group-based means, and group-based means with group-based heteroscedasticity. Finally we note that with a relatively large sample size, we prefer the use of noninformative priors via `prior="flat"`, and we specify `min.levels=1`, as we have no prior information on the number of levels of `agecat` that may be grouped together and we wish to consider a comprehensive set of candidate models. The number of unique classes can always be obtained as `length(usermodels)+sum(het)`.

```
smell.out <- lm.slgf(df=smell, response="olf", lgf="agecat",
usermodels=smellmodels, prior="flat", het=c(0,0,1), min.levels=1)
```

The output is a list of class `slgf`, with six elements when `prior="flat"` and seven when `prior="zs"`:

- `results`, an $M \times 11$ matrix containing the model selection results and information

for each model, including `Model`, the `formula` associated with each model; `Scheme`, the grouping scheme associated with each model; `Variance`, a string of either `Homosk` or `Heterosk` associated with each model; `logFlik`, the log-factional marginal likelihood associated with each model; `Mod.Prior`, the model prior associated with each model; `Fmodprob`, the posterior fractional model probability associated with each model; `Cumulative`, the cumulative posterior model probability of the models aggregated in descending order; `df.Index`, an indicator of which element of `group.dfs` contains the corresponding `data.frame` associated with each model's scheme; `mle.index`, `mle.index`, an indicator of which element of `coefficients`, `variances`, and `gs` contains the estimators associated with each model; `Model.Index`, an indicator of each model's rank by posterior model probability; and `Class`, denoting the model class associated with each model

- `group.dfs`, a list containing dataframes associated with each model class containing the appropriate effects, including group effects
- `scheme.Probs`, a $S \times 1$ `data.frame` containing the total probability for each grouping scheme considered
- `class.Probs`, a $C \times 1$ `data.frame` containing the total probability for each model class considered
- `coefficients`, MLEs for each model's regression effects
- `variances`, REML estimates for each model's variance(s)
- `gs`, REML estimates for each model's g; only included if `prior="zs"`

We summarize the five most probable models:

```
> smell.out$results[1:5,c(1:3,6,11)]
       Model    Scheme Variance modprob.FBF                Class
```

```
1  olf~group 4,5:1,2,3 Heterosk       1e+00 olf~group, Heterosk
2 olf~agecat      None   Homosk       1e-08  olf~agecat, Homosk
3  olf~group 5:1,2,3,4 Heterosk       0e+00 olf~group, Heterosk
4  olf~group 4:1,2,3,5 Heterosk       0e+00 olf~group, Heterosk
5  olf~group 4,5:1,2,3   Homosk       0e+00   olf~group, Homosk
```

Note we overwhelmingly favor the model with group-based mean effects and variances via scheme 4,5:1,2,3. Recall that our variance MLEs were $\hat{\sigma}^2_{\text{Null}} = 0.04344$ and $\hat{\sigma}^2_{\text{Alt}} = 0.03211$. However, with our method we obtain two REML estimators:

```
> smell.hats <- extract.hats(smell.out, model.index=1)
> print(smell.hats$`sigsq.4,5`)
[1] 0.05869897
> print(smell.hats$`sigsq.1,2,3`)
[1] 0.01211161
```

The function `extract.hats` provides the estimators for the coefficient(s) and variance(s) for a given `model.index`, as well as g if `prior="zs"`. That is, we compute REML estimates $\hat{\sigma}^2_{1,2,3} = 0.05869897$ and $\hat{\sigma}^2_{4,5} = 0.01211161$. This suggests that the standard methods overestimated the variance for three levels of the categorical predictor, and underestimated the variance for the other two.

Let us also consider the case where `het=c(1,1,1)`; that is, we include two additional classes: group-based variances and a single global mean, and group-based variances with means by `agecat`.

```
> smell.out <- lm.slgf(df=smell, response="olf", lgf="agecat",
usermodels=smellmodels, prior="flat", het=c(1,1,1), min.levels=1)
> smell.out$results[1:5,c(1:3,6,11)]
       Model    Scheme Variance modprob.FBF                Class
```

```
1  olf~group 4,5:1,2,3 Heterosk  0.76422666  olf~group, Heterosk
2 olf~agecat 4,5:1,2,3 Heterosk  0.23577298 olf~agecat, Heterosk
3 olf~agecat 1,2:3,4,5 Heterosk  0.00000012 olf~agecat, Heterosk
4 olf~agecat 5:1,2,3,4 Heterosk  0.00000010 olf~agecat, Heterosk
5 olf~agecat 1:2,3,4,5 Heterosk  0.00000007 olf~agecat, Heterosk
```

Now the most probable models are a bit less conclusive, as the distinct category-means model with scheme 4,5:1,2,3 group-based heteroscedasticity accounts for a meaningful amount of posterior model probability. We can easily summarize the scheme and class probabilities, which overwhelmingly favor scheme 4,5:1,2,3 and moderately favor the group-based means and variances model class:

```
> smell.out$scheme.Probs
          Scheme.Prob
4,5:1,2,3  0.99999968
1,2:3,4,5  0.00000012
5:1,2,3,4  0.00000010
1:2,3,4,5  0.00000007
2:1,3,4,5  0.00000001
None       0.00000001
1,3:2,4,5  0.00000000
1,4:2,3,5  0.00000000
1,5:2,3,4  0.00000000
2,3:1,4,5  0.00000000
2,4:1,3,5  0.00000000
2,5:1,3,4  0.00000000
3,4:1,2,5  0.00000000
3,5:1,2,4  0.00000000
3:1,2,4,5  0.00000000
```

```
4:1,2,3,5  0.00000000
> smell.out$class.Probs
                     Class.Prob
olf~group, Heterosk  0.76422666
olf~agecat, Heterosk 0.23577328
olf~1, Heterosk      0.00000004
olf~agecat, Homosk   0.00000001
olf~1, Homosk        0.00000000
olf~group, Homosk    0.00000000
```

3.4.2 Analysis of Covariance (ANCOVA)

Next consider the data of Flurry [1939] in which the breaking strength of a starch chip (measured in grams) is analyzed according to the thickness of the chip (measured in 10^{-4} inches) and the type of starch used to create the chip (canna, corn, or potato starch). As usual, we begin by plotting the data to ascertain whether there is a latent grouping factor present. By inspection we note that the potato chips, represented by squares, appear to have a higher variability than the corn (triangles) and canna (squares) chips. A researcher's first approach might be analysis of covariance (ANCOVA), in which three parallel lines for each level of starch would be fit with a common error variance:

```
break.anc <- lm(chips$strength ~ chips$starch + chips$film)
break.coefs <- break.anc$coefficients
```

This leads to the following fit and residuals: We also note that a researcher's first inclination when faced with this classic cone-shaped residual pattern is to use a logarithmic transformation on the response. Let us implement this transformation and consider whether it has improved the apparent heteroscedasticity within the data:

```
logbreak.anc <- lm(log(chips$strength) ~ chips$starch + chips$film)
```

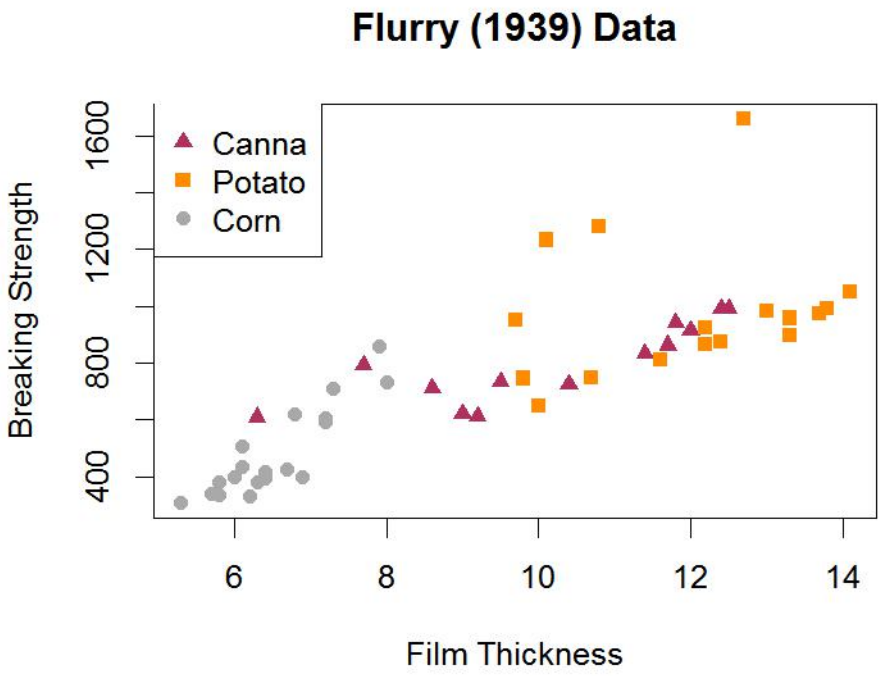

Figure 3.3: The Flurry [1939] data suggests potential heteroscedasticity or a latent group interaction, with scheme potato:corn,canna.

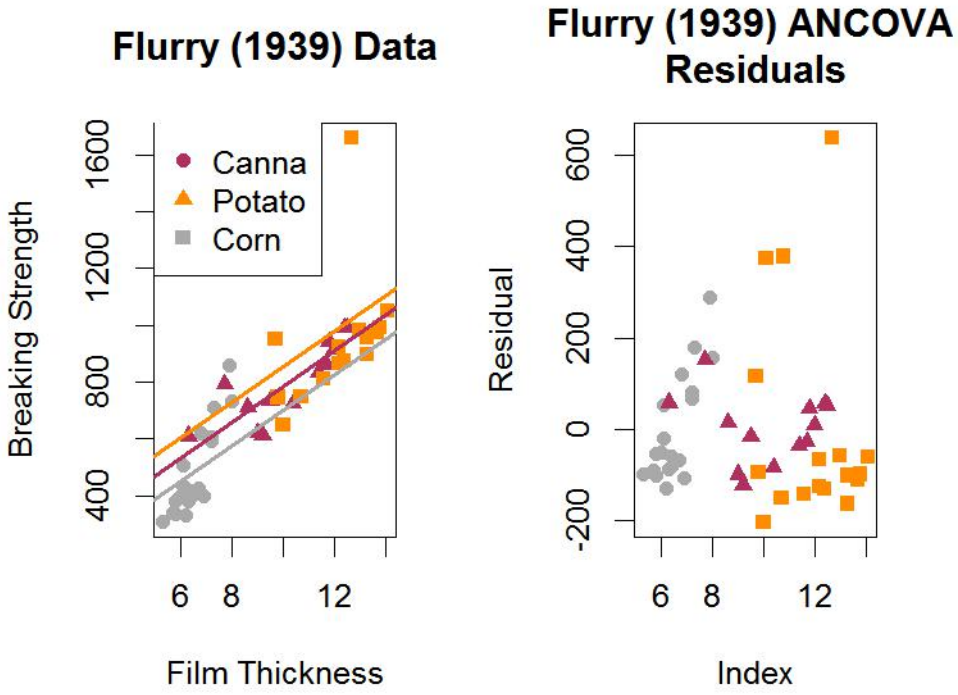

Figure 3.4: An ANCOVA model fit to the Flurry [1939] data leads to a problematic residual pattern.

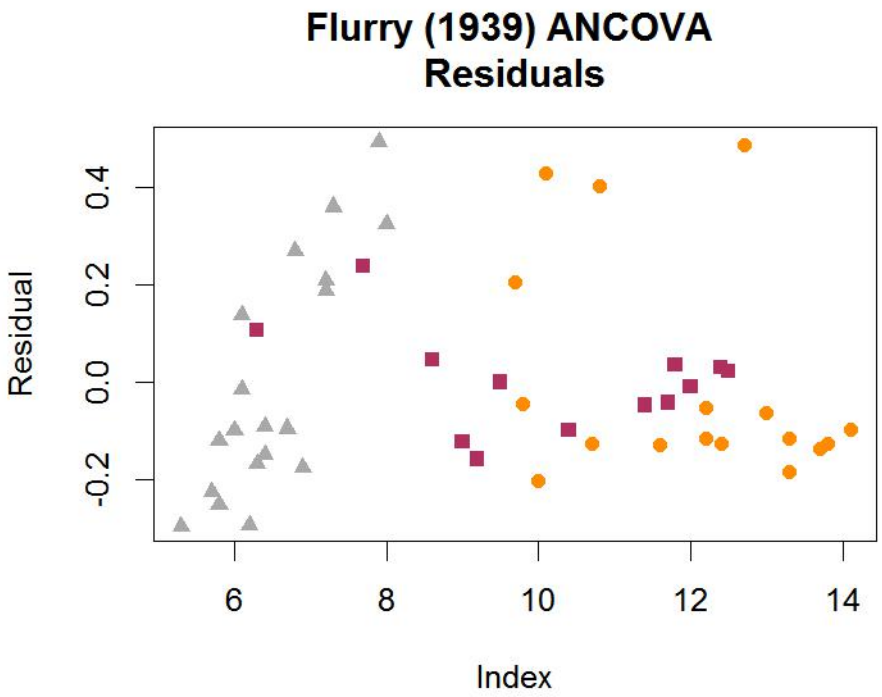

Figure 3.5: Residuals on the log-transformed Flurry [1939] data set show an alleviated cone shape, but the heteroscedasticity pattern remains by the levels of `starch`.

