

6-30-2016

## Development and Application of Bayesian Semiparametric Models For Dependent Data

Junshu Bao  
*University of South Carolina*

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>



Part of the [Statistics and Probability Commons](#)

---

### Recommended Citation

Bao, J.(2016). *Development and Application of Bayesian Semiparametric Models For Dependent Data*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/3489>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [digres@mailbox.sc.edu](mailto:digres@mailbox.sc.edu).

DEVELOPMENT AND APPLICATION OF BAYESIAN SEMIPARAMETRIC MODELS  
FOR DEPENDENT DATA

by

Junshu Bao

Bachelor of Arts  
Dongbei University of Finance and Economics 2000

Master of Science  
Columbia University 2003

Master of Arts  
Wayne State University 2007

---

Submitted in Partial Fulfillment of the Requirements  
for the Degree of Doctor of Philosophy in  
Statistics

College of Arts and Sciences  
University of South Carolina

2016

Accepted by:

Timothy E. Hanson, Major Professor

Xianzheng Huang, Committee Member

Lianming Wang, Committee Member

Jiajia Zhang, Committee Member

Lacy Ford, Senior Vice Provost and Dean of Graduate Studies

© Copyright by Junshu Bao, 2016  
All Rights Reserved.

## ACKNOWLEDGMENTS

First and foremost, I am forever grateful to my advisor Dr. Timothy Hanson, for his constant support of my Ph.D. study and for his patience, motivation, and immense knowledge. His relentless passion and research vision has greatly influenced me. It has always been a great pleasure working with him and I am truly fortunate and at the same time proud to have him as my advisor.

Besides my advisor, I would like to thank my other dissertation committee members: Dr. Xianzheng Huang, Dr. Lianming Wang, and Dr. Jiajia Zhang for their insightful comments and encouragement. My sincere thanks also go to all professors in the Department of Statistics. I also thank my fellow graduate students in the department who always support and encourage me.

I give my whole-hearted thanks to my dear husband and my beautiful daughters. I would not have completed this long journey without their unconditional love and support. They have always been by my side.

## ABSTRACT

Dependent data are very common in many research fields, such as medicine (repeated measures), finance (time series), traffic (clustered), etc. Effective control/modeling of the dependency among data can enhance the performance of the models and result in better prediction. In many cases, the correlation itself may be of great interest. In this dissertation, we develop novel Bayesian semi-/nonparametric regression models to analyze data with various dependence structures. In Chapter 2, a Bayesian nonparametric multivariate ordinal regression model is proposed to fit drinking behavior survey data from DWI offenders. The responses are two-dimensional ordinal data, drinking frequency and drinking quantity at each occasion. In Chapter 3, we develop a hierarchical Gaussian process model to analyze repeated hearing test data of pediatric cancer patients. A penalized B-spline is used to capture the overall trend of the curve. Individual intercepts and slopes as random effects are allowed to model individual deviation from the population mean. Since the curves are theoretically smooth, a hierarchical Gaussian process is assumed on top of the individual-specific mean function. In Chapter 4, we propose a constructive approach to imposing a mean constraint in a finite mixture of multivariate normal densities. Implemented in a linear mixed model, the effectiveness of the constraint is verified by both simulation and data analysis using longitudinal cholesterol data.

# TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . . .	iii
ABSTRACT . . . . .	iv
LIST OF TABLES . . . . .	viii
LIST OF FIGURES . . . . .	x
CHAPTER 1 INTRODUCTION . . . . .	1
1.1 Dependent Data . . . . .	1
1.2 Multivariate Probit Models . . . . .	2
1.3 Linear Mixed Models . . . . .	3
1.4 Curve Fitting . . . . .	4
1.5 Stick-Breaking Mixtures . . . . .	6
1.6 Summary . . . . .	8
CHAPTER 2 BAYESIAN NONPARAMETRIC MULTIVARIATE ORDINAL RE- GRESSION . . . . .	10
2.1 Introduction . . . . .	10
2.2 Multivariate Probit Model . . . . .	12
2.3 Nonparametric extension . . . . .	17
2.4 Interpretation and model comparison . . . . .	23

2.5	Simulation study . . . . .	26
2.6	Data analysis . . . . .	28
2.7	Conclusions . . . . .	32
CHAPTER 3 ASSESSMENT OF DPOAE TEST-RETEST DIFFERENCE CURVES VIA HIERARCHICAL GAUSSIAN PROCESSES . . . . .		38
3.1	Introduction . . . . .	38
3.2	Hierarchical Gaussian process model . . . . .	45
3.3	Volume tube method for obtaining reference chart . . . . .	51
3.4	Posterior contour probabilities for finding children with abnormal DP-grams . . . . .	53
3.5	Age-gender-specific model . . . . .	54
3.6	Data analysis . . . . .	55
3.7	Concluding remarks . . . . .	61
CHAPTER 4 A MEAN-CONSTRAINED FINITE MIXTURE OF NORMALS . . . . .		62
4.1	Introduction . . . . .	62
4.2	Model . . . . .	63
4.3	A Simulation study . . . . .	67
4.4	Data analysis . . . . .	69
4.5	Conclusions . . . . .	71
BIBLIOGRAPHY . . . . .		73
APPENDIX A SUPPLEMENT TO CHAPTER 2 . . . . .		82
A.1	Derivation of the Parametric Model . . . . .	82

A.2	Derivation of the Nonparametric Model . . . . .	88
A.3	Derivation of the Nonparametric Model with Hyperpriors . . . . .	96
A.4	Sample Fortran 90 code . . . . .	107
APPENDIX B SUPPLEMENT TO CHAPTER 3 . . . . .		125
B.1	Supplementary Table and Figures . . . . .	125
B.2	Hierarchical Model . . . . .	134
B.3	Age-Gender Model . . . . .	144
B.4	Linear Mixed Model with Individual Variances . . . . .	148
B.5	Linear Mixed Model with Overall Common Variance . . . . .	153
B.6	Hierarchical Model with Correlation Among f2 . . . . .	153
B.7	Age-gender Model with Correlation Among f2 . . . . .	158
B.8	Sample FORTRAN 90 Code . . . . .	163
APPENDIX C PERMISSION TO REPRINT OF CHAPTER 2 . . . . .		176



## LIST OF TABLES

Table 2.1	LPML and MED of parametric (P) and nonparametric (NP) models; simulated data. . . . .	34
Table 2.2	Posterior means and standard deviations from parametric and nonparametric models (only non-empty components); simulated bivariate normal data, $T = 2000$ . . . . .	35
Table 2.3	Posterior means and standard deviations from parametric and nonparametric models (only non-empty components); simulated mixture of bivariate normal data, $T = 2000$ . . . . .	35
Table 2.4	DWI offender alcohol quantity-frequency data by age, gender, and history of physical/sexual abuse. . . . .	36
Table 2.5	Posterior means and standard deviations from reduced parametric and nonparametric models (only non-empty components are included); drinking data. . . . .	36
Table 2.6	Posterior means and standard deviations of average causal effects from reduced parametric and nonparametric models; drinking data. . . . .	37
Table 2.7	Posterior means and standard deviations of polychoric correlations from reduced parametric and nonparametric models; drinking data. . . . .	37
Table 3.1	Subject characteristics . . . . .	46
Table 4.1	Posterior means and standard errors (in brackets) from MCFM LMM; simulated data, $n = 200$ subjects, weights=(0.7, 0.3) . . . . .	68
Table 4.2	Posterior means and standard errors (in brackets) from the MCFM LMM; simulated data, $n = 200$ subjects, weights=(0.5, 0.5) . . . . .	69
Table 4.3	Posterior means and standard errors (in brackets) by the Gaussian LMM and MCFM LMM; Framingham cholesterol data. . . . .	70

Table B.1	Number of measurements by f2 Primary . . . . .	125
-----------	--	-----