Although the conical shape appears to have been alleviated, we note that a more subtle pattern remains. As film thickness increases, the canna chips do not appear to show a similar increase in variance, so a log-transformation has failed to account for the variance structure that is a function of the categorical covariate, rather than the continuous film thickness. Thus we instead consider starch as a latent grouping factor. Additionally, we consider models with and without interactions between the categorical and continuous predictors:

```
break.models <- list("strength~1", "strength~film",  "strength~film+starch",
"strength~film+group", "strength~film+starch+film*starch",
"strength~film+group+film*group")
break.out <- lm.slgf(df=chips, response="strength", lgf="starch",
usermodels=break.models, het=c(0,0,0,1,0,1), prior="flat", min.levels=1)
```

We note that the posterior model probabilities favor group-based heteroscedasticity with the scheme canna,corn:potato:

```
> break.out$results[1:5, c(1:3,6)]
                            Model            Scheme Variance modprob.FBF
1              strength~film+group potato:canna,corn Heterosk  0.68750513
2   strength~film+group+film*group potato:canna,corn Heterosk  0.27416483
3                    strength~film              None   Homosk  0.01314140
4   strength~film+group+film*group corn:canna,potato   Homosk  0.00864297
5 strength~film+starch+film*starch              None   Homosk  0.00455575
> break.out$scheme.Probs
                  Scheme.Prob
potato:canna,corn  0.96646025
None               0.02066285
corn:canna,potato  0.01080258
canna:corn,potato  0.00207431
> break.out$class.Probs
                                         Class.Prob
strength~film+group, Heterosk            0.68750513
strength~film+group+film*group, Heterosk 0.27416483
strength~film, Homosk                    0.01314140
strength~film+group+film*group, Homosk   0.01180418
strength~film+group, Homosk              0.00586300
strength~film+starch+film*starch, Homosk 0.00455575
strength~film+starch, Homosk             0.00296570
strength~1, Homosk                       0.00000000
```

Thus we strongly conclude group-based heteroscedasticity with scheme potato:canna,corn along with main effects for `film` and `group`. The inclusion of a `film*group` interaction is

less certain; we briefly revisit this uncertainty in Chapter **??**.

3.4.3 Two-way Replicated Layouts

We revisit the two-way replicated layout of Meek and Ozgur [1991] illustrated in the right panel of Figure 3.1, where the torque required to tighten a locknut was measured as a function of a plating process and a threading method. As usual, we begin by plotting the data to assess for a latent grouping structure in Figure 3.6. Perhaps the first characteristic we note is

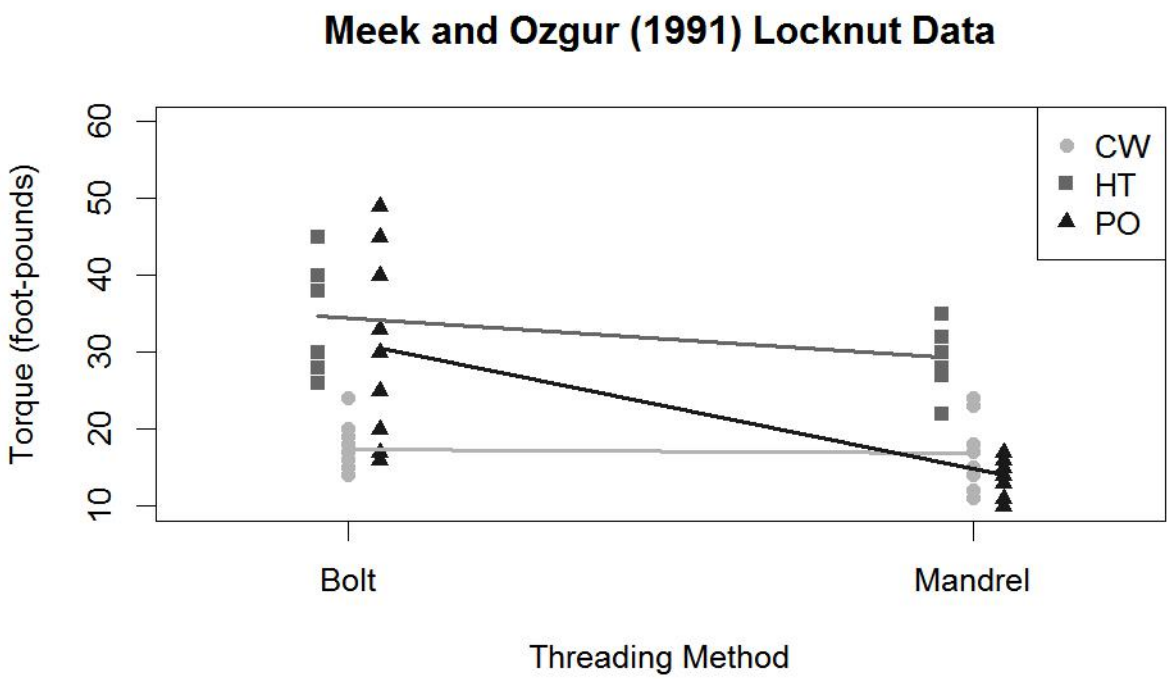

Figure 3.6: The torque (y-axis) required to tighten a locknut is presented as a function of two predictors. First, the fixture on which the nut was threaded, either a bolt or a conical mandrel, and second, one of three plating methods including cadmium and wax, heat treating, and phosphate and oil, denoted CW, HT, and PO, respectively. Lines connect the means of the plating treatments at each level of fixture. The observations at level bolt appear to have a higher error variance than those of mandrel.

that the variance of the torque measured on locknuts fitted on bolts appear higher than those fitted on a mandrel. If we neglected to account for this potential heteroscedasticity, and fit

only a standard homoscedastic interaction model, we would obtain the following coefficients and error variance estimate:

```
> locknut.twoway <- lm(Torque~Fixture+Plating+Fixture*Plating, data=locknut)
> locknut.twoway$coefficients
             (Intercept)           Fixturemandrel                PlatingHT
                    17.4                     -0.5                     17.3
Fixturemandrel:PlatingPO
                   -15.9
> summary(locknut.twoway)$sigma^2
[1] 36.57778
```

Instead we treat the fixture as the latent grouping factor. Note that with only two levels of the categorical predictor, there is only one trivial grouping scheme, bold:mandrel. Moreover, we can only consider distinct variances by group, as a group effect with one degree of freedom would be isomorphic to the standard fixture effect; thus it is important that we do not include the `group` term in the `usermodels` argument, but rather specify heteroscedastic models and `lgf="Fixture"`. Regarding the regression effects, there appears to be an interaction between fixture and plating, as indicated by the relatively non-parallel lines. Thus we consider `usermodels=list("Torque∼1", "Torque∼Fixture", "Torque∼Fixture+Plating", "Torque∼Fixture*Plating")` and `het=c(1,1,1,1)`. Because there are 10 replicates at each combination of factor levels of fixture and plating, we use the noninformative prior `prior="flat"`.

```
locknutmodels <- list("Torque~1",  "Torque~Fixture", "Torque~Fixture+Plating",
                      "Torque~Fixture+Plating+Fixture*Plating")
locknut.out <- lm.slgf(df=locknut, response="Torque", lgf="Fixture",
                       usermodels=locknutmodels, het=c(1,1,1,1),
                       prior="flat", min.levels=1)
```

With relatively few models, and only one trivial grouping scheme in heteroscedastic classes, we view all eight candidate models' posterior probability. Because each model class contains only a single model, each model probability is also its corresponding class probability:

```
> locknut.out$results[, c(1:3,6)]
                                   Model       Scheme Variance modprob.FBF
1 Torque~Fixture+Plating+Fixture*Plating bolt:mandrel Heterosk  0.97976751
2                 Torque~Fixture+Plating bolt:mandrel Heterosk  0.01655156
3 Torque~Fixture+Plating+Fixture*Plating         None   Homosk  0.00364617
4                 Torque~Fixture+Plating         None   Homosk  0.00003476
5                        Torque~Fixture         None   Homosk  0.00000000
6                        Torque~Fixture bolt:mandrel Heterosk  0.00000000
7                              Torque~1         None   Homosk  0.00000000
8                              Torque~1 bolt:mandrel Heterosk  0.00000000
> locknut.out$scheme.Probs
             Scheme.Prob
bolt:mandrel  0.99631907
None          0.00368093
```

The most probable model accounts for approximately 98% of the posterior mass, indicating distinct variances between bolt and mandrel, along with a fixture by plating interaction. Our estimates for the coefficients are equivalent to the least squares estimates obtained from `lm`:

```
> locknut.out$coefficients[[9]]
             (Intercept)           Fixturemandrel                 PlatingHT
                    17.4                     -0.5                      17.3
               PlatingPO Fixturemandrel:PlatingHT Fixturemandrel:PlatingPO
                    13.1                     -4.8                     -15.9
```

On the other hand, let us inspect this data more carefully. The most probable model, in the set `locknut.models` as parametrized earlier, is `Torque~Fixture+Plating+Fixture*Plating` with heteroscedasticity between bolt and mandrel. We observe the residuals of this model; for the sake of plotting the data efficiently, we append a factor onto the `locknut` data that denotes the interaction using either "b" or "m" for bolt and mandrel:

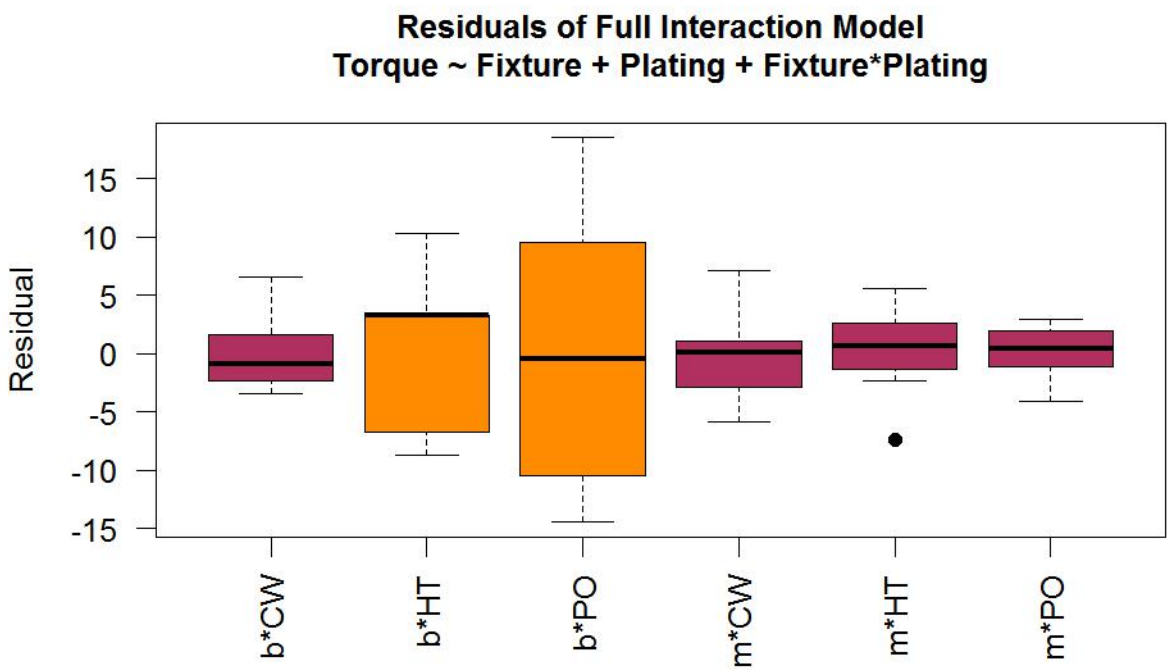

Figure 3.7: The residuals (y-axis) under the full interaction model Torque~Fixture+Plating + Fixture*Plating indicate there may be a more subtle heteroscedasticity structure than `bolt:mandrel`; we color-code for emphasis. We consider Fixture * Plating as the `lgf` instead.

```
locknut$`Fixture*Plating` <- paste0(substr(as.character(locknut$Fixture), 1, 1),
 "*", locknut$Plating)
```

It appears the latent grouping factor may in fact be the levels of the interaction itself, with equivalent means and distinct variances by scheme b*CW,m*CW,m*HT,m*PO:b*HT,b*PO. So alternatively we let `usermodels` contain only the full interaction model, and let

```
lgf="Fixture*Plating".

> locknut.int <- list("Torque~Fixture+Plating+Fixture*Plating")
> locknut.int.out <- lm.slgf(df=locknut, response="Torque",
lgf="Fixture*Plating", usermodels=locknut.int, het=1, prior="flat", min.levels=1)
> locknut.int.out$results[1,c(2,3,6)]
                         Scheme Variance modprob.FBF
1 b*HT,b*PO:b*CW,m*CW,m*HT,m*PO Heterosk   0.7939723
```

As expected from the residual boxplot, we place nearly 80% of the posterior mass on the model we suspected from the visual inspection. This illustrates the flexibility of the SLGF method, even allowing for latent groupings by the levels of an interaction effect. Additionally, our method can be implemented as a post-hoc analysis on the residuals of the levels of a categorical predictor based on an existing model.