## LIST OF FIGURES

Figure 2.1	Scatter plots of $T = 1000$ simulated latent $\tilde{z}$ . Left: mixture of two bivariate normals. Right: bivariate normal with covariate-dependent correlation $\rho = \frac{1-e^x}{1+e^x}$ . . . . .	28
Figure 2.2	Contour plots of true and estimated densities from three models; latent variables simulated from bivariate normal with covariate-dependent correlation ( $E[(Z_1, Z_2)] = (0.5, 0.5)$ , $\rho = \frac{1-e^x}{1+e^x}$ where $x = -2$ ). . . . .	29
Figure 2.3	Contour plots of estimated densities; parametric reduced model; drinking data. . . . .	31
Figure 2.4	Contour plots of estimated densities; nonparametric reduced model, drinking data. . . . .	31
Figure 3.1	Example DP-grams for 18-month old boy treated with cisplatin at baseline and about 4, 10, and 56 weeks later. . . . .	41
Figure 3.2	95% credible bands and 10 sample DP-grams . . . . .	56
Figure 3.3	10 sample DP-grams of test-retest differences of 5 children and 95% simultaneous credible band; followup time = 1 month. . . . .	57
Figure 3.4	Half widths of credible bands of test-retest differences for all children and for girls. . . . .	58
Figure 3.5	Half widths of credible bands of test-retest differences for all children and for boys. . . . .	59
Figure 3.6	Test-retest difference simultaneous bands at 4, 10, and 56 weeks (upper left, upper right, lower left) along with actual difference for 18-month male cancer patient; credible band for one DP-gram with corresponding DP-grams (lower right). . . . .	60

Figure 4.1	MCFM LMM estimates of random intercepts and slopes, with sample size $n = 50$ and $200$ , weights=( $0.7, 0.3$ ) (top four panels) and ( $0.5, 0.5$ ) (bottom four panels). Estimated distributions (dashed line) are superimposed to the true distributions (solid line). . . . .	69
Figure 4.2	Estimated random effects distribution from the MCFM LMM fitted to the cholesterol data. . . . .	71
Figure B.1	Followup time (months). . . . .	126
Figure B.2	Individual DP-grams (IDs 1 to 10). . . . .	127
Figure B.3	Individual DP-grams (IDs 11 to 20). . . . .	128
Figure B.4	Individual DP-grams (IDs 21 to 30). . . . .	129
Figure B.5	Individual DP-grams (IDs 31 to 38). . . . .	130
Figure B.6	Posterior means of $b_0, b_1, \log(\sigma^2), \log(\theta_t), \log(\theta_e)$ versus age. . . . .	131
Figure B.7	Extreme test-retest DP-gram differences with contour probabilities; Case 1: oscillatory. . . . .	131
Figure B.8	Extreme test-retest DP-gram differences with contour probabilities; Case 2: overall vertical shift. . . . .	132
Figure B.9	Extreme test-retest DP-gram differences with contour probabilities; Case 3: monotone. . . . .	132
Figure B.10	Test-retest differences in DP-grams from baseline. . . . .	133

# CHAPTER 1

## INTRODUCTION

### 1.1 DEPENDENT DATA

Independence among data is a common assumption in regression analysis. Nevertheless, in reality, data can be dependent in various ways. Generally speaking, there are two types of dependency in data. Firstly, subjects could be dependent on each other in the sample. For example, a researcher wants to estimate the average base salary of all employees in a company. Instead of doing random sampling, he collects data from one particular research group. Conditioning on that they are from this group, the employees' data could be treated as independent. However, marginally, the data are highly correlated because they are from the same group. The second type of dependency is within subjects. For instance, a patient takes a hearing test on a series of tone frequencies for both left and right ears. Test results of each subject are naturally correlated. This is called repeated measures. In particular, if the measurements are taken over time, it is so-called longitudinal data. For example, a patient's blood pressure is tested repeatedly over a period of time during treatment. Another type of dependence is when one subject is measured on different outcomes. For example, a person may fill out a survey asking about her exercise behavior which includes information about both exercise duration and frequency.

These dependent data are problematic for the classical methods of statistical analysis. There is a rich literature of methodology dealing with dependent data. One branch is multivariate analysis. Tiao and Zellner (1964) use a Bayesian approach to

analyze correlated responses. The model is quite flexible allowing the set of predictors to be different for different response variables. Generalized versions of multivariate models were also proposed to analyze categorical and binary data (Chib and Greenberg, 1998 and O’Brian and Dunson, 2004). In longitudinal studies, data collected may be unbalanced. For example, Individual A may be measured at  $t_1, t_2, \dots, t_n$  but Individual B are not measured at some time points. In this case, it is difficult to apply multivariate models with a general covariance matrix. Laird and Ware (1982) propose a family of mixed models, which assume an overall mean for the whole population but allow variation of individual parameters (random effects) that induce correlation among repeated measures within each subject.

In this introduction, I discuss selected models and tools that are used in the development of the proposed models. Section 1.2 contains a brief introduction to multivariate probit models that are used on dependent data for subjects measured on the same responses. Linear mixed models are discussed in Section 1.3; these models can handle unbalanced data as well as high-dimensional repeated measurements. Section 1.4 extends the discussion to curve fitting methods and models. In particular, Gaussian Process models and the B-spline approach are introduced. A general overview of the three projects comprising this dissertation is contained in Section 1.6.

## 1.2 MULTIVARIATE PROBIT MODELS

Tiao and Zellner (1964) propose the general  $m$ -dimensional multivariate regression model as follows:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, i = 1, \dots, m, \tag{1.1}$$

where  $\mathbf{y}_i$  is a  $T \times 1$  vector of all  $T$  responses on outcome  $i$ ,  $\mathbf{X}_i$  a  $T \times p_i$  matrix of regressors of rank  $p_i$  including intercepts,  $\boldsymbol{\beta}_i$  the corresponding vector of regression coefficients and  $\mathbf{e}_i$  a  $T \times 1$  vector of random errors. Let  $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_m)'$  and

assume  $\mathbf{e}$  is normally distributed with zero mean and covariance matrix  $\boldsymbol{\Sigma} \otimes \mathbf{I}_T$  where  $\boldsymbol{\Sigma}$  is a  $m \times m$  positive definite matrix,  $\mathbf{I}_T$  a  $T \times T$  identity matrix and  $\otimes$  denotes the Kronecker product. Write  $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_m)'$ ,  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_m)$  and  $\mathbf{X} = \text{blockdiag}(\mathbf{X}_1, \dots, \mathbf{X}_m)_{Tm \times p+}$  where  $p+ = \sum_{i=1}^m p_i$ . Then, the multivariate model is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where  $\mathbf{e} \sim N_{Tm}(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_T)$ . If  $\mathbf{X}_1 = \mathbf{X}_2 = \dots = \mathbf{X}_m$ , the model can be reduced to the traditional multivariate regression model.

### 1.3 LINEAR MIXED MODELS

Laird and Ware (1982) propose a two-stage random-effects model for the analysis of longitudinal data. Let  $\boldsymbol{\beta}$  denote the population parameter vector and  $\boldsymbol{\gamma}_i$  denote the individual effects. Let  $\mathbf{X}_i$  and  $\mathbf{Z}_i$  denote the design matrices that link  $\mathbf{y}_i$  with  $\boldsymbol{\beta}$  and  $\boldsymbol{\gamma}_i$ , respectively. For each subject  $i$ ,

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\gamma}_i + \mathbf{e}_i,$$

where  $\mathbf{e}_i$  is distributed as  $N(\mathbf{0}, \boldsymbol{\Sigma}_i)$ .  $\boldsymbol{\beta}$  is treated as fixed effects while  $\boldsymbol{\gamma}_i$  are random effects,

$$\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \mathbf{D}).$$

Then, marginally,  $\mathbf{y}_i$  are independent normals with mean and covariance matrix as follows:

$$E(\mathbf{y}_i) = \mathbf{X}_i\boldsymbol{\beta}$$

and

$$\text{Var}(\mathbf{y}_i) = \boldsymbol{\Sigma}_i + \mathbf{Z}_i\mathbf{D}\mathbf{Z}'_i$$

Often times, it is assumed that  $\boldsymbol{\Sigma}_i = \sigma_i^2 \mathbf{I}_{n_i \times n_i}$ , where  $\mathbf{I}$  denotes the identity matrix and  $n_i$  is the total number of observations of individual  $i$ .

Note that linear mixed models may be easily generalized to models which can handle categorical or Poisson data, termed generalized linear mixed models. They can also be generalized in another direction: relaxing the normal assumption of the error terms. These semi-parametric or non-parametric models can handle data with more complicated structures, such as clustered, severely skewed, bimodal, etc.