3.4.4 Two-way Unreplicated Layouts

Next we restrict our two-way approach to an unreplicated context, which was the inspiration for the SLGF method. Consider the data analyzed by Franck [2018], where six dogs with lymphoma were studied. Two individual samples were taken from healthy and tumor tissue within each dog, and the copy number variation was measured for each sample. Without loss of generality, we arrange dogs into rows and tissue types into columns of a two-way layout. We first plot the data to determine whether a latent grouping structure underlies the data: We strongly suspect that dogs 1, 2, and 5 behave distinctly from dogs 3, 4, and 6. The telltale non-parallel lines suggest an underlying interaction, but with only a single observation in each cell, we lack the degrees of freedom to fit a standard row by column interaction term. Instead, we let `dog` represent the SLGF, exclude the column effect, and specify a group-by-column interaction term, isomorphic to fitting distinct column effects by group. Thus we consider four reasonable model classes: a dog and tissue effect, a dog and tissue-by-column

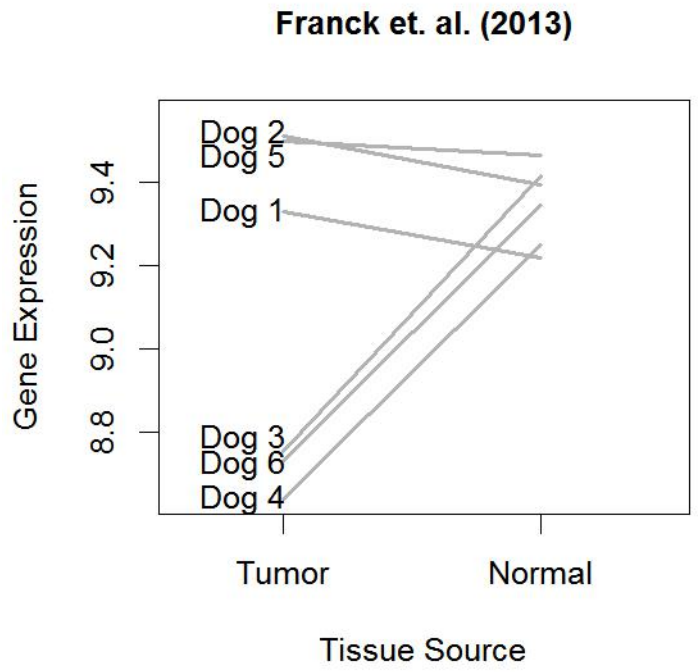

Figure 3.8: The data analyzed by Franck et al. [2013] measured a continuous gene expression response (y-axis) in six dogs with lymphoma. Tissue samples were taken from tumor and normal tissue in each dog.

interaction, and the heteroscedastic counterparts of each class. We accomplish this with the arguments `usermodels=list("gene~dog+tissue", "gene~dog+group:tissue")` and `het=c(1,1)`. Additionally, we must specify `min.levels=2` or `min.levels=3` to ensure sufficient degrees of freedom to estimate the `group:col` effect. Here we consider `min.levels=2` in the interest of assessing a more complete set of candidate models.

Because of the limited amount of data, our choice of prior is more impactful in this case. We impose `prior="zs"` to utilize the Zellner-Siow mixture g-prior, as the fractional Bayes factor exponent would require a prohibitively high proportion of the data for model training.

We first put the two-way layout into a `data.frame` format compatible with the `slgf` function, and then implement this approach:

```
> lymphoma.tall <- maketall(lymphoma)
> lymphoma.tall <- data.frame("gene"=lymphoma.tall[,1], "dog"=lymphoma.tall[,2],
 "tissue"=lymphoma.tall[,3])
> lymphoma.models <- list("gene~dog+tissue", "gene~dog+group:tissue")
> lymphoma.out <- lm.slgf(df=lymphoma.tall, response="gene", lgf="dog",
 usermodels=lymphoma.models, prior="zs", het=c(1,1), min.levels=2)
```

Note the `:` operator in the `usermodels` syntax, which does not automatically include the main effects `group` and `tissue` which are not both estimable. As expected, we conclude with high probability that scheme 1,2,5:3,4,6 underlies the data. The five most probable models are given by:

```
> lymphoma.out$results[1:5, c(1:3,6)]
                  Model      Scheme Variance modprob.FBF
1 gene~dog+group:tissue 1,2,5:3,4,6 Heterosk  0.77731156
2 gene~dog+group:tissue 1,2,5:3,4,6   Homosk  0.18489612
3 gene~dog+group:tissue 1,2:3,4,5,6 Heterosk  0.00683924
4 gene~dog+group:tissue 2,5:1,3,4,6 Heterosk  0.00580652
```

```
5       gene~dog+tissue        None   Homosk  0.00429596
```

while the class and five highest scheme probabilities are

```
> lymphoma.out$class.Probs
                                Class.Prob
gene~dog+group:tissue, Heterosk 0.79638854
gene~dog+group:tissue, Homosk   0.19069556
gene~dog+tissue, Heterosk       0.00861996
gene~dog+tissue, Homosk         0.00429596
> head(lymphoma.out$scheme.Probs, 5)
            Scheme.Prob
1,2,5:3,4,6  0.96631751
1,2:3,4,5,6  0.00790504
2,5:1,3,4,6  0.00723547
None         0.00429596
1,5:2,3,4,6  0.00365555
```

The `group.dfs` element of `lymphoma.out` contains `data.frames` associated with each model and grouping scheme. We first determine which element of `lymphoma.out$group.dfs` contains the `data.frame` of interest via the column `df.Index`:

```
> lymphoma.out$results[1,c(1:3,6,8)]
                  Model      Scheme Variance modprob.FBF df.Index
1 gene~dog+group:tissue 1,2,5:3,4,6 Heterosk   0.7773116       18
```

This tells us that element 18 of `lymphoma.out$group.dfs` contains the `data.frame` with the 1,2,5:3,4,6 group effect:

```
> lymphoma.out$group.dfs[[18]]
```

```
     gene dog tissue group
1  9.3278   1      1 1,2,5
2  9.2168   1      2 1,2,5
3  9.5108   2      1 1,2,5
4  9.3942   2      2 1,2,5
5  8.7535   3      1 3,4,6
6  9.4158   3      2 3,4,6
7  8.6372   4      1 3,4,6
8  9.2480   4      2 3,4,6
9  9.4981   5      1 1,2,5
10 9.4626   5      2 1,2,5
11 8.7322   6      1 3,4,6
12 9.3439   6      2 3,4,6
```

3.4.5 Balanced Incomplete Block Design (BIBD)

Finally we illustrate the utility of the Zellner-Siow mixture g-prior in another context with limited data, the balanced incomplete block design (BIBD). Recall the data described in the center panel of Figure 3.1 by Davies [1954], which measured the wear on tires as a function of the compound from which they were manufactured. Four tires were used in the study and treated as blocks. Four compounds were studied, but due to manufacturing limitations, each tire was divided into thirds and the sections made from three different compounds. Let us first visualize the data: We note that the abrasion appears higher for compounds C and D, and for tire 4. Compounds C and D behave quite similarly over tires as well, as indicated by their similar V-shapes. A standard frequentist analysis considers a linear model with factors `compound` and `as.factor(tire)`:

```
> roadwear.lm <- lm(abrasion ~ compound + as.factor(tire), data=roadwear)
> anova(roadwear.lm)
```

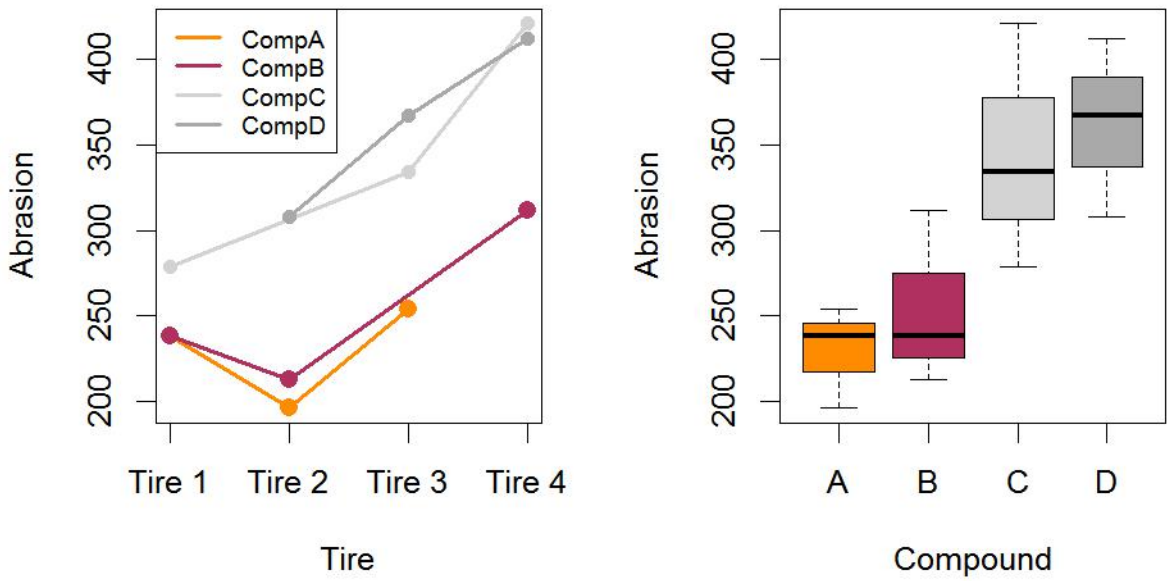

Figure 3.9: The data analyzed by Davies [1954] measured the abrasion on four tires (blocks), each manufactured from three of four compounds, in a road wear test.

```
Analysis of Variance Table

Response: abrasion
                Df Sum Sq Mean Sq F value    Pr(>F)
compound         3  38814 12938.0  36.946 0.0007763 ***
as.factor(tire)  3  21038  7012.6  20.026 0.0032406 **
Residuals        5   1751   350.2
```

The analysis of variance indicates that the blocking factor `tire` was and important inclusion to account for variability in the data. Additionally, compound is a significant source of variation in the model. A reasonable follow-up is to conduct multiple comparisons with a post-hoc pairwise comparisons method; here we illustrate the Tukey's honest significant difference method, which indicates that compounds A and B are statistically equivalent, C and D are statistically equivalent, but all other pairwise comparisons (A vs. C, A vs. D, B vs. C, and B vs. D) show a significant difference.

```
> TukeyHSD(aov(roadwear.lm), "compound")
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = roadwear.lm)

$`compound`
         diff       lwr       upr     p adj
B-A  25.00000 -31.37911  81.37911 0.4385215
C-A 115.33333  58.95422 171.71244 0.0024729
D-A 133.00000  76.62089 189.37911 0.0012750
C-B  90.33333  33.95422 146.71244 0.0074160
D-B 108.00000  51.62089 164.37911 0.0033397
```

```
D-C  17.66667 -38.71244  74.04578 0.6755282
```

Alternatively, we consider our proposed method with `compound` treated as the SLGF:

```
roadwear.models <- list("abrasion~1", "abrasion~tire", "abrasion~compound",
"abrasion~group", "abrasion~tire+compound", "abrasion~tire+group")
roadwear.out <- lm.slgf(df=roadwear, response="abrasion", lgf="compound",
 usermodels=roadwear.models, het=c(0,0,1,1,1,1), prior="zs", min.levels=1)
```

Many models receive a non-negligible proportion of the posterior model probability. The ten most probable models are given by:

```
> roadwear.out$results[1:10, c(1:3, 6)]
                    Model  Scheme Variance modprob.FBF
1     abrasion~tire+group A,B:C,D Heterosk  0.24842160
2  abrasion~tire+compound    None   Homosk  0.21495758
3     abrasion~tire+group A,B:C,D   Homosk  0.12682158
4  abrasion~tire+compound D:A,B,C Heterosk  0.05072768
5  abrasion~tire+compound A,B:C,D Heterosk  0.04032375
6  abrasion~tire+compound A:B,C,D Heterosk  0.04031812
7  abrasion~tire+compound A,C:B,D Heterosk  0.03974173
8  abrasion~tire+compound C:A,B,D Heterosk  0.03718444
9  abrasion~tire+compound A,D:B,C Heterosk  0.03625826
10 abrasion~tire+compound B:A,C,D Heterosk  0.03478117
```

Note the most probable grouping structure, A,B:C,D, is the same one implied by the results of the TukeyHSD post-hoc comparison. However, note that nearly 60% of the posterior class probability belongs to heteroscedastic classes:

```
> roadwear.hetprob <- sum(roadwear.out$class.Probs[which(grepl("Heterosk",
```

```
rownames(roadwear.out$class.Probs))),])
> print(roadwear.hetprob)
[1] 0.5989437
```

In this context, our fully Bayesian analysis mostly agrees with the standard frequentist approach, but the package `slgf` easily allows the user to explore a red flag regarding potential latent heteroscedasticity.