## 1.4 CURVE FITTING

Generally speaking, there are three major components in curve fitting models: a regression function for the overall trend, individual parameters modeling the differences among subjects or curves (e.g. random effects), and the variance-covariance structure for the repeated within-subject observations. Different approaches emphasize different components, but they have gradually evolved in a more flexible and non-parametric direction. In this section, I introduce nonlinear models that are employed in my projects.

Instead of assuming  $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$  in linear models, let  $E(\mathbf{Y}) = \mu(\mathbf{X}, \boldsymbol{\beta})$  where  $\mu(\cdot)$  is an unknown nonlinear function. In particular, a basis function expansion assumes

$$\mu(x) = \sum_{h=1}^H \beta_h b_h(x),$$

where  $b_h$  are prespecified basis functions and  $\beta_h$  are basis coefficients.  $b_h$  can take a variety of forms. We will focus on the penalized B-spline approach in this introduction. The B-spline is piecewise continuous function which is defined over some prespecified set of *knots*. Often times, these knots are equispaced over the range of the predictor  $x$ . For example, the “mother” quadratic B-spline basis function on  $[0, 3]$



is defined as follows:

$$\phi(x) = \begin{cases} 0.5x^2 & 0 \leq x \leq 1 \\ 0.75 - (x - 1.5)^2 & 1 \leq x \leq 2 \\ 0.5(3 - x)^2 & 2 \leq x \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

And the basis functions  $b_h(\cdot)$  on the range of predictor  $x$  are

$$b_h(x) = \phi\left(\frac{x - a}{\delta} + 3 - h\right)$$

where  $\delta = (b - a)/(H - 2)$ ,  $a = \min(x_1, \dots, x_n)$  and  $b = \max(x_1, \dots, x_n)$ .

The number of knots is determined by how oscillatory the curve is. A small number of knots will result in a function space that is not flexible enough to sufficiently model the fluctuation of the curve. Too many knots will make the model too complicated and computation-intensive and may also result in overfitting. To solve this tricky problem efficiently, Eilers and Marx (1996) propose a roughness penalty approach. First, a relatively large number (e.g. 20) of equally spaced knots are chosen to ensure the flexibility of the function space. Then the roughness is penalized through a difference penalty on adjacent B-spline coefficients (Lang and Brezger 2004), equivalent to penalizing  $\int_a^b |\mu''(x)|^2 dx$ .

The fitting of the B-spline models may be sensitive to the location and number of knots. An alternative choice is to set up a prior on the regression function directly using a Gaussian process:

$$\mu \sim \text{GP}(m, k),$$

where  $m$  is a mean function and  $K$  is a covariance function. For any prespecified set  $\{x_1, \dots, x_n\}$ ,  $\{\mu(x_i)\}_{i=1}^n$  follow a multivariate normal distribution as follows:

$$\mu(x_1), \dots, \mu(x_n) \sim N_n((m(x_1), \dots, m(x_n)), K(x_1, \dots, x_n)).$$

The mean function is an initial guess at the regression function (Gelman, 2014), which could be the vector zero, a linear function  $\mathbf{X}\boldsymbol{\beta}$ , or even a B-spline function, depending on how much information is available about the curve. The function  $k$  determines the covariance between any two functional values  $\mu(x)$  and  $\mu(x')$ . A common choice is the Gaussian covariance function

$$k(x, x') = \gamma^2 \exp(-\theta|x - x'|^2),$$

where  $\gamma$  and  $\theta$  are unknown parameters and  $|x - x'|^2$  is the squared Euclidean distance between  $x$  and  $x'$ .  $\gamma$  controls the magnitude and  $\theta$  the smoothness of the function.

## 1.5 STICK-BREAKING MIXTURES

In classical regression models, error terms are assumed to be normally distributed. This assumption is strong and often times inappropriate. To model data with uncertain or non-standard distributions, a variety of semi-/non-parametric approaches have been proposed, among which are models based on mixtures of standard components, such as normal mixtures. In this dissertation, we adopt mixture models which assume stick-breaking priors.

Stick-breaking priors are random probability measures  $G$  defined as

$$G(\cdot) = \sum_{k=1}^N w_k \delta_{\theta_k}(\cdot), \quad (1.2)$$

where  $\delta_{\theta_k}(\cdot)$  denotes a point mass at  $\theta_k$  and  $w_k$  are random weights with special structure (below) that are independent of  $\theta_k$  and such that  $0 \leq w_k \leq 1$  and  $\sum_{k=1}^N w_k = 1$  almost surely (Ishwaran and James, 2001). Let  $\mathbf{a} = (a_1, a_2, \dots)$  and  $\mathbf{b} = (b_1, b_2, \dots)$  be positive vectors.  $G = \mathbf{G}_N(\mathbf{a}, \mathbf{b})$  is called a stick-breaking random measure if it is of form (1.2) and

$$w_1 = V_1 \text{ and } w_k = V_k \prod_{l < k} (1 - V_l), \quad k \geq 2,$$

where  $V_k \sim \text{Beta}(a_k, b_k)$ . Note that if  $a_1 = a_2 = \dots$  and  $b_1 = b_2 = \dots$  the weights are stochastically ordered, i.e.  $E(w_1) \geq E(w_2) \geq \dots$

The number of  $V_k$ 's,  $N$ , could be either finite or infinite. In the case that  $N < \infty$ , setting  $V_N = 1$  guarantees that  $\sum_{k=1}^N w_k = 1$ . Random weights defined in this way have the generalized Dirichlet distribution, which is conjugate to multinomial distribution. In the case that  $N = \infty$ , the prior is infinite dimensional and can be denoted as  $\mathbf{G}_\infty(\mathbf{a}, \mathbf{b})$ . One example of infinite-dimensional priors is the Dirichlet process.

Ishwaran and James (2001) discuss two Gibbs sampling methods for stick-breaking priors. In the first method, samples are drawn from the posterior of a hierarchical model formed by marginalizing over the prior. In the case of mixtures of normals,

$$P(\mathbf{y}|\text{else}) = \int \phi(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma})dG(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

This is called ‘‘Polya urn Gibbs sampler’’ and can be applied to any random probability measure with a known prediction rule.

The second method, the ‘‘blocked Gibbs sampler’’, avoids marginalizing over the prior but directly samples from the nonparametric posterior. This method works even if the prediction rule is unknown. The prior  $G$  is assumed to be a finite dimensional  $\mathbf{G}_N(\mathbf{a}, \mathbf{b})$  so that the model can be expressed entirely in terms of a finite number of components. In particular, the mixture model can be written as follows:

$$\begin{aligned} Y_i|\boldsymbol{\theta}, \mathbf{s} &\stackrel{ind}{\sim} \pi(Y_i|\theta_{s_i}), \quad i = 1, \dots, n \\ s_i|\mathbf{w} &\stackrel{iid}{\sim} \sum_{k=1}^N w_k \delta_k(\cdot) \\ \mathbf{w}, \boldsymbol{\theta} &\sim \pi(\mathbf{w}) \times H^N(\boldsymbol{\theta}) \end{aligned}$$

where  $\mathbf{s} = (s_1, \dots, s_n)$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_N)$  and  $\mathbf{w} = (w_1, \dots, w_N)$ .  $s_i$  are component membership indicators that identify the  $\theta_k$  associated with  $Y_i$ ,  $\theta_k$  are iid  $H$ , and  $w_k$  follow generalized Dirichlet distribution.

## 1.6 SUMMARY

In this dissertation, we develop Bayesian semiparametric models to analyze dependent data. In Chapter 2, multivariate ordinal data are modeled as a finite stick-breaking mixture of multivariate probit models. Parametric multivariate probit models are first developed for ordinal data, then generalized to finite mixtures of multivariate probit models. Specific recommendations for prior settings are found to work well in simulations and data analyses. Interpretation of the model is carried out by examining aspects of the mixture components as well as through averaged effects focusing on the mean responses. A simulation verifies that the fitting technique works, and an analysis of alcohol drinking behavior data illustrates the usefulness of the proposed model.