3.5 Summary

We provide a user-friendly and intuitive R package, `slgf`, for researchers to implement the SLGF methodology in practice. The SLGF approach provides an intuitive, easy to implement analysis that has shown to be reasonable and beneficial in many common linear model examples. By partitioning the levels of a categorical predictor into two groups, we can detect latent structures including group-based regression effects, interactions, heteroscedasticity, and combinations of these. By selecting probable model classes and partitioning schemes with common priors, fractional Bayes factors, and Bayesian models selection, we provide a straightforward Bayesian approach. The package uses syntax similar to the `lm` function and allows for users to specify a broad range of model parametrizations that are highly customizable for varying needs, scientific contexts, and prior knowledge.

Chapter 4

Conclusions

4.1 Summary

This work has aimed to provide an intuitive, flexible, mathematically and computationally feasible alternative to classical linear model analysis methods. Through the use of a combinatoric grouping approach, fractional Bayes factors, and Bayesian model selection, we can detect a wide variety of latent grouping structures underlying a dataset.

We first noted that there is a lack of literature addressing the specific concerns of latent group-based effects describe here. A foundation was then laid supporting our approach, including derivation of the posterior model probabilities, a fractional Bayes factor approach that accounts for our noninformative comparison between homoscedastic and heteroscedastic models, and a computational strategy for efficient and accurate calculations. We have demonstrated the applicability of our method through numerous real datasets that manifest the group-based effects we have described, illustrating a great utility for this work in a variety of applications. A simulation study was also undertaken to show our method's ability to detect a range of group-based effects in linear models of diverse types. Finally, an intuitive R package, `slgf`, was written and thoroughly described so that researchers from many fields,

particularly those outside of statistics, might be able to implement our methodology on their data.

4.2 Future Work

Future avenues of research are numerous and promising. We revisit the data of Flurry [1939] in Figure 4.1 to motivate our first future goal. Recall that we placed approximately 69% posterior model probability on the model with the film and a group-based effect, and group-based variances, with scheme canna,corn:potato. The next most probable model had the same heteroscedasticity scheme, with a film by group interaction as well. Thus we are quite confident in the group-based heteroscedasticity, but the regression effects are less certain. As we inspect Figure 4.1, we suspect that an issue may be that corn actually has a distinct regression effect from canna and potato. We thus propose a more flexible SLGF approach whereby there are two distinct suspected latent grouping factors: a regression effect SLGF and a variance SLGF. This could allow us not only to model the regression effects and variances with completely different categorical effects, but also do so with the same categorical effect but distinct schemes. We are optimistic that the existing methodology and R code is sufficient to make this extension with relative simplicity.

Next, we hope to explore statistical inference on parameters under the SLGF framework; promising preliminary results have been described by Franck [2018]. In many cases, the most difficult component of model averaging is computing model weights; that is, posterior model probabilities for the model in question. We believe the method described by Hoeting et al. [1999] has promise in our context.

A third future goal is to continually improve the R package. The R package `slgf` was developed over several months by aggregating three separate functions that had originally been written to accommodate only one-way ANOVA, ANCOVA, and two-way unreplicated data, respectively. These functions were then generalized so that they could accommodate

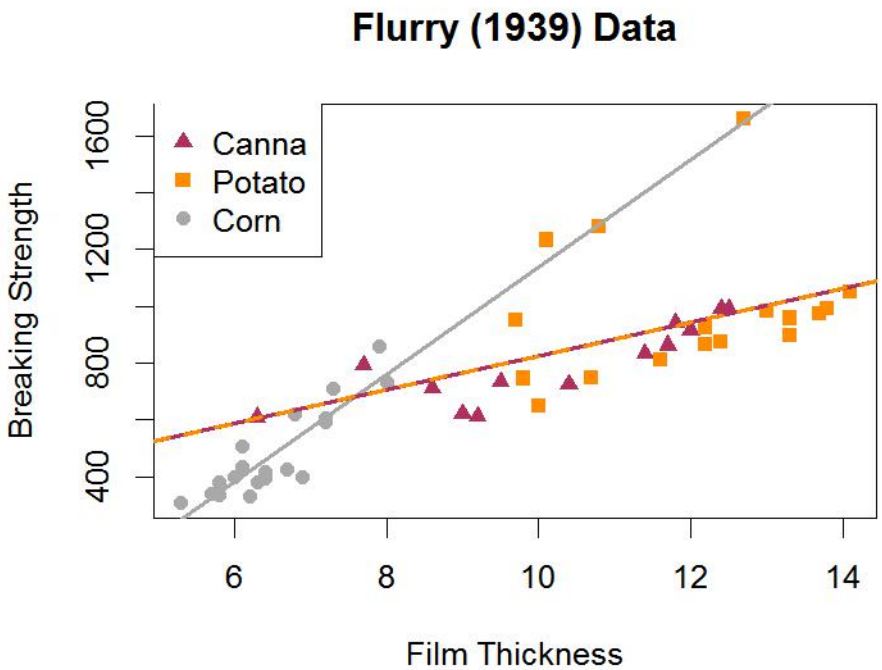

Figure 4.1: On closer examination of the Flurry data [Flurry, 1939], it appears that the latent heteroscedasticity by scheme canna,corn:potato can be improved upon if we also model regression effects via the scheme corn:canna,potato.

a broader class of linear models with a categorical predictor, as we have carefully developed our method so that it is very general and flexible. In its current form, although user-friendly and functional, there are many places to make the code and computations more efficient. For example, a first sequence computes the design matrices and regression effect MLEs for each model. Next a second sequence performs two tasks: it first estimates the variance and, if necessary, g, for each model, and then computes the posterior model probability for each model. This computation requires a relatively costly optimization over two or three dimensions, possibly for many model classes and grouping schemes. These sequences could be combined into one `for` loop, and, more efficient methods of optimization could be employed.

Generalization to three latent groups is another reasonable extension. Mathematically, this would be straightforward and could allow for more complex grouping structures to be accounted for. We are currently in search of datasets that reasonably illustrate such a latent grouping structure.

Finally, Gaussian process (GP) modeling is a burgeoning area of statistics. Latent categorical (qualitative) factors were recently discussed by Zhang et al. [2019], where they are modeled similarly to numeric variables with the goal of dimension reduction and eliciting parsimony. With our approach, we might consider distinct GPs according to the levels of a categorical predictor, perhaps with distinct group-based lengthscales, covariance functions, or nuggets.

It is our belief that the SLGF methodology outlined in this work can become a staple of many researchers' statistical analysis repertoires. Because of its broad scope, clear application in real datasets, and straightforward implementation in practice, we are optimistic that this work can assist researchers in many fields.

Bibliography

M. S. Bartlett. Properties of sufficiency and statistical tests. *Proceedings of the Royal Statistical Society, Series A*, 160(901):268–282, May 1937.

James O. Berger and Luis R. Pericchi. The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122, 03 1996.

W. D. Berry. *Quantitative Applications in the Social Sciences: Understanding Regression Assumptions*. SAGE Publications Ltd., Thousand Oaks, CA, 1993.

P. J. Bickel. Using residuals robustly i: Tests for heteroscedasticity, nonlinearity. *Annals of Statistics*, 6:266–291, 1978.

W. John Boscardin and Andrew Gelman. Bayesian computation for parametric models of heteroscedasticity in the linear model, 1994.

G. E. Box and D. R. Cox. An analysis of transformations. *Journal of the Royal Statistical Society*, 26:211–252, 1964.

G. E. P. Box and W. J. Hill. Correcting inhomogeneity of variance with power transformation weighting. *Technometrics*, 16:385–389, 1974.

T. S. Breusch and A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica (pre-1986)*, 47(5):1287, 09 1979. URL `http://login.ezproxy.lib.vt.edu/login?url=https://search-proquest-com.`

ezproxy.lib.vt.edu/docview/214655264?accountid=14826. Copyright - Copyright Econometric Society Sep 1979; Last updated - 2011-10-07; CODEN - ECMTA7.

Morton B. Brown and Alan B. Forsythe. Robust tests for the equality of variances. *Journal of the American Statistical Association*, Vol 69(346):364–367, June 1974.

Raymond J. Carroll and David Ruppert. Robust estimation in heteroscedastic linear models. *The Annals of Statistics*, 10(2):429–441, June 1982.

Raymond J. Carroll and David Ruppert. *Transformation and weighting in regression*. Chapman and Hall, New York, 1988.

R. Dennis Cook and Sanford Weisberg. Diagnostics for heteroscedasticity in regression. *Biometrika*, 70(1):1–10, April 1983. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(85)90158-7. URL http://www.sciencedirect.com/science/article/pii/0304407685901587.

J. G. Cragg. More efficient estimation in the presence of heteroscedasticity of unknown form. *Econometrica*, 51(3):751–763, May 1983.

F. Cribari-Neto and S. G. Zarkos. Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, 18(2):211–228, 1999. doi: 10.1080/07474939908800440. URL https://doi.org/10.1080/07474939908800440.

Edilberto Cepeda Cuervo and Jorge Alberto Achcar. Regression models with heteroscedasticity using bayesian approach. *Revista Colombiana de Estadística*, 32(2):267–287, December 2009.

O. L. Davies. *The Design and Analysis of Industrial Experiments*. Oliver & Boyd, London, 1954.

John Deegan. The consequences of model misspecification in regression analysis. *Multivariate Behavioral Research*, 11(2):237 – 248, 1976. ISSN 0378-3758. doi: 10.1207/s15327906mbr1102_9. URL https://doi.org/10.1207/s15327906mbr1102_9.

Philippe J. Deschamps. On the estimated variances of regression coefficients in misspecified error components models. *Econometric Theory*, 7(3):369–384, September 1991.

Bianca Dumitrascu, Gregory Darnell, Julien Ayroles, and Barbara E. Engelhardt. A bayesian test to identify variance effects. *Cornell University Library*, 2015.

Friedhelm Eicker. Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 59–82, Berkeley, Calif., 1967. University of California Press. URL `https://projecteuclid.org/euclid.bsmsp/1200512981`.

M. S. Flurry. Breaking strength, elongation and folding endurance of films of starches and gelatin used in sizing. *Technical Bulletin (United States Department of Agriculture)*, 674: 1–36, 1939.

Christopher Franck. Detection of hidden additivity and inference under model uncertainty for unreplicated factorial studies via bayesian model selection and averaging. *Technometrics*, 0(ja):1–30, 2018. doi: 10.1080/00401706.2018.1518791. URL `https://doi.org/10.1080/00401706.2018.1518791`.

Christopher T. Franck and Jason A. Osborne. Exploring Interaction Effects in Two-Factor Studies using the hiddenf Package in R. *The R Journal*, 8(1):159–172, 2016. URL `https://journal.r-project.org/archive/2016/RJ-2016-011/index.html`.

Christopher T. Franck, Dahlia M. Nielsen, and Jason A. Osborne. A method for detecting hidden additivity in two-factor unreplicated experiments. *Computational Statistics & Data Analysis*, 67(Supplement C):95 – 104, 2013. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2013.05.002. URL `http://www.sciencedirect.com/science/article/pii/S0167947313001618`.

Paul Gilbert and Ravi Varadhan. *numDeriv: Accurate Numerical Derivatives*, 2016. URL `https://CRAN.R-project.org/package=numDeriv`. R package version 2016.8-1.

H. Glejser. A new test for heteroscedasticity. *Journal of the American Statistical Association*, 64(325):316–323, March 1969.

Stephen M. Goldfeld and Richard E. Quandt. Some tests for homoscedasticity. *Journal of the American Statistical Association*, 60(310):539–547, 1965. doi: 10.1080/01621459.1965.10480811. URL `http://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480811`.

H. O. Hartley. The maximum f-ratio as a short cut test for homogeneity of variance. *Biometrika*, 37(3/4):308–312, 1950.

Andrew F. Hayes and Li Cai. Using heteroskedasticity-consistent standard error estimators in ols regression: An introduction and software implementation. *Behavior Research Methods*, 39(4):709–722, 2007.

Clifford Hildreth and James P. Houck. Some estimators for a linear model with random coefficients. *Journal of the American Statistical Association*, 63(322):584–595, June 1968.

Jennifer A. Hoeting, David Madigan, Adrian E. Raftery, and Chris T. Volinsky. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and e. i. george, and a rejoinder by the authors. *Statist. Sci.*, 14(4):382–417, 11 1999. doi: 10.1214/ss/1009212519. URL `https://doi.org/10.1214/ss/1009212519`.

Peter J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 221–233, Berkeley, Calif., 1967. University of California Press. URL `https://projecteuclid.org/euclid.bsmsp/1200512988`.

J. D. Jobson and W. A. Fuller. Least squares estimation when the covariance matrix and parameter vector are functionally relate. *Journal of the American Statistical Association*, 75:176–181, 1980.

M. Kharrati-Kopaei and S. M. Sadooghi-Alvandi. A new method for testing interaction in unreplicated two-way analysis of variance. *Communications in Statistics - Theory and Methods*, 36(15):2787–2803, 2007. doi: 10.1080/03610920701386851. URL `http://dx.doi.org/10.1080/03610920701386851`.

Howard Levene. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, chapter Robust tests for equality of variances 1960, pages 278–292. Stanford University Press, 1960.

Feng Liang, Rui Paulo, German Molina, Merlise A Clyde, and Jim O Berger. Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423, 2008. doi: 10.1198/016214507000001337. URL `https://doi.org/10.1198/016214507000001337`.

J. Scott Long and Laurie H. Ervin. Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*, 54(3):217–224, 2000. doi: 10.1080/00031305.2000.10474549. URL `https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474549`.

James G MacKinnon and Halbert White. Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29 (3):305 – 325, 1985. ISSN 0304-4076. doi: https://doi.org/10.1016/0304-4076(85)90158-7. URL `http://www.sciencedirect.com/science/article/pii/0304407685901587`.

G.E. Meek and C.O. Ozgur. Torque variation analysis. *Journal of the Industrial Mathematics Society*, 41:1–16, 1991.

Thomas A. Metzger and Christopher T. Franck. Detection of latent heteroscedasticity and group-based regression effects in linear models via Bayesian model selection. *arXiv e-prints*, art. arXiv:1903.01035, Mar 2019.

Donald F. Morrison. *Applied Linear Statistical Methods*. Prentice-Hall, 1983.

R. G. O'Brien and M. W. Heft. New discrimination indexes and models for studying sensory functioning in aging. *Journal of Applied Statistics*, 22:9–27, 1995.

Anthony O'Hagan. Fractional bayes factors for model comparison. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):99–138, 1995. ISSN 00359246. URL `http://www.jstor.org/stable/2346088`.

R. E. Park. Estimation with heteroscedastic error terms. *Econometrica*, 34(4):888, 1966.

Frieder. Andreas Perthes. Carolus fridericus gauss theoria motus corporum coelestium in sectionibus conicis solem ambientium. *Astronomische Nachrichten*, 40(25):391–392, 1855. ISSN 1521-3994. doi: 10.1002/asna.18550402508. URL `http://dx.doi.org/10.1002/asna.18550402508`.

W. Polasek, S. Liu, and S. Jin. Heteroskedastic linear regression models a bayesian analysis. In Ingo Balderjahn, Rudolf Mathar, and Martin Schader, editors, *Classification, Data Analysis, and Data Highways*, pages 182–191, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg. ISBN 978-3-642-72087-1.

Wolfgang Polasek and Klaus Pötzelberger. Robust bayesian methods in simple anova models. *Journal of Statistical Planning and Inference*, 40(2):295 – 311, 1994. ISSN 0378-3758. doi: https://doi.org/10.1016/0378-3758(94)90127-9. URL `http://www.sciencedirect.com/science/article/pii/0378375894901279`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL `https://www.R-project.org/`.

Potluri Rao. Some notes on misspecification in multiple regressions. *The American Statistician*, 25(5):37–39, December 1971.

Alvin C. Rencher and G. Bruce Schaalje. *Linear Models in Statistics*. John Wiley and Sons, 2nd edition, 2008. ISBN 978-0-471-75498-5.

Henry Scheffé. *The Analysis of Variance*. Wiley Publications in Statistics, 1959.

John W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2): 99–114, 1949. ISSN 0006341X, 15410420. URL `http://www.jstor.org/stable/3001913`.

B. L. Welch. On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3/4):330–336, 1951. ISSN 00063444. URL `http://www.jstor.org/stable/2332579`.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, May 1980.

A. Zellner and A. Siow. Posterior odds ratios for selected regression hypotheses. *Trabajos de Estadistica y de Investigacion Operativa*, 31(1):585–603, Feb 1980. ISSN 0041-0241. doi: 10.1007/BF02888369. URL `https://doi.org/10.1007/BF02888369`.

Arnold Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. In P. Goel and Arnold Zellner, editors, *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, pages 233–243. Elsevier Science Publishers, Inc., 1986.

Yichi Zhang, Siyu Tao, Wei Chen, and Daniel W Apley. A latent variable approach to gaussian process modeling with qualitative and quantitative factors. *Technometrics*, (just-accepted):1–19, 2019.