Distortion product otoacoustic emissions (DPOAE) testing is a promising alternative to behavioral hearing tests and auditory brainstem response testing of pediatric cancer patients. The central goal of the study in Chapter 3 is to assess whether significant changes in the DPOAE frequency/emissions curve (DP-gram) occur in pediatric patients in a test-retest scenario. This is accomplished through the construction of normal reference charts, or credible regions, that DP-gram differences lie in, as well as contour probabilities that measure how abnormal (or in a certain sense rare) a test-retest difference is. A challenge is that the data were collected over varying frequencies, at different time points from baseline, and on possibly one or both ears. A hierarchical structural equation Gaussian process model is proposed to handle the different sources of correlation in the emissions measurements, wherein both subject-specific random effects and variance components governing the smoothness and variability of each child's Gaussian process are coupled together.

In Chapter 4, a simple constructive approach to imposing a mean constraint in a finite mixture of multivariate Gaussian densities is proposed. All parameters in the model except for one have closed-form full conditional distributions and are fit

through a simple Markov chain Monte Carlo algorithm. For illustration, the mean-constrained finite mixture is implemented in a linear mixed model. Simulations reveal that the mean-constrained model is able to precisely estimate the regression coefficients and mean-constrained random effects distribution simultaneously. An analysis of the Framingham cholesterol data shows that, with relatively simple structure, the model has competitive predictive power with earlier approaches.

# CHAPTER 2

## BAYESIAN NONPARAMETRIC MULTIVARIATE ORDINAL REGRESSION<sup>1</sup>

### 2.1 INTRODUCTION

This chapter proposes a nonparametric regression model and associated inferential methods for multivariate ordinal data. The development is quite general, allowing for different sets of covariates to be tied to different ordinal outcomes. Specifically, we observe  $m$  ordinal outcomes from each of  $T$  subjects in a study. The model is immediately generalized to include mixtures of ordinal and continuous responses. Our approach makes use of recent work in dependent prior processes, notably the dependent Dirichlet process (MacEachern, 1999), but is easily simplified to an interpretable finite mixture of multivariate probit models, as we demonstrate with our data analysis in Section 2.6. The model presented here builds on and generalizes work by Tiao and Zellner (1964); Kottas, Müller, and Quintana (2005); Lawrence, Bingham, Liuc, and Naird (2008); and De Iorio, Johnson, Müller, and Rosner (2009).

Historically, there are two distinct approaches to regression modeling for multivariate multinomial data, models that assume underlying continuous latent traits (threshold models), and models that focus on the discrete joint distribution of the responses directly in terms of cumulative logits and/or odds ratios. So-called threshold models have their origins with Pearson (1904) who assumed latent bivariate normal

---

<sup>1</sup>The content in this chapter is a reprint by permission of John Wiley & Sons, Inc. for “Junshu Bao and Timothy Hanson (2015). Bayesian Nonparametric Multivariate Ordinal Regression. Canadian Journal of Statistics.”

random variables underlying a two-way contingency table, leading to the tetrachoric correlation measure of association for  $2 \times 2$  tables and the more general polychoric correlation for larger tables. In contrast, Yule (1900) proposed a direct measure, Yule’s  $Q$ , for  $2 \times 2$  tables that is a measure of concordance among two binary variables. Agresti (2013, Chapter 17) reviews the history of categorical data analysis including early disputes among Pearson, Yule, and Fisher. An example of a recent regression formulation for multivariate ordinal data based on the threshold approach is Lesaffre and Molenberghs (1991) and references therein; the direct approach is nicely typified by Dale (1986).

Although some work has been done on nonparametric approaches to univariate ordinal data (e.g. Hastie, Botha, and Schnitzler, 1989; Chu and Ghahramani, 2005; Srijith, Shevade, and Sundararajan, 2013), there is very little in terms of nonparametric approaches to multivariate ordinal data. Kottas et al. (2005) consider a Dirichlet process mixture (DPM) of normals for modeling underlying latent traits without covariates using the threshold approach. Yang (2006) also posited an underlying Bayesian nonparametric prior, a mixture of multivariate Polya trees, adding group offsets yielding a semiparametric MANOVA model, with application to bivariate ordinal drinking behavior data. De Yoreo and Kottas (2014) consider the threshold model applied to the multivariate regression model of Müller, Erkanli, and West (1995). Their model specifies a joint DPM model for the predictors and latent continuous response  $(\tilde{\mathbf{x}}, \tilde{\mathbf{z}})$  that induces a conditional model  $\tilde{\mathbf{z}}|\tilde{\mathbf{x}}$ ; this approach is most natural for strictly continuous  $\tilde{\mathbf{x}}$ . We consider a generalization of Kottas et al. (2005) allowing for the incorporation of categorical and continuous covariates in a linear dependent Dirichlet process mixture (LDDPM) model that can be viewed as a discrete mixture of multivariate probit regressions.

Following Chib and Greenberg (1998) the most general case where each ordinal outcome is modeled using a different set of predictor variables is initially considered;

this general parametric model is then further extended to mixtures of probit models. However, we restrict ourselves to the case where each ordinal outcome has at least three levels. Our model can handle the case where one or more outcomes are dichotomous, but the incorporation of additional constraints are necessary (e.g. Hanson and Pearson, 2007) or a parameter expanded model needs to be considered (e.g. Lawrence et al., 2008).

Section 2.2 develops the multivariate probit model in detail; Section 2.3 generalizes the parametric model to a nonparametric finite mixture. Model interpretation and comparison is discussed in Section 2.4. Section 2.5 gives a simulation study for the parametric and nonparametric models; Section 2.6 offers an analysis on bivariate ordinal data on beer drinking quantity and frequency. Finally, Section 2.7 concludes the chapter with a discussion.

## 2.2 MULTIVARIATE PROBIT MODEL

### Parametric multivariate probit model for latent traits

There are many ways to parameterize a threshold ordinal regression model; one approach constrains the covariance matrix of the underlying latent trait vector to be a correlation matrix (e.g. Chib and Greenberg, 1998). This approach disallows easy posterior computation, and so we instead follow Chen, Shao, and Ibrahim (2000) and fix the two outer cutoffs. Kottas et al. (2005) consider a flexible mixture model underlying the table and so are able to fix *all* cutoffs in the model. The  $i$ th ordinal outcome from subject  $j$ ,  $y_{ij}$ , follows a latent trait model:

$$y_{ij} = r \Leftrightarrow \alpha_{i,r-1} < z_{ij} < \alpha_{i,r}, \quad (2.1)$$

where  $\alpha_{i0} = -\infty$ ,  $\alpha_{i1} = 0$ ,  $\alpha_{i,c_i-1} = 1$  and  $\alpha_{i,c_i} = \infty$  ( $c_i$  is the number of levels of the  $i$ th ordinal outcome). Let  $\mathbf{y}_i = (y_{i1}, \dots, y_{iT})'$  be the ordinal outcomes for measurement  $i$  on individuals  $j = 1, \dots, T$  and let  $\mathbf{z}_i = (z_{i1}, \dots, z_{iT})'$  be the underlying latent



variables associated with  $\mathbf{y}_i$ . For each outcome  $i$  there are  $c_i - 3$  free parameters  $\alpha_{i2} < \dots < \alpha_{i,c_i-2}$  to estimate.

Consider the general multivariate regression model for the latent  $\mathbf{z}_i$  described in Tiao and Zellner (1964). Let  $\mathbf{X}_i$  be a  $T \times p_i$  matrix of regressors of rank  $p_i$  including intercepts, and  $\boldsymbol{\beta}_i$  the corresponding vector of regression coefficients for ordinal outcome  $i$ . Let  $\mathbf{e} = (\mathbf{e}'_1, \dots, \mathbf{e}'_m)'$ . The model is

$$\mathbf{z}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{e}_i, \quad \mathbf{e} \sim N_{Tm}(\mathbf{0}_{Tm}, \boldsymbol{\Sigma}_{m \times m} \otimes \mathbf{I}_T), \quad (2.2)$$

where  $N_{Tm}$  stands for a  $T \times m$ -dimensional multivariate normal,  $\mathbf{0}_{Tm}$  is a  $T \times m$ -dimensional vector of zero,  $\boldsymbol{\Sigma}_{m \times m}$  is the variance-covariance matrix of latent  $\mathbf{z}_i$ , and  $\mathbf{I}_T$  is the  $T$ -dimensional identity matrix. Note that this allows for different predictors for each latent outcome unlike the approaches of Lawrence et al. (2008) and Biswas and Das (2002). Now take  $\mathbf{E} = [\mathbf{e}_1 \cdots \mathbf{e}_m]$  to be the  $T \times m$  matrix of errors. The  $i = 1, \dots, m$  columns are of course  $\mathbf{e}_i$ ; let the  $j = 1, \dots, T$  row vectors be denoted  $\tilde{\mathbf{e}}'_j$  so that  $\mathbf{E}' = [\tilde{\mathbf{e}}_1 \cdots \tilde{\mathbf{e}}_T]$ . Note that  $\mathbf{E}$  is matrix-normal; the matrix-normal formulation was used by Geisser (1965) for the simplified model where each outcome has the same set of predictors  $\mathbf{X}_1 = \dots = \mathbf{X}_m$ .

Let  $\tilde{\boldsymbol{\mu}}_j = E\{\tilde{\mathbf{z}}_j | \boldsymbol{\beta}\}$  be the mean of  $m$  latent outcomes  $\tilde{\mathbf{z}}_j = (z_{1j}, \dots, z_{mj})'$  from subject  $j$  given  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_m)'$ . Let  $\mathbf{X}_i$  have  $j$ th row vector  $\tilde{\mathbf{x}}_{ij}$  and let  $\tilde{\mathbf{X}}_j = \text{blockdiag}(\tilde{\mathbf{x}}_{1j}, \dots, \tilde{\mathbf{x}}_{mj})_{m \times p_+}$  where  $p_+ = \sum_{i=1}^m p_i$ ; then  $\tilde{\boldsymbol{\mu}}_j = \tilde{\mathbf{X}}_j \boldsymbol{\beta}$ . An alternative formulation of (2.2) is

$$\tilde{\mathbf{z}}_j = \tilde{\mathbf{X}}_j \boldsymbol{\beta} + \tilde{\mathbf{e}}_j, \quad \tilde{\mathbf{e}}_1, \dots, \tilde{\mathbf{e}}_T \stackrel{iid}{\sim} N_m(\mathbf{0}_m, \boldsymbol{\Sigma}). \quad (2.3)$$

## Posterior updating under Jeffreys' prior

As our approach is Bayesian, we obtain posterior inference through Gibbs sampling. By now, Gibbs sampling is a standard tool in the statistician's toolbox and we refer the reader to textbooks such as Christensen et al. (2010) or Robert

and Casella (2004). A Gibbs sampler is characterized by the full conditional distributions of all model parameters. Once these full conditional distributions are derived, a Gibbs sampler simply samples from them in turn, updating parameter values as they are sampled. The updated parameter values comprise a dependent sample from the posterior, albeit one whose ergodic averages converge to posterior means and empirical quantiles converge to posterior quantiles. Under the parametric model the posterior distribution augmented with the latent  $\{\mathbf{z}_i\}$  is denoted  $[\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{z}_1, \dots, \mathbf{z}_m, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m | \mathbf{y}_1, \dots, \mathbf{y}_m]$ . A Gibbs sampler yields iterates from the augmented posterior  $\{(\boldsymbol{\beta}^{(d)}, \boldsymbol{\Sigma}^{(d)}, \mathbf{z}_1^{(d)}, \dots, \mathbf{z}_m^{(d)}, \boldsymbol{\alpha}_1^{(d)}, \dots, \boldsymbol{\alpha}_m^{(d)})\}_{d=1}^D$  where  $D$  is some large number of iterates after an initial number are discarded (the burn-in).

Let  $\mathbf{B} = \text{blockdiag}(\mathbf{X}_1, \dots, \mathbf{X}_m)$  and  $\mathbf{z} = (\mathbf{z}'_1, \dots, \mathbf{z}'_m)'$ . Under Jeffreys' prior

$$p(\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-\frac{1}{2}(m+1)},$$

Tiao and Zellner (1964) show that

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}, \mathbf{z} \sim N_{p+}(\hat{\boldsymbol{\beta}}, \mathbf{V}^{-1}),$$

where  $\mathbf{V} = [\mathbf{B}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{B}]$  and  $\hat{\boldsymbol{\beta}} = \mathbf{V}^{-1}\mathbf{B}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{z}$ . Note that

$$\mathbf{V} = \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1\boldsymbol{\Sigma}_{11}^{-1} & \mathbf{X}'_1\mathbf{X}_2\boldsymbol{\Sigma}_{12}^{-1} & \cdots & \mathbf{X}'_1\mathbf{X}_m\boldsymbol{\Sigma}_{1m}^{-1} \\ \mathbf{X}'_2\mathbf{X}_1\boldsymbol{\Sigma}_{21}^{-1} & \mathbf{X}'_2\mathbf{X}_2\boldsymbol{\Sigma}_{22}^{-1} & \cdots & \mathbf{X}'_2\mathbf{X}_m\boldsymbol{\Sigma}_{2m}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}'_m\mathbf{X}_1\boldsymbol{\Sigma}_{m1}^{-1} & \mathbf{X}'_m\mathbf{X}_2\boldsymbol{\Sigma}_{m2}^{-1} & \cdots & \mathbf{X}'_m\mathbf{X}_m\boldsymbol{\Sigma}_{mm}^{-1} \end{bmatrix}_{p_+ \times p_+}.$$

Here,  $\boldsymbol{\Sigma}_{ij}^{-1}$  is the  $ij$ th element of the precision matrix  $\boldsymbol{\Sigma}^{-1}$ . In the simplifying case where  $\mathbf{X}_1 = \dots = \mathbf{X}_m = \mathbf{X}$  of dimension  $T \times p$  (e.g. Lawrence et al., 2008) this reduces to

$$\mathbf{V} = [\boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X})]_{p_+ \times p_+} \text{ with inverse } \mathbf{V}^{-1} = [\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}]_{p_+ \times p_+}.$$

In general,

$$\mathbf{B}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T)\mathbf{z} = \begin{bmatrix} \mathbf{X}_1 \sum_{i=1}^m \mathbf{z}_i \boldsymbol{\Sigma}_{i1}^{-1} \\ \vdots \\ \mathbf{X}_m \sum_{i=1}^m \mathbf{z}_i \boldsymbol{\Sigma}_{im}^{-1} \end{bmatrix}.$$

The error matrix as a function of  $\boldsymbol{\beta}$  and  $\mathbf{z}$  is

$$\mathbf{E} = [\mathbf{z}_1 - \mathbf{X}_1 \boldsymbol{\beta}_1 | \cdots | \mathbf{z}_m - \mathbf{X}_m \boldsymbol{\beta}_m]_{T \times m}.$$

Then

$$\boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \mathbf{z} \sim \text{Wish}_m([\mathbf{E}'\mathbf{E}]^{-1}, T).$$

We will use Geweke's (1991) method to sample the latent  $\{z_{ij}\}$ . Let  $\tilde{\mathbf{z}}_{j,-i}$  be the vector  $\tilde{\mathbf{z}}_j$  with the  $i$ th element omitted. Then

$$z_{ij} | \tilde{\mathbf{z}}_{j,-i}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tilde{\mathbf{y}}_j \sim N \left\{ \tilde{\mu}_{ij} - \frac{1}{\boldsymbol{\Sigma}_{ii}^{-1}} \boldsymbol{\Sigma}_{i,-i}^{-1} (\tilde{\mathbf{z}}_{j,-i} - \tilde{\boldsymbol{\mu}}_{j,-i}), \frac{1}{\boldsymbol{\Sigma}_{ii}^{-1}} \right\}$$

truncated to  $(\alpha_{i,y_{ij-1}}, \alpha_{i,y_{ij}})$ . Here,  $\boldsymbol{\Sigma}_{ii}^{-1}$  is the  $i$ th diagonal element of the precision matrix  $\boldsymbol{\Sigma}^{-1}$ ,  $\boldsymbol{\Sigma}_{i,-i}^{-1}$  is the  $(m-1)$ -dimensional  $i$ th row of  $\boldsymbol{\Sigma}^{-1}$  with the  $i$ th element omitted and  $\tilde{\boldsymbol{\mu}}_{j,-i}$  is the  $(m-1)$ -dimensional vector  $\tilde{\boldsymbol{\mu}}_j$  with the  $i$ th element omitted.