Appendix A

Appendix: lm.slgf Code

```
lm.slgf <- function(df,response,lgf=NA,
                    usermodels,prior="flat",
                    het=rep(0,length(usermodels)),
                    min.levels=1){

  #Ensure variable types are correct
  df[,which(colnames(df)==response)] <- as.numeric(
      as.character(df[,which(colnames(df)==response)]))
  df[,which(colnames(df)==lgf)] <- as.factor(df[,which(colnames(df)==lgf)])

  #Create a vector for the LGF
  lgf <- as.factor(df[,which(colnames(df)==as.character(lgf))])
  #K = number of levels of LGF
  K <- length(levels(lgf))

  #Create a centered vector with the response
  y <- df[,which(colnames(df)==as.character(response))]
```

```
ybar <- mean(y)
# y <- y-ybar

#Sample size
N <- length(y)

if(min.levels>=3){
  d_groups <- groupings(matrix(NA,nrow=K,ncol=1))
  while(length(d_groups[[1]][,1])>min.levels){
    d_groups <- (d_groups[-length(d_groups)])
  }  }
if(min.levels==2){
  d_groups <- groupings(matrix(NA,nrow=K,ncol=1))
  if(K==3){d_groups[[1]][,3] <- c(2,3)}
}
if(min.levels==1){
  d_groups <- as.list(c(NA,groupings(matrix(NA,nrow=K,ncol=1))))
  d_groups[[1]] <- as.matrix(combn(K,1))
}
if(min.levels<1){print("Enter a valid value for min.levels.")}
if(K==2){
  d_groups <- list(matrix(1,nrow=1,ncol=1))
}

#Deterimine the number of possible groupings
if(min.levels>=3){ngroups <- sum(unlist(lapply(d_groups, ncol)))}
if(min.levels==2){ngroups <- (2^(K-1))-K-1}
if(min.levels==1){ngroups <- (2^(K-1))-1}
```

```
#break if there are too many grouping schemes
if(ngroups>1024){
  paste0("There are too many grouping schemes (", ngroups,");
          try increasing min.levels to reduce the number of schemes.")
  break
}

#Create a minimial training sample size of 0 and
#fractional exponent b = m0/N that will be
#updated by more complex models as needed
m0 <- 2
b <- m0/N

#Fix for heteroscedastic models but no group-based FEs
fix <- 0
if(sum(het)>0 & sum(grepl("group", usermodels))==0){
  fix.model <- paste0(response,"~group")
  usermodels <- as.list(c(unlist(usermodels), fix.model))
  het.orig <- het
  het=c(het, 0)
  fix=1
}

#Determine which models do and do not contain a group-based fixed effect

index.lgf <- which(grepl("group",usermodels))
index.no.lgf <- which(!(1:length(usermodels)%in%index.lgf))
```

```
#Determine which models should have a heteroscedastic counterpart
if(sum(het)>0){
  index.het <- seq(1:sum(het)) + length(usermodels)
}else{
  index.het <- NA
}

allmodels <- unlist(usermodels)
allmodels <- c(allmodels,paste0(allmodels[which(het==1)]))

#Create vector indicating how many models are in each class
repvec <- rep(1,length(allmodels))
repvec[c(index.lgf,index.het)] <- ngroups

#Create vector indicating which models will be fit, including
#duplicates for group-based models
modelvec <- rep(allmodels,times=repvec)

#Create indicator for heteroscedastic models
if(sum(het)>0){
  hetvec <- rep(FALSE,length(modelvec))
  hetvec[(length(modelvec)-sum(het==1)*ngroups+1):length(modelvec)] <- TRUE
}else{
  hetvec <- rep(FALSE, length(modelvec))
}

#Create vector for each model's scheme
```

```
schemevec <- rep("None",length(modelvec))

#Create vector to store each model's marginal probability
marginalvec <- rep(0,length(modelvec))

classvec <- rep(1:length(repvec),times=repvec)

#Create result output matrix
result.mat <- as.data.frame(cbind(modelvec,schemevec,classvec,hetvec,
                                  marginalvec))
result.mat$marginalvec <- as.numeric(as.character(result.mat$marginalvec))
result.mat$schemevec <- as.character(result.mat$schemevec)

fitted.models <- as.list(rep(NA,length(unique(usermodels))))
names(fitted.models) <- unlist(unique(usermodels))

#How can we determine whether the model results in an
#expansion or a contraction?
group.dfs <- as.list(rep(NA,ngroups + length(index.no.lgf)))
all.coefs <- as.list(rep(NA, nrow(result.mat)))
all.vars <- as.list(rep(NA, nrow(result.mat)))

#Create list for the models with a group effect

#Fit the models,  compute m0 - noninformative prior case
for(tempmod in unlist(usermodels)){

  m <- which(tempmod==names(fitted.models))
```

```
#Determine the class of the model in question.
#If it's a class that contains only one model,
#there is no LGF so the model is fit accordingly.

if(grepl("group",tempmod)==FALSE){

  templm <- lm(as.formula(tempmod),data=df)
  tempmm <- model.frame(templm)

  group.dfs[[ngroups+which(index.no.lgf==which(tempmod==
             usermodels))]] <- tempmm
  names(group.dfs)[[ngroups+which(index.no.lgf==
             which(tempmod==usermodels))]] <- tempmod

  #Update m0 if model is more complex
  if((length(templm$coefficients)+2)>m0){
    m0 <- length(templm$coefficients)+2
    b <- m0/N
  }

  #For group-based variances only
  if(het[m]==TRUE){

    index <- 1
    templist <- as.list(rep(NA,ngroups))

    for(i in 1:length(d_groups)){
```

```
for(j in 1:ncol(d_groups[[i]])){
  {
    #Create the distinct group effect
    group <- as.numeric(as.numeric(lgf)%in%d_groups[[i]][,j])
    tempscheme1 <- paste(unique(lgf[group==1]),collapse=",")
    tempscheme0 <- paste(unique(lgf[group==0]),collapse=",")
    group[group==1] <- tempscheme1
    group[group==0] <- tempscheme0

    #Update m0 if necessary
    if((length(templm$coefficients)+4)>m0){
      m0=length(templm$coefficients)+4
      b=m0/N
    }

    #Name the model
    tempscheme <- paste(c(tempscheme1,tempscheme0),collapse=":")

    result.mat$schemevec[which(modelvec==tempmod&hetvec==
                         TRUE)[index]] <-
      paste0(tempscheme1,":",tempscheme0)

    names(all.coefs)[[which(modelvec==tempmod&result.mat$
                      schemevec==tempscheme)]] <- tempscheme
    names(all.vars)[[which(modelvec==tempmod&result.mat$
                      schemevec==tempscheme)]] <- tempscheme

    tempscheme <- paste0("scheme=",tempscheme)
```

```
          names(templist)[[index]] <- paste(tempscheme)

          index=index+1
        }  }  }  }
  fitted.models[[m]] <- templm
}

#Group-based fixed effect models
if(grepl("group",tempmod)==TRUE){

  index <- 1
  templist <- as.list(rep(NA,ngroups))

  for(i in 1:length(d_groups)){
    for(j in 1:ncol(d_groups[[i]])){
      {
        #Create the distinct group effect
        group <- as.numeric(as.numeric(lgf)%in%d_groups[[i]][,j])
        tempscheme1 <- paste(unique(lgf[group==1]),collapse=",")
        tempscheme0 <- paste(unique(lgf[group==0]),collapse=",")
        group[group==1] <- tempscheme1
        group[group==0] <- tempscheme0

        tempdf <- data.frame(df,"group"=as.factor(group))

        templm <- lm(as.formula(tempmod),data=tempdf)

        templist[[index]] <- templm
```

```
tempdf[,which(colnames(tempdf)==response)] <- tempdf[,
        which(colnames(tempdf)==response)] # + ybar
group.dfs[[index]] <- tempdf
names(group.dfs)[index] <- paste(tempscheme1, tempscheme0, sep=":")

#Update m0 if necessary
if((length(templm$coefficients)+2)>m0){
  m0=length(templm$coefficients)+2 #<-need to delete coefs with NA
  b=m0/N
}

#Name the model
tempscheme <- paste(c(tempscheme1,tempscheme0),collapse=":")

result.mat$schemevec[which(modelvec==tempmod&hetvec==FALSE)[index]]
 <- tempscheme

result.mat$schemevec[which(modelvec==tempmod&hetvec==
                     TRUE)[index]] <- tempscheme

nameindex <- which(modelvec==tempmod&result.mat$schemevec==
                   tempscheme)

names(all.coefs)[nameindex] <- rep(tempscheme, length(nameindex))
names(all.vars)[nameindex] <- rep(tempscheme, length(nameindex))

# names(all.coefs)[[which(modelvec==tempmod&hetvec==
        TRUE&result.mat$schemevec==tempscheme)]] <- tempscheme
```

```
          # names(all.coefs)[[which(modelvec==tempmod&hetvec==
                  FALSE&result.mat$schemevec==tempscheme)]] <- tempscheme
          #
          # names(all.vars)[[which(modelvec==tempmod&hetvec==TRUE&
            result.mat$schemevec==tempscheme)]] <- tempscheme
          # names(all.vars)[[which(modelvec==tempmod&hetvec==FALSE&
            result.mat$schemevec==tempscheme)]] <- tempscheme

          tempscheme <- paste0("scheme=",tempscheme)
          names(templist)[[index]] <- paste(tempscheme)

          index <- index+1
        }   }   }
    fitted.models[[m]] <- templist
  }  }

#Have smaller m0 if using ZS prior
if(prior=="zs"){
  m0 <- 2 + 1 + 1 #two variances + g + intercept
  b <- m0/N
  all.gs <- as.list(rep(NA, nrow(result.mat)))
  names(all.gs) <- paste0(result.mat$modelvec,", ", result.mat$schemevec)
}

names(all.coefs) <- paste0(result.mat$modelvec,", ", result.mat$schemevec)
names(all.vars) <- paste0(result.mat$modelvec,", ", result.mat$schemevec)

#Compute marginal model probabilities - noninformative prior case
```

```
#Store fixed effect and variance MLEs

if(prior=="flat"){
  for(tempmod in unlist(usermodels)){

    m <- which(tempmod==names(fitted.models))
    tempfit <- fitted.models[[m]]

    #Compute marginal model probability for homoscedastic
    #models with global fixed effects
    if(grepl("group",tempmod)==FALSE){

      tempSSResid <- sum(tempfit$residuals^2)
      tempP <- length(tempfit$coefficients)
      templogPY <- (-N*(1-b)/2)*log(pi) + log(b^((N*b-1)/2)) +
        (lgamma((N-tempP)/2)) - (lgamma((N*b-tempP)/2)) +
        ((-N*(1-b)/2)*log(tempSSResid))

      result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                                   result.mat$hetvec==FALSE)] <- templogPY
      all.coefs[[which(result.mat$modelvec==tempmod&result.mat$
                 schemevec=="None")]] <- tempfit$coefficients
      all.vars[[which(result.mat$modelvec==tempmod&result.mat$
                 schemevec=="None")]] <- summary(tempfit)$sigma^2
    }

    #Heteroscedastic models with global fixed effects
    if(grepl("group",tempmod)==FALSE & het[m]==TRUE){
```

```
index <- 1

for(i in 1:length(d_groups)){
  for(j in 1:ncol(d_groups[[i]])){
    {
      #Create the distinct group effect
      MM <- model.matrix(tempfit)
      tempdf <- group.dfs[[index]]

      group <- as.numeric(as.numeric(lgf)%in%d_groups[[i]][,j])
      tempscheme1 <- paste(unique(lgf[group==1]),collapse=",")
      tempscheme0 <- paste(unique(lgf[group==0]),collapse=",")
      tempscheme <- paste(c(tempscheme1,tempscheme0),collapse=":")

      level1 <- tempscheme0
      y1 <- tempdf[tempdf$group==level1,which(names(tempdf)==response)]
      n1 <- sum(tempdf$group==level1)
      level2 <- tempscheme1
      y2 <- tempdf[tempdf$group==level2,which(names(tempdf)==response)]
      n2 <- sum(tempdf$group==level2)

      tempP <- ncol(MM)

      f3 =function(arg){
        why=c(y1,y2)
        bee=1
        gam1=arg[1]
```

```
  gam2=arg[2]
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e)
}
f3b=function(arg){
  why=c(y1,y2)
  bee=b
  gam1=arg[1]
  gam2=arg[2]
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e)
}

lvf3 =function(arg){
  phi1=arg[1]
  phi2=arg[2]
```

```
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
  bee=1
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e*J)
}
lvf3b=function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
  bee=b
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
```

```
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e*J)
}

lvlogf3 =function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
  bee=1
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%solve(t(MM)%*%
     Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(log(d)+log(s)+(-.5*as.numeric(e1-e2))+log(J))
}
lvlogf3b=function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
```

```
  bee=b
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(log(d)+log(s)+(-.5*as.numeric(e1-e2))+log(J))
}

gam1hat <- sum(fitted.models[[m]]$residuals^2)/(N-tempP)
gam2hat <- gam1hat
lvhat <- log(c(gam1hat, gam2hat))

mode3=optim(lvhat,lvlogf3,
            control=list(fnscale=-1),
            method="Nelder-Mead")$par

mode3b=optim(lvhat,lvlogf3b,
             control=list(fnscale=-1),
             method="Nelder-Mead")$par

H3=det(-1*optim(mode3,lvlogf3,
                control=list(fnscale=-1),hessian=TRUE,
                method="Nelder-Mead")$hessian)^-.5
H3b=det(-1*optim(mode3b,lvlogf3b,
                 control=list(fnscale=-1),hessian=TRUE,
```

```
                        method="Nelder-Mead")$hessian)^-.5

        templogPY <- log(((2*pi)^(-(N-tempP)/2)))-
          log(((2*pi)^(-(N*b-tempP)/2)))+
          # log(H3*(lvf3(mode3))/(H3b*lvf3b(mode3b)))
          log(H3) + lvlogf3(mode3) - log(H3b) - lvlogf3b(mode3b)

        result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                                     result.mat$hetvec==