That is,

$$z_{ij} | \{z_{lj}\}_{l \neq i}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \tilde{\mathbf{y}}_j \sim N \left\{ \tilde{\mu}_{ij} - \frac{1}{\boldsymbol{\Sigma}_{ii}^{-1}} \sum_{l \neq i} \boldsymbol{\Sigma}_{il}^{-1} (z_{lj} - \tilde{\mu}_{lj}), \frac{1}{\boldsymbol{\Sigma}_{ii}^{-1}} \right\}$$

truncated to  $(\alpha_{i,y_{ij-1}}, \alpha_{i,y_{ij}})$ .

Finally, as usual, given the independent proper prior

$$p(\alpha_{i2}, \dots, \alpha_{i,c_i-2}) \propto I\{0 < \alpha_{i2} < \cdots < \alpha_{i,c_i-2} < 1\},$$

we have

$$\alpha_{ir} | \{z_{ij}\}_{j=1}^T, \tilde{\mathbf{y}}_j \sim \mathcal{U}(\max\{z_{ij} : y_{ij} = r\}, \min\{z_{ij} : y_{ij} = r+1\})$$

for  $i = 1, \dots, m$  and  $2 \leq r \leq c_i - 2$ .  $\mathcal{U}(a, b)$  stands for the uniform distribution with support  $(a, b)$ .

## Posterior updating under proper priors

The mixture approach in Section 2.3 requires proper priors for the posterior to be proper (unless we force a few observations to come from each mixture component). Furthermore, sensible proper priors are easy to develop due to parameterization  $\alpha_{i1} = 0$ ,  $\alpha_{i,c_i-1} = 1$ . So, proper priors were used throughout the simulation study and data analysis for model comparison purpose. Consider conditionally conjugate priors

$$\boldsymbol{\beta} \sim N_{p_+}(\mathbf{b}, \mathbf{M}^{-1}) \text{ ind. } \boldsymbol{\Sigma}^{-1} \sim \text{Wish}_m(\mathbf{S}, \nu). \quad (2.4)$$

Then

$$\boldsymbol{\beta} | \boldsymbol{\Sigma}^{-1}, \mathbf{z} \sim N_{p_+} \{ [\mathbf{V} + \mathbf{M}]^{-1} [\mathbf{B}'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_T) \mathbf{z} + \mathbf{M}\mathbf{b}], [\mathbf{V} + \mathbf{M}]^{-1} \},$$

$$\boldsymbol{\Sigma}^{-1} | \boldsymbol{\beta}, \mathbf{z} \sim \text{Wish}_m([\mathbf{E}'\mathbf{E} + \mathbf{S}^{-1}]^{-1}, T + \nu).$$

See Section A.1 for more detail on posterior derivation for the parametric model.

By fixing the outer cutoffs  $\alpha_{i1} = 0$ ,  $\alpha_{i,c_i-1} = 1$  we constrain the  $\{z_{ij}\}$  to “live” in an interval containing, but not much larger than  $[0, 1]$ . In the worst-case scenario where no covariate affects the mean of  $y_{ij}$ , a typical value of  $z_{ij}$  is simply the intercept  $\beta_{i0}$ . An average value is then the middle of the interval  $[0, 1]$ , i.e. 0.5. Thus, in the absence of real prior information, we center the  $m$  intercepts  $\beta_{10}, \dots, \beta_{m0}$  at 0.5. Prior mass should allow a typical value of  $y_{ij}$  to be either 1 or  $c_i$ , so we set the prior standard deviation of  $\beta_{i0}$  to be unity. At the other extreme, if a covariate is highly predictive, the mean should span  $\{1, \dots, c_i\}$  over the range of covariate values. Since we standardize continuous covariates, and otherwise use zero/one dummy variables for categorical predictors, a standard deviation of unity for the non-intercept elements of  $\{\boldsymbol{\beta}_i\}$  allows for such highly predictive covariates. Finally, considering the spacing of unity between the outer non-infinite cutoffs  $\alpha_{i1}$  and  $\alpha_{i,c_i-1}$ , we roughly center each diagonal element of  $\boldsymbol{\Sigma}$  at  $0.2^2$ , corresponding to a precision of 25. Thus, the following

prior is considered throughout

$$\mathbf{S} = \frac{25}{m} \mathbf{I}_{m \times m}, \quad \nu = m, \quad \mathbf{b} = (0.5, 0, \dots, 0)' \in \mathbb{R}^{p+}, \quad \mathbf{M} = \mathbf{I}_{p+ \times p+}.$$

Note that the  $i$ th diagonal element of  $\boldsymbol{\Sigma}^{-1}$ ,  $\Sigma_{ii}^{-1} \sim s_{ii} \chi_{\nu}^2$  where  $s_{ii}$  is the  $i$ th diagonal element of  $\mathbf{S}$ . For  $m = 2$  this gives a 95% equal-tailed probability interval for the standard deviation of  $z_{ij}$  of (0.1, 1.3). We have experimented with many alternative prior settings, including flat priors, and have found inference to be quite robust to prior choice is reasonable sample sizes.

### 2.3 NONPARAMETRIC EXTENSION

## Stick-breaking mixture of probit models

A Dirichlet process generalization of the parametric model (2.3) is given by

$$\tilde{\mathbf{z}}_j | G \stackrel{ind.}{\sim} \int \phi_m(\cdot | \tilde{\mathbf{X}}_j \boldsymbol{\beta}, \boldsymbol{\Sigma}) dG(\boldsymbol{\beta}, \boldsymbol{\Sigma}), \quad G | \alpha, G_0 \sim DP(\alpha G_0), \quad (2.5)$$

where

$$G_0 = N_{p+}(\mathbf{b}, \mathbf{M}^{-1}) \times \text{Wish}_m^{-1}(\mathbf{S}, \nu),$$

where  $DP$  stands for Dirichlet process,  $\phi_m$  is the density function of  $m$ -dimensional normal, and  $\text{Wish}_m^{-1}$  is  $m$ -dimensional inverse-Wishart distribution. The parametric model obtains in the limit  $\alpha \rightarrow 0^+$ . A univariate version of this model on the latent traits, the LDDPM model, was developed by De Iorio et al. (2009) for survival data. Accordingly, the model developed in this section for the latent  $\tilde{\mathbf{z}}_j$  can be easily modified to handle multivariate interval-censored survival data. Definitions of dependent Dirichlet process (DDP) are found in Barrientos et al. (2012, Definitions 1, 2, and 3). A connection between the LDDPM and one special case (Definition 2) of the DDP was illustrated in De Iorio et al. (2009).

There are two approaches to obtaining inference in Dirichlet process mixture models, marginalized or unmarginalized. Marginalized versions make use of the Polya urn

representation of the Dirichlet process predictive measure (Blackwell and McQueen, 1973) whereas unmarginalized versions use the Sethuraman (1994) stick-breaking representation of the Dirichlet process; we concentrate on the latter here as the resulting algorithm is easier to implement. A truncated approximation

$$G^N(\cdot) = \sum_{k=1}^N w_k \delta_{(\beta(k), \Sigma(k))}(\cdot)$$

can instead be used (Ishwaran and Zarepour, 2000; Chung and Dunson, 2009), where  $N$  is chosen to be large enough to adequately model the data. Alternatively, the full model with  $N = \infty$  can be implemented using the slice-sampling trick of Kalli, Griffin, and Walker (2011). Here  $V_N = 1$  and

$$w_k = V_k \prod_{l < k} (1 - V_l) \text{ where } V_1, \dots, V_{N-1} | \alpha \stackrel{iid}{\sim} \text{beta}(1, \alpha).$$

A general stick-breaking prior simply allows a more general  $\text{beta}(\theta, \alpha)$  distribution for the stick-pieces. Introduce component membership indicators such that  $s_j = k$  when  $\tilde{\mathbf{z}}_j$  comes from  $N_m(\tilde{\mathbf{X}}_j' \boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k))$ . The nonparametric model is written hierarchically

$$\begin{aligned} \tilde{\mathbf{z}}_j &\stackrel{ind.}{\sim} N_m\{\tilde{\mathbf{X}}_j \boldsymbol{\beta}(s_j), \boldsymbol{\Sigma}(s_j)\}, \\ P(s_j = k) &= V_k \prod_{l=1}^{k-1} (1 - V_l), \quad k = 1, \dots, N, \\ \{\boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k)\} &\stackrel{iid}{\sim} N_{p+}(\mathbf{b}, \mathbf{M}^{-1}) \times \text{Wish}_m^{-1}(\mathbf{S}^{-1}, \nu), \quad k = 1, \dots, N, \\ V_1, \dots, V_{N-1} &\stackrel{iid}{\sim} \text{beta}(1, \alpha). \end{aligned}$$

The proper parametric prior (2.4) is generalized to

$$\boldsymbol{\beta}(1), \dots, \boldsymbol{\beta}(N) \stackrel{iid}{\sim} N_{p+}(\mathbf{b}, \mathbf{M}^{-1}) \text{ ind. } \boldsymbol{\Sigma}(1), \dots, \boldsymbol{\Sigma}(N) \stackrel{iid}{\sim} \text{Wish}_m^{-1}(\mathbf{S}, \nu). \quad (2.6)$$

A referee has asked about centering the LDDPM at a parametric model, such as the Gaussian. First note that as  $\alpha \rightarrow 0^+$ , the Gaussian model is obtained. Griffin (2011) discusses centering the DPM prior at an overall multivariate Gaussian distribution when  $\alpha > 0$ . Since the number of parameters grows rather quickly with  $N$ , we follow

Hanson (2006) and opt to fix  $\alpha$  at a relatively small value; in the sequel this is  $\alpha = 1$ , but we have found little difference in predictive inference by considering other small values of  $\alpha$ . The truncation level  $N$  can be set by bounding the truncation error of those components “left off” of the sum, i.e. the tail probability  $U_N = \sum_{k=N+1}^{\infty} w_k$ . Ishwaran and Zarepour (2000) show  $E(U_N) = \alpha^N / (1 + \alpha)^N$ . Setting  $N = 10$  and  $\alpha = 1$  gives  $E(U_N) \approx 0.00098$ , so the truncation error is about 0.1%. The precision  $\alpha$  is directly related to the number of distinct clusters  $1 \leq d \leq N$ . In particular,  $E(d|\alpha) \approx \alpha \log\{(\alpha + T)/\alpha\}$ . For instance the sample size  $T = 2000$ , slightly larger than the drinking data, and  $\alpha = 1$  imply that  $E(d) \approx 7.6$ . Under the posterior for all simulations and data analyses the number of active components is  $E\{d|\text{data}\} < 5$ . Therefore, the truncation level  $N = 10$  is adequate for the simulations and data analysis presented herein.

For finite mixtures, the model is not interpretable without further constraints on model parameters; see e.g. McLachlan and Peel (2000) for a discussion. A natural restriction in the multivariate setting is to order the stick-breaking weights  $w_1 \geq w_2 \geq \dots \geq w_N$ . This model offers enhanced interpretability over the traditional model (where the weights are only *stochastically* ordered for  $\alpha \geq 1$  when  $N < \infty$ ), as the first component is associated with the largest subpopulation, the second component associated with the next-largest, etc. Recently a stick-breaking prior that orders the weights has been introduced, the “geometric stick-breaking prior” (Mena, Ruggiero, and Walker, 2011; Fuentes-García, Mena, and Walker, 2009) where  $w_k$  is simply replaced by  $E(w_k) = \lambda(1 - \lambda)^{k-1}$ , where  $\lambda = \frac{1}{1+\alpha}$ . The resulting  $N = \infty$  process still enjoys full support with respect to the weak topology and has many other appealing properties, especially in terms of posterior computation. As we choose relatively small values for  $\alpha$  and the truncation level  $N = 10$ , we want the stick-breaking weights to be more flexible, and so do not model them via just one parameter  $\lambda$ , but instead simply constrain  $w_1 \geq w_2 \geq \dots \geq w_N$ . This is accomplished in the MCMC algorithm

by simply rejecting values of  $V_k$  that destroy the ordering, i.e. simple acceptance sampling.

Note that truncated model (2.5) for the latent  $\{\tilde{\mathbf{z}}_j\}_{j=1}^T$  can be re-cast as

$$\tilde{\mathbf{z}}_j | \boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k) \sim \sum_{k=1}^N w_k N_m(\cdot | \tilde{\mathbf{X}}_j \boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k)),$$

where the stick-breaking prior induces the prior on  $\{(\boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k), w_k)\}_{k=1}^N$ . This is a finite mixture of multivariate probit regression models. The infinite  $N = \infty$  version of this model is commonly termed the “single weights” DDP. The single weights LDDPM was used for univariate outcomes by De Iorio et al. (2004), De Iorio et al. (2009), de Carvalho et al. (2013), and Hanson and Jara (2013); Jara et al. (2011) consider a multivariate version but constrain the covariance matrix to be the same across components. Gelfand et al. (2005), Dunson and Herring (2006), and Barrientos et al. (2012) consider the DDPM model but with atoms that are Gaussian processes rather than hyperplanes; Griffin and Steel (2006) and Chung and Dunson (2009) consider predictor-dependent weights.

The implied mean under the LDDPM model is linear:

$$E(\tilde{\mathbf{z}}_j | \boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k)) = \sum_{k=1}^N w_k \tilde{\mathbf{X}} \boldsymbol{\beta}(k) \stackrel{def}{=} \tilde{\mathbf{X}} \bar{\boldsymbol{\beta}},$$

so although the model does allow for heteroscedasticity in  $\tilde{\mathbf{z}}_j$  across values of  $\tilde{\mathbf{X}}$ , the mean is only as flexible as the functions entering into the design  $\tilde{\mathbf{X}}$ . This naturally leads to the question of how flexible the model actually is. For the model without predictors, Kottas et al. (2005) argue that the discrete mixture of normals can come arbitrarily close to placing any configuration of mass on the simplex into a single contingency table with underlying probabilities  $\mathcal{P} = \{\pi_{y_1 \dots y_m}\}$  where  $\pi_{y_1 \dots y_m} = (Y_{1j} = y_1, \dots, Y_{mj} = y_m)$ . Barrientos et al. (2012) proves full Kullback-Liebler or Hellinger support for the conditional distribution of  $\tilde{\mathbf{z}} | \tilde{\mathbf{X}}$  (thus implying full support on  $\mathcal{P} | \tilde{\mathbf{X}}$ ), but only for predictors that are stochastic processes. de Carvalho et al. (2013)



suggest that linear combinations of flexible basis functions, e.g. B-splines, can well-approximate Gaussian processes, and so should also roughly achieve full Hellinger support. Although this argument also holds for the proposed model, a potentially more relevant result is Theorem 2.1 in Viele and Tong (2002). They use a bracketing entropy technique to show Hellinger consistency of  $p(\tilde{\mathbf{z}}|\tilde{\mathbf{X}})$  for the finite mixture of regressions model under mild conditions. Their theorem essentially states that a finite mixture of regressions can consistently estimate a finite mixture of regressions. In essence, the model is as flexible as desired, but requires the addition of, e.g., B-splines in the predictor.