                              TRUE)[index]] <- templogPY

        all.coefs[[which(result.mat$modelvec==tempmod&
                          result.mat$schemevec==tempscheme&
                          result.mat$hetvec==TRUE)]] <-
                          tempfit$coefficients
        all.vars[[which(result.mat$modelvec==tempmod&
                         result.mat$schemevec==tempscheme&
                         result.mat$hetvec==TRUE)]] <-
                         exp(c(mode3[2], mode3[1]))
        index=index+1
      }   }   }
}

#Homoscedastic models, group-based fixed effects
if(grepl("group",tempmod)==TRUE){

  index <- 1
```

```
result.mat.rows <- which(result.mat$modelvec==tempmod &
                    result.mat$hetvec==FALSE)

for(i in 1:length(d_groups)){
  for(j in 1:ncol(d_groups[[i]])){
    {
      group <- as.numeric(as.numeric(lgf)%in%d_groups[[i]][,j])
      tempscheme1 <- paste(unique(lgf[group==1]),collapse=",")
      tempscheme0 <- paste(unique(lgf[group==0]),collapse=",")
      tempscheme <- paste(c(tempscheme1,tempscheme0),collapse=":")

      tempfit.scheme <- tempfit[[index]]
      tempSSResid <- sum(tempfit.scheme$residuals^2)
      tempP <- length(tempfit.scheme$coefficients)
      templogPY <- (-N*(1-b)/2)*log(pi)+((N*b-1)/2)*log(b)+
        (lgamma((N-tempP)/2))-(lgamma((N*b-tempP)/2))+
        ((-N*(1-b)/2)*log(tempSSResid))

      result.mat$marginalvec[result.mat.rows[index]] <- templogPY

      index <- index+1

      all.coefs[[which(result.mat$modelvec==tempmod&
                        result.mat$schemevec==tempscheme&
                        result.mat$hetvec==FALSE)]] <-
                        tempfit.scheme$coefficients
      all.vars[[which(result.mat$modelvec==tempmod&
                       result.mat$schemevec==tempscheme&
```

```
                          result.mat$hetvec==FALSE)]] <-
                          summary(tempfit.scheme)$sigma^2
      }   }   }
}

#Heteroscedastic models, group-based fixed effects
if(grepl("group",tempmod)==TRUE & het[m]==TRUE){

  index <- 1

  result.mat.rows <- which(result.mat$modelvec==tempmod &
                           result.mat$hetvec==TRUE)

  for(i in 1:length(d_groups)){
    for(j in 1:ncol(d_groups[[i]])){
      {
        #Create the distinct group effect
        MM <- model.matrix(tempfit[[index]])
        tempdf <- group.dfs[[index]]

        group <- as.numeric(as.numeric(lgf)%in%d_groups[[i]][,j])
        tempscheme1 <- paste(unique(lgf[group==1]),collapse=",")
        tempscheme0 <- paste(unique(lgf[group==0]),collapse=",")
        tempscheme <- paste(c(tempscheme1,tempscheme0),collapse=":")

        level1 <- tempscheme0
        y1 <- tempdf[tempdf$group==level1,which(names(tempdf)==response)]
        n1 <- sum(tempdf$group==level1)
```

```
level2 <- tempscheme1
y2 <- tempdf[tempdf$group==level2,which(names(tempdf)==response)]
n2 <- sum(tempdf$group==level2)

tempP <- ncol(MM)

f3 =function(arg){
  why=c(y1,y2)
  bee=1
  gam1=arg[1]
  gam2=arg[2]
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e)
}
f3b=function(arg){
  why=c(y1,y2)
  bee=b
  gam1=arg[1]
  gam2=arg[2]
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
```

```
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e)
}

lvf3 =function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
  bee=1
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e*J)
}
lvf3b=function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
```

```
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  why=c(y1,y2)
  bee=b
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(d*s*e*J)
}

lvlogf3 =function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  # J=exp(-(phi1+phi2))
  logJ=-(phi1+phi2)
  why=c(y1,y2)
  bee=1
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  # d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  logd=-.5*log(det(bee*t(MM)%*%Sigma%*%MM))
  # s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  logs=((bee*n1/2)-1)*log(gam1) + ((bee*n2/2)-1)*log(gam2)
```

```
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(logd+logs+(-.5*as.numeric(e1-e2))+logJ)
}
lvlogf3b=function(arg){
  phi1=arg[1]
  phi2=arg[2]
  gam1=exp(-phi1)
  gam2=exp(-phi2)
  J=exp(-(phi1+phi2))
  logJ=-(phi1+phi2)
  why=c(y1,y2)
  bee=b
  Sigma=diag(c(rep(gam1,n1),rep(gam2,n2)))
  # d=det(bee*t(MM)%*%Sigma%*%MM)^-.5
  logd=-.5*log(det(bee*t(MM)%*%Sigma%*%MM))
  # s=(gam1^((bee*n1/2)-1))*(gam2^((bee*n2/2)-1))
  logs=((bee*n1/2)-1)*log(gam1) + ((bee*n2/2)-1)*log(gam2)
  e1=bee*t(why)%*%Sigma%*%why
  e2=bee*t(why)%*%Sigma%*%MM%*%
     solve(t(MM)%*%Sigma%*%MM)%*%t(MM)%*%Sigma%*%why
  e=exp(-.5*as.numeric(e1-e2))
  return(logd+logs+(-.5*as.numeric(e1-e2))+logJ)
}

lv1hat <- log(sum(tempfit[[index]]$residuals
```

```
              [tempdf$group==level1]^2))
lv2hat <- log(sum(tempfit[[index]]$residuals
              [tempdf$group==level2]^2))

mode3=optim(c(lv1hat,lv2hat),lvlogf3,
            control=list(fnscale=-1),
            method="Nelder-Mead")$par

mode3b=optim(c(lv1hat,lv2hat),lvlogf3b,
             control=list(fnscale=-1),
             method="L-BFGS-B")$par

H3=det(-1*optim(mode3,lvlogf3,
                control=list(fnscale=-1),hessian=TRUE,
                method="Nelder-Mead")$hessian)^-.5
H3b=det(-1*optim(mode3b,lvlogf3b,
                 control=list(fnscale=-1),hessian=TRUE,
                 method="Nelder-Mead")$hessian)^-.5

templogPY <- log(((2*pi)^(-(N-tempP)/2)))-
  log(((2*pi)^(-(N*b-tempP)/2)))+
  # log(H3*(lvf3(mode3))/(H3b*lvf3b(mode3b)))
  log(H3) + lvlogf3(mode3) - log(H3b) - lvlogf3b(mode3b)

result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                             result.mat$hetvec==TRUE)
                       [index]] <- templogPY
```

```
            tempfit.scheme <- tempfit[[index]]

            all.coefs[[which(result.mat$modelvec==tempmod&
                              result.mat$schemevec==tempscheme&
                              result.mat$hetvec==TRUE)]] <-
                              tempfit.scheme$coefficients
            all.vars[[which(result.mat$modelvec==tempmod&
                             result.mat$schemevec==tempscheme&
                             result.mat$hetvec==TRUE)]] <-
                             exp(c(mode3[2], mode3[1]))
            index=index+1
          }   }   }   }   }
}

#total sum of squares
SST <- sum((y-mean(y))^2)
Z <- rep(1,N)

#Compute marginal model probabilities - Zellner-Siow mixture g-prior case

if(prior=="zs"){
  for(tempmod in unlist(usermodels)){

    m <- which(tempmod==names(fitted.models))
    tempfit <- fitted.models[[m]]

    #done 3/23
```

```
#Homoscedsatic models, group-based effects
if(grepl("group",tempmod)==FALSE){

  # tempSSResid <- sum(tempfit$residuals^2)
  tempP <- length(tempfit$coefficients)
  R2A <- summary(tempfit)$r.squared

  #Mode of additive marginal distribution
  faddmode <- function(gee){
    Q <- 1-R2A
    value <- -Q*(tempP+3)*(gee^3)+
      (N-tempP-4-2*(1-R2A))*(gee^2)+
      ((N*(2-R2A)-3)*gee)+N
    return(value)
  }
  #Mode of fractional additive marginal distribution
  faddbmode=function(gee){
    Q <- 1-R2A
    value=-Q*(b^2)*(tempP+3)*(gee^3)+
      (b*(N*b-tempP-4)-2*Q)*(gee^2)+
      (N*b*(2-R2A)-3)*gee+N
    return(value)
  }

  addmode <- uniroot(faddmode,c(0,1e9),check.conv=TRUE,
                     tol=1e-10)$root
  addbmode <- uniroot(faddbmode,c(1e-9,1e9),check.conv=TRUE,
                      tol=1e-10)$root
```

```
logfadd <- function(gee){
  value <- ((N-tempP-1)/2)*log(1+gee) +
    (-(N-1)/2)*log(1+(1-R2A)*gee) +
    -1.5*log(gee) +
    (-N/(2*gee))
  return(value)
}

logfaddb <- function(gee){  #Put on log-scale
  value <- ((N*b-1-tempP)/2)*log(1+b*gee) +
    (-(N*b-1)/2)*log(1+b*gee*(1-R2A)) +
    -1.5*log(gee) + (-N/(2*gee))
  return(value)
}

faddH <- function(gee){
  value <- .5*((((N-1)*(1-R2A)^2)/((1+gee*(1-R2A))^2))-
                ((N-tempP-1)/((1+gee)^2))+
                (3/(gee^2))-
                ((2*N)/(gee^3)))
  return(value)
}
faddHb <- function(gee){
  value <- .5*((((N*b-1)*(b^2)*(1-R2A)^2)/((1+gee*b*(1-R2A))^2))-
                ((N*b-tempP-1)*(b^2)/((1+b*gee)^2))+
                (3/(gee^2))-
                ((2*N)/(gee^3)))
```

```
    return(value)
  }

  addH <- (-1*faddH(addmode))^-.5
  addHb <- (-1*faddHb(addbmode))^-.5

  qaddLA <- lgamma((N-1)/2) + .5*log(N/2) + (-(N-1)/2)*log(SST) +
    .5*log(N) - ((N-1)/2)*log(pi) - lgamma(.5) + log(sqrt(2*pi)) +
    log(addH) + logfadd(addmode)

  qaddbLA <- lgamma((N*b-1)/2) + .5*log(N/2) + (-(N*b-1)/2)*log(SST) +
    .5*log(N) - ((N*b-1)/2)*log(pi) - lgamma(.5) + log(sqrt(2*pi)) +
    log(addHb) + logfaddb(addbmode) + (-N*b/2)*log(b)

  templogPY <- qaddLA - qaddbLA

  result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                                result.mat$hetvec==FALSE)] <- templogPY
  all.coefs[[which(result.mat$modelvec==tempmod&result.mat$
            schemevec=="None")]] <- tempfit$coefficients
  all.vars[[which(result.mat$modelvec==tempmod&result.mat$
           schemevec=="None")]] <- summary(tempfit)$sigma^2
  all.gs[[which(result.mat$modelvec==tempmod&result.mat$
         schemevec=="None")]] <- addmode
}

#Heteroscedastic models with global fixed effects
if(grepl("group",tempmod)==FALSE & het[m]==TRUE){
```

```
index <- 1

#done 3/24
for(i in 1:length(d_groups)){
  for(j in 1:ncol(d_groups[[i]])){
    {
      #Create the distinct group effect
      MM <- model.matrix(tempfit)
      tempdf <- group.dfs[[index]]
      tempP <- ncol(MM)
      R2A <- summary(tempfit)$r.squared
      gamhat <- 1/(((1-R2A)*SST)/(N-tempP))
      lvhat <- log(1/gamhat)
      level1 <- levels(tempdf$group)[1]
      y1 <- tempdf[tempdf$group==level1,which(names(tempdf)==response)]
      n1 <- sum(tempdf$group==level1)
      level2 <- levels(tempdf$group)[2]
      y2 <- tempdf[tempdf$group==level2,which(names(tempdf)==response)]
      n2 <- sum(tempdf$group==level2)
      scheme <- as.numeric(tempdf$group==level1)
      tempscheme1 <- levels(tempdf$group)[1]
      tempscheme0 <- levels(tempdf$group)[2]

      lvlogf3  <- function(arg){
        gee <- arg[1]
        u <- arg[2]
        v <- arg[3]
```

```
  gam1 <- exp(-u)
  gam2 <- exp(-v)
  J <- exp(-(u+v))
  logJ <- -(u+v)
  gamvec <- rep(NA,N)
  gamvec[scheme==1] <- gam1
  gamvec[scheme==0] <- gam2
  S <- diag(gamvec)
  Z_S <- S%*%Z%*%solve(t(Z)%*%S%*%Z)%*%t(Z)%*%S

  value <- (n1/2-1)*(-u)+(n2/2-1)*(-v)-((tempP/2)*log(gee))+
    .5*determinant(t(MM)%*%S%*%MM, logarithm=TRUE)$modulus -
    .5*determinant(t(Z)%*%S%*%Z, logarithm=TRUE)$modulus -
    .5*determinant(((gee+1)/gee)*t(MM)%*%S%*%MM-
                   t(MM)%*%Z_S%*%MM,
                   logarithm=TRUE)$modulus -
    .5*(t(y)%*%S%*%y-t(y)%*%Z_S%*%y-
          t(y)%*%(S-Z_S)%*%MM%*%solve(
            ((gee+1)/gee)*t(MM)%*%S%*%MM-
              t(MM)%*%Z_S%*%MM)%*%t(MM)%*%(S-Z_S)%*%y)-
    1.5*log(gee)-(N/(2*gee))+logJ
  return(value)
}
lvlogf3b <- function(arg){
  gee <- arg[1]
  u <- arg[2]
  v <- arg[3]
  gam1 <- exp(-u)
```

```
  gam2 <- exp(-v)
  J <- exp(-(u+v))
  logJ <- -(u+v)
  gamvec <- rep(NA,N)
  gamvec[scheme==1] <- gam1
  gamvec[scheme==0] <- gam2
  S <- diag(gamvec)
  Z_S <- S%*%Z%*%solve(t(Z)%*%S%*%Z)%*%t(Z)%*%S
  bg <- b*gee

  value <- (n1*b/2-1)*(-u)+(n2*b/2-1)*(-v)-
           ((tempP/2)*log(gee))+
    .5*determinant(t(MM)%*%S%*%MM, logarithm=TRUE)$modulus -
    .5*determinant(t(Z)%*%S%*%Z, logarithm=TRUE)$modulus -
    .5*determinant(((bg+1)/bg)*t(MM)%*%S%*%MM-
                   t(MM)%*%Z_S%*%MM,logarithm=TRUE)$modulus -
    .5*b*(t(y)%*%S%*%y-t(y)%*%Z_S%*%y-
            t(y)%*%(S-Z_S)%*%MM%*%solve(
              ((bg+1)/bg)*t(MM)%*%S%*%MM-
                t(MM)%*%Z_S%*%MM)%*%t(MM)%*%(S-Z_S)%*%y)-
    1.5*log(gee)-(N/(2*gee))+logJ
  return(value)
}

lvmode3 <- optim(c(N,lvhat,lvhat),
                 lvlogf3,
                 control=list(fnscale=-1),
                 method="L-BFGS-B",
```

```
                 lower=c(1e-9,-1e6,-1e6),
                 upper=c(1e9,1e6,1e6))$par
lvmode3b <- optim(c(N,lvhat,lvhat),
                  lvlogf3b,
                  control=list(fnscale=-1),
                  method="L-BFGS-B",
                  lower=c(1e-6,-1e6,-1e6),
                  upper=c(1e9,1e6,1e6))$par

H3 <- (-1*det(hessian(lvlogf3,lvmode3)))^-.5
logH3 <- -.5*determinant(-1*hessian(lvlogf3,lvmode3),
             logarithm=TRUE)$modulus
H3b <- (-1*det(hessian(lvlogf3b,lvmode3b)))^-.5
logH3b <- -.5*determinant(-1*hessian(lvlogf3b,lvmode3b),
              logarithm=TRUE)$modulus