## Posterior updating for nonparametric model

The Gibbs sampler now samples the model parameters

$[\boldsymbol{\beta}(1), \dots, \boldsymbol{\beta}(N), \boldsymbol{\Sigma}(1), \dots, \boldsymbol{\Sigma}(N), w_1, \dots, w_N, \mathbf{z}_1, \dots, \mathbf{z}_m | \mathbf{y}_1, \dots, \mathbf{y}_m]$ . Following Kottas et al. (2005), we fix the cutoff points at  $\alpha_{i,r} = \frac{r-1}{c_i-2}$ , for  $r = 1, \dots, c_i - 1$ . We now detail the updating of the  $\{V_k\}$ ,  $\{s_j\}$ ,  $\{\boldsymbol{\beta}(k)\}$ , and  $\{\boldsymbol{\Sigma}(k)\}$ . The stick-breaking pieces  $\{V_k\}$  are updated according to

$$V_k | \text{else} \sim \text{beta} \left( n_k + 1, \alpha + \sum_{l>k} n_l \right),$$

where  $n_k = \sum_{j=1}^T I\{s_j = k\}$  are number of observations from component  $k$ . The constraint  $w_1 \geq \dots \geq w_N$  is incorporated through accept/reject. Component membership indicators are multinomial with probabilities proportional to

$$P(s_j = k | \text{else}) \propto w_k \phi_m(\tilde{\mathbf{z}}_j | \tilde{\mathbf{X}}_j \boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k)),$$

for  $k = 1, \dots, N$ . The component covariance matrices are updated independently as

$$\boldsymbol{\Sigma}^{-1}(k) | \text{else} \sim \text{Wish}_m \left( \left[ \sum_{j:s_j=k} (\tilde{\mathbf{z}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\beta}(k)) (\tilde{\mathbf{z}}_j - \tilde{\mathbf{X}}_j \boldsymbol{\beta}(k))' + \mathbf{S}^{-1} \right]^{-1}, n_k + \nu \right).$$

Let  $\mathbf{X}_i(k)$  be an  $n_k \times p_i$  matrix with rows comprising only those individuals  $j$  with  $s_j = k$ , where  $k = 1, \dots, N$ . Let  $\mathbf{B}(k) = \text{blockdiag}(\mathbf{X}_1(k), \dots, \mathbf{X}_m(k))$ , and let

$$\mathbf{V}(k) = \begin{bmatrix} \Sigma_{11}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{1j} \tilde{\mathbf{x}}_{1j} & \Sigma_{12}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{1j} \tilde{\mathbf{x}}_{2j} & \cdots & \Sigma_{1m}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{1j} \tilde{\mathbf{x}}_{mj} \\ \Sigma_{21}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{2j} \tilde{\mathbf{x}}_{1j} & \Sigma_{22}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{2j} \tilde{\mathbf{x}}_{2j} & \cdots & \Sigma_{2m}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{2j} \tilde{\mathbf{x}}_{mj} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{mj} \tilde{\mathbf{x}}_{1j} & \Sigma_{m2}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{mj} \tilde{\mathbf{x}}_{2j} & \cdots & \Sigma_{mm}^{-1}(k) \sum_{j:s_j=k} \tilde{\mathbf{x}}'_{mj} \tilde{\mathbf{x}}_{mj} \end{bmatrix}.$$

Note that one can show that

$$\mathbf{V}(k) = \mathbf{B}'(k)(\Sigma^{-1}(k) \otimes \mathbf{I}_{n_k})\mathbf{B}(k).$$

Let  $\mathbf{z}_i(k)$  be those elements of  $\mathbf{z}_i$  such that  $s_j = k$  in lexicographical order, and let  $\mathbf{z}(k) = (\mathbf{z}_1(k)', \dots, \mathbf{z}_m(k)')'$ . Then the regression effects are updated independently as

$$\boldsymbol{\beta}(k)|\text{else} \sim N_{p+} \left( [\mathbf{V}(k) + \mathbf{M}]^{-1} \mathbf{B}(k)' (\Sigma^{-1}(k) \otimes \mathbf{I}_{n_k}) \mathbf{z}(k), [\mathbf{V}(k) + \mathbf{M}]^{-1} \right).$$

Finally, the  $\{z_{ij}\}$  are updated

$$z_{ij}|\text{else} \sim N \left\{ \tilde{\mu}_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)} \sum_{l \neq i} \Sigma_{il}^{-1}(s_j) (z_{lj} - \tilde{\mu}_{lj}), \frac{1}{\Sigma_{ii}^{-1}(s_j)} \right\}$$

truncated to  $(\alpha_{i,y_{ij-1}}, \alpha_{i,y_{ij}})$ . Here,  $\tilde{\mu}_{ij} = \tilde{\mathbf{x}}_{ij} \boldsymbol{\beta}_i(s_j)$ . See Section A.2 for derivation details of the nonparametric model.

We also considered placing hyperpriors on  $\mathbf{b}$ ,  $\mathbf{M}$ , and  $\mathbf{S}$ .

$$\mathbf{b} \sim N_{p+}(\mathbf{b}_0, \mathbf{M}_0^{-1}), \quad \mathbf{M} \sim \text{Wish}_{p+}(\mathbf{M}_0, m_0), \quad \mathbf{S}^{-1} \sim \text{Wish}_m(\mathbf{S}_0, \nu_0).$$

This model shrinks the regression effects and covariance matrices toward common targets, borrowing strength across mixture components. The full conditional distributions are

$$\begin{aligned} \mathbf{b}|\{\boldsymbol{\beta}(k)\}, \mathbf{M} &\sim N_{p+} \left( [\mathbf{M}_0 + N\mathbf{M}]^{-1} \left[ \mathbf{M}_0 \mathbf{b}_0 + \mathbf{M} \sum_{k=1}^N \boldsymbol{\beta}(k) \right], \mathbf{M}_0 + N\mathbf{M} \right)^{-1}, \\ \mathbf{M}|\{\boldsymbol{\beta}(k)\}, \mathbf{b} &\sim \text{Wish}_{p+} \left( \left[ \sum_{k=1}^N (\boldsymbol{\beta}(k) - \mathbf{b})(\boldsymbol{\beta}(k) - \mathbf{b})' + \mathbf{M}_0^{-1} \right]^{-1}, N + m_0 \right), \\ \mathbf{S}^{-1}|\{\Sigma(k)\} &\sim \text{Wish}_m \left( \left[ \sum_{k=1}^N \Sigma^{-1}(k) + \mathbf{S}_0^{-1} \right]^{-1}, N\nu + \nu_0 \right). \end{aligned}$$

Analogous to Section 2.2, a reasonable prior is

$$\mathbf{S}_0 = \frac{1}{25} \mathbf{I}_{m \times m}, \quad \nu_0 = m, \quad m_0 = p+, \quad \mathbf{b}_0 = (0.5, 0, \dots, 0)' \in \mathbb{R}^{p+}, \quad \mathbf{M}_0 = \frac{1}{p+} \mathbf{I}_{p+ \times p+}.$$

More details on the nonparametric model with hyperpriors are available in Section A.3.

## 2.4 INTERPRETATION AND MODEL COMPARISON

### Interpretation of parametric model

The direction (positive or negative) and significance of regression effects are directly interpretable in terms of the latent variables, and hence indirectly interpretable in terms of the observed ordinal variables. For example if the posterior mean of  $\beta_{12}$  is 1.4 with posterior standard deviation 0.2, then the corresponding predictor significantly increases the mean response of the first ordinal outcome. In terms of hypothesis testing, this is enough.

Quantifying the effect of covariates on the mean responses takes a bit more thought. If the observed data  $\{(\tilde{\mathbf{X}}_j, \tilde{\mathbf{y}}_j)\}_{j=1}^T$  are a random sample, we can consider averaged effects; for Bernoulli data see Agresti (2013; Secs. 5.4.7 and 5.4.8). Consider one of the  $m$  ordinal outcomes of subject  $j$ ,  $Y_{ij}$  with corresponding predictor  $\tilde{\mathbf{x}}_{ij}$ . Under the parametric model

$$E(Y_{ij} | \tilde{\mathbf{x}}_{ij}) = \sum_{r=1}^{c_i} r \left[ \Phi \left( \frac{\alpha_{ir} - \tilde{\mathbf{x}}_{ij} \boldsymbol{\beta}_i}{\sqrt{\sigma_{ii}}} \right) - \Phi \left( \frac{\alpha_{i,r-1} - \tilde{\mathbf{x}}_{ij} \boldsymbol{\beta}_i}{\sqrt{\sigma_{ii}}} \right) \right], \quad (2.7)$$

where  $\boldsymbol{\Sigma} = [\sigma_{i_1 i_2}]_{m \times m}$ . Thus we are able to get mean responses for any covariate vector. Unfortunately, unlike the regression on  $z_{ij}$ , this is nonlinear in covariates. To obtain easily-digestible inferential statements, we look at average causal effects. For the drinking behavior data there are two 4-level outcomes per individual ( $Y_{1j}, Y_{2j}$ ), drinking frequency