q3LA <- (-(N-1)/2)*log(2*pi) + .5*log(N/2) - lgamma(.5) +
  1.5*log(2*pi) + logH3 + lvlogf3(lvmode3)
q3bLA <- (-(N*b-1)/2)*log(2*pi) + (-(tempP+1)/2)*log(b) +
  .5*log(N/2) - lgamma(.5) + 1.5*log(2*pi) +
  logH3b + lvlogf3b(lvmode3b)

templogPY <- q3LA - q3bLA

result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                               result.mat$hetvec==TRUE)[index]]
                        <- templogPY
```

```
        tempscheme <- names(group.dfs[index])

        all.coefs[[which(result.mat$modelvec==tempmod&
                          result.mat$schemevec==tempscheme&
                          result.mat$hetvec==TRUE)]] <-
                          tempfit$coefficients
        all.vars[[which(result.mat$modelvec==tempmod&
                         result.mat$schemevec==tempscheme&
                         result.mat$hetvec==TRUE)]] <-
                         exp(c(lvmode3[2], lvmode3[1]))
        all.gs[[which(result.mat$modelvec==tempmod&
                       result.mat$schemevec==tempscheme&
                       result.mat$hetvec==TRUE)]] <- lvmode3[1]
        index=index+1
      }   }   }
}

#Group-based fixed effects
if(grepl("group",tempmod)==TRUE){

  index <- 1

  result.mat.rows=which(result.mat$modelvec==tempmod &
                        result.mat$hetvec==FALSE)

  #homoscedastic group-based effects
  for(i in 1:length(d_groups)){
    for(j in 1:ncol(d_groups[[i]])){
```

```
{
  {
    tempscheme <- gsub("scheme=","",names(tempfit)[index])

    tempfit.scheme <- tempfit[[index]]
    tempSSResid <- sum(tempfit.scheme$residuals^2)
    tempP <- length(tempfit.scheme$coefficients)

    R2A=1-(tempSSResid/SST)
    Q <- 1-R2A

    #Mode of additive marginal distribution
    faddmode <- function(gee){
      Q <- 1-R2A
      value <- -Q*(tempP+3)*(gee^3)+
        (N-tempP-4-2*(1-R2A))*(gee^2)+
        ((N*(2-R2A)-3)*gee)+N
      return(value)
    }
    #Mode of fractional additive marginal distribution
    faddbmode=function(gee){
      Q <- 1-R2A
      value=-Q*(b^2)*(tempP+3)*(gee^3)+
        (b*(N*b-tempP-4)-2*Q)*(gee^2)+
        (N*b*(2-R2A)-3)*gee+N
      return(value)
    }
```

```
addmode <- uniroot(faddmode,c(0,1e9),check.conv=TRUE,
                   tol=1e-10)$root
addbmode <- uniroot(faddbmode,c(1e-9,1e9),check.conv=TRUE,
                    tol=1e-10)$root

logfadd <- function(gee){
  value <- ((N-tempP-1)/2)*log(1+gee) +
    (-(N-1)/2)*log(1+(1-R2A)*gee) +
    -1.5*log(gee) +
    (-N/(2*gee))
  return(value)
}

logfaddb <- function(gee){  #Put on log-scale
  value <- ((N*b-1-tempP)/2)*log(1+b*gee) +
    (-(N*b-1)/2)*log(1+b*gee*(1-R2A)) +
    -1.5*log(gee) + (-N/(2*gee))
  return(value)
}

faddH <- function(gee){
  value <- .5*((((N-1)*(1-R2A)^2)/((1+gee*(1-R2A))^2))-
                 ((N-tempP-1)/((1+gee)^2))+
                 (3/(gee^2))-
                 ((2*N)/(gee^3)))
  return(value)
}
faddHb <- function(gee){
```

```
  value <- .5*((((N*b-1)*(b^2)*(1-R2A)^2)/((1+gee*b*(1-R2A))^2))-
                ((N*b-tempP-1)*(b^2)/((1+b*gee)^2))+
                (3/(gee^2))-
                ((2*N)/(gee^3)))
  return(value)
}

addH <- (-1*faddH(addmode))^-.5
addHb <- (-1*faddHb(addbmode))^-.5

qaddLA <- lgamma((N-1)/2) + .5*log(N/2) +
          (-(N-1)/2)*log(SST) +
  .5*log(N) - ((N-1)/2)*log(pi) - lgamma(.5) +
          log(sqrt(2*pi)) +
  log(addH) + logfadd(addmode)

qaddbLA <- lgamma((N*b-1)/2) + .5*log(N/2) +
           (-(N*b-1)/2)*log(SST) +
  .5*log(N) - ((N*b-1)/2)*log(pi) - lgamma(.5) +
           log(sqrt(2*pi)) +
  log(addHb) + logfaddb(addbmode) + (-N*b/2)*log(b)

templogPY <- qaddLA - qaddbLA

result.mat$marginalvec[result.mat.rows[index]] <- templogPY

all.coefs[[which(result.mat$modelvec==tempmod&
                    result.mat$schemevec==tempscheme&
```

```
                          result.mat$hetvec==FALSE)]] <-
                          tempfit.scheme$coefficients
        all.vars[[which(result.mat$modelvec==tempmod&
                         result.mat$schemevec==tempscheme&
                         result.mat$hetvec==FALSE)]] <-
                         summary(tempfit.scheme)$sigma^2
        all.gs[[which(result.mat$modelvec==tempmod&
                       result.mat$schemevec==tempscheme&
                       result.mat$hetvec==FALSE)]] <- addmode

        index <- index+1
      }   }   }
}

#Heteroscedastic models, group-based fixed effects
if(grepl("group",tempmod)==TRUE & het[m]==TRUE){

  index <- 1

  result.mat.rows=which(result.mat$modelvec==tempmod &
                       result.mat$hetvec==TRUE)

  for(i in 1:length(d_groups)){
    for(j in 1:ncol(d_groups[[i]])){
      {
        tempscheme <- gsub("scheme=","",names(tempfit)[index])

        tempfit.scheme <- tempfit[[index]]
```

```
tempSSResid <- sum(tempfit.scheme$residuals^2)
tempP <- length(tempfit.scheme$coefficients)

MM <- model.matrix(tempfit.scheme)
tempdf <- group.dfs[[index]]
R2A <- summary(tempfit.scheme)$r.squared
gamhat <- 1/(((1-R2A)*SST)/(N-tempP))
lvhat <- log(1/gamhat)
level1 <- levels(tempdf$group)[1]
y1 <- tempdf[tempdf$group==level1,
             which(names(tempdf)==response)]
n1 <- sum(tempdf$group==level1)
level2 <- levels(tempdf$group)[2]
y2 <- tempdf[tempdf$group==level2,
             which(names(tempdf)==response)]
n2 <- sum(tempdf$group==level2)
scheme <- as.numeric(tempdf$group==level1)

lvlogf3  <- function(arg){
  gee <- arg[1]
  u <- arg[2]
  v <- arg[3]
  gam1 <- exp(-u)
  gam2 <- exp(-v)
  J <- exp(-(u+v))
  logJ <- -(u+v)
  gamvec <- rep(NA,N)
  gamvec[scheme==1] <- gam1
```

```
  gamvec[scheme==0] <- gam2
  S <- diag(gamvec)
  Z_S <- S%*%Z%*%solve(t(Z)%*%S%*%Z)%*%t(Z)%*%S

  value <- (n1/2-1)*(-u)+(n2/2-1)*(-v)-((tempP/2)*log(gee))+
    .5*determinant(t(MM)%*%S%*%MM, logarithm=TRUE)$modulus -
    .5*determinant(t(Z)%*%S%*%Z, logarithm=TRUE)$modulus -
    .5*determinant(((gee+1)/gee)*t(MM)%*%S%*%MM-
                   t(MM)%*%Z_S%*%MM,
                   logarithm=TRUE)$modulus -
    .5*(t(y)%*%S%*%y-t(y)%*%Z_S%*%y-
          t(y)%*%(S-Z_S)%*%MM%*%solve(
            ((gee+1)/gee)*t(MM)%*%S%*%MM-
              t(MM)%*%Z_S%*%MM)%*%t(MM)%*%(S-Z_S)%*%y)-
    1.5*log(gee)-(N/(2*gee))+logJ
  return(value)
}
lvlogf3b<- function(arg){
  gee <- arg[1]
  u <- arg[2]
  v <- arg[3]
  gam1 <- exp(-u)
  gam2 <- exp(-v)
  J <- exp(-(u+v))
  logJ <- -(u+v)
  gamvec <- rep(NA,N)
  gamvec[scheme==1] <- gam1
  gamvec[scheme==0] <- gam2
```

```
  S <- diag(gamvec)
  Z_S <- S%*%Z%*%solve(t(Z)%*%S%*%Z)%*%t(Z)%*%S
  bg <- b*gee

  value <- (n1*b/2-1)*(-u)+(n2*b/2-1)*(-v)-
     ((tempP/2)*log(gee))+
    .5*determinant(t(MM)%*%S%*%MM, logarithm=TRUE)$modulus -
    .5*determinant(t(Z)%*%S%*%Z, logarithm=TRUE)$modulus -
    .5*determinant(((bg+1)/bg)*t(MM)%*%S%*%MM-t(MM)%*%Z_S%*%MM,
                   logarithm=TRUE)$modulus -
    .5*b*(t(y)%*%S%*%y-t(y)%*%Z_S%*%y-
            t(y)%*%(S-Z_S)%*%MM%*%solve(
              ((bg+1)/bg)*t(MM)%*%S%*%MM-
                t(MM)%*%Z_S%*%MM)%*%t(MM)%*%(S-Z_S)%*%y)-
    1.5*log(gee)-(N/(2*gee))+logJ
  return(value)
}

lvmode3 <- optim(c(N,lvhat,lvhat),
                 lvlogf3,
                 control=list(fnscale=-1),
                 method="L-BFGS-B",
                 lower=c(1e-9,-1e6,-1e6),
                 upper=c(1e9,abs(lvhat),abs(lvhat)))$par
lvmode3b <- optim(c(N,lvhat,lvhat),
                  lvlogf3b,
                  control=list(fnscale=-1),
                  method="L-BFGS-B",
```

```
                  lower=c(1e-3,-1e2,-1e2),
                  upper=c(1e6,abs(lvhat),abs(lvhat)))$par

H3 <- (-1*det(hessian(lvlogf3,lvmode3)))^-.5
logH3 <- -.5*determinant(-1*hessian(lvlogf3,lvmode3),
                         logarithm=TRUE)$modulus
H3b <- (-1*det(hessian(lvlogf3b,lvmode3b)))^-.5
logH3b <- -.5*determinant(-1*hessian(lvlogf3b,lvmode3b),
                          logarithm=TRUE)$modulus

q3LA <- (-(N-1)/2)*log(2*pi) + .5*log(N/2) - lgamma(.5) +
  1.5*log(2*pi) + logH3 + lvlogf3(lvmode3)
q3bLA <- (-(N*b-1)/2)*log(2*pi) + (-(tempP+1)/2)*log(b) +
  .5*log(N/2) - lgamma(.5) + 1.5*log(2*pi) +
  logH3b + lvlogf3b(lvmode3b)

templogPY <- q3LA - q3bLA

result.mat$marginalvec[which(result.mat$modelvec==tempmod&
                               result.mat$hetvec==TRUE)[index]] <- templ
tempscheme <- names(group.dfs[index])

all.coefs[[which(result.mat$modelvec==tempmod&
                   result.mat$schemevec==tempscheme&
                   result.mat$hetvec==TRUE)]] <-
                   tempfit.scheme$coefficients
all.vars[[which(result.mat$modelvec==tempmod&
                  result.mat$schemevec==tempscheme&
```

```
                                result.mat$hetvec==TRUE)]] <-
                                exp(c(lvmode3[2], lvmode3[1]))
              all.gs[[which(result.mat$modelvec==tempmod&
                              result.mat$schemevec==tempscheme&
                              result.mat$hetvec==TRUE)]] <- lvmode3[1]
              index=index+1
            }   }   }   }   }  }
}


#remove fix model if necessary
if(fix==1){
  delete <- which(as.character(result.mat$modelvec)==
                 paste0(response,"~group"))
  result.mat <- result.mat[-delete,]
  row.names(result.mat) <- 1:nrow(result.mat)
  result.mat$classvec <- droplevels(result.mat$classvec)
  result.mat$classvec <- as.factor(rep(1:length(
  levels(result.mat$classvec)), times=table(result.mat$classvec)))
}

modevs <- exp(result.mat$marginalvec-(max(result.mat$marginalvec)))

#Create uniform prior by class
modpriors <- rep(NA,length(modevs))
modprobs <- rep(NA,length(modevs))

modpriors[result.mat$scheme=="None"] <-
```

```
1/length(levels(unique(result.mat$classvec)))
modpriors[result.mat$scheme!="None"] <-
1/(ngroups*length(levels(unique(result.mat$classvec))))

for(i in 1:length(modevs)){
  modprobs[i]=(modevs[i]*modpriors[i])/(modevs%*%modpriors)
}

result.mat$modpriors <- modpriors
result.mat$modprobs <- round(modprobs,8)

#Aggregate probabilities by grouping
#scheme to obtain class probabilities
classprobs <- rep(NA,length(unique(result.mat$classvec)))
names(classprobs) <- levels(result.mat$classvec)
for(c in as.numeric(as.character(unique(result.mat$classvec)))){
  classprobs[c] <- sum(modprobs[as.numeric(
  as.character(result.mat$classvec))==c])
}

result.mat$mle.index <- seq(1:nrow(result.mat))
sorted <- result.mat[order(-modprobs),]
mle.index <- sorted$mle.index
sorted$cumuprob <- cumsum(sorted$modprobs)

sorted <- sorted[,-c(3,8)]

variance <- rep(NA, length(result.mat$hetvec))
```

```
variance[sorted$hetvec==TRUE] <- c("Heterosk")
variance[sorted$hetvec==FALSE] <- c("Homosk")

sorted$hetvec <- variance

sorted$df.Index <- rep(NA, nrow(result.mat)) #nrow(result.mat)
for(i in 1:nrow(result.mat)){
  if(sorted$schemevec[i]!="None"){
    sorted$df.Index[i] <- which(names(group.dfs)==sorted$schemevec[i])
  }
  if(sorted$schemevec[i]=="None"){
    sorted$df.Index[i] <- which(names(group.dfs)==sorted$modelvec[i])
  }
}

sorted$mle.index <- mle.index
sorted$model.index <- 1:nrow(sorted)
sorted$class <- paste0(sorted$modelvec, ", ", sorted$hetvec)
names(sorted) <- c("Model", "Scheme", "Variance", "logFlik",
                   "Mod.Prior", "Fmodprob", "Cumulative",
                    "df.Index", "mle.index", "Model.Index", "Class")
row.names(sorted) <- NULL

classprobs <- rep(NA, length(levels(as.factor(sorted$Class))))
names(classprobs) <- levels(as.factor(sorted$Class))
for(class in levels(as.factor(sorted$Class))){
  tempprobs <- sum(as.numeric(as.character(
  sorted$modprob.FBF[sorted$Class==class])))
```

```
  classprobs[which(names(classprobs)==class)] <- tempprobs
}

schemeprobs <- rep(NA, length(levels(as.factor(sorted$Scheme))))
names(schemeprobs) <- levels(as.factor(sorted$Scheme))
for(scheme in levels(as.factor(sorted$Scheme))){
  tempprobs <- sum(as.numeric(as.character(
  sorted$modprob.FBF[sorted$Scheme==scheme])))
  schemeprobs[which(names(schemeprobs)==scheme)] <- tempprobs
}

schemeprobs <- data.frame("Scheme.Prob"=sort(schemeprobs, decreasing=TRUE))
classprobs <- data.frame("Class.Prob"=sort(classprobs, decreasing=TRUE))

if(prior=="flat"){
  out <- list("results"=sorted, "group.dfs"=group.dfs,
              "class.Probs"=round(classprobs, 8),
              "scheme.Probs"=schemeprobs,
              "coefficients"=all.coefs, "variances"=all.vars)
}
if(prior=="zs"){
  out <- list("results"=sorted, "group.dfs"=group.dfs,
              "class.Probs"=round(classprobs, 8),
              "scheme.Probs"=schemeprobs,
              "coefficients"=all.coefs, "variances"=all.vars, "gs"=all.gs)
}
class(out) <- "slgf"
return(out)
```

```
}
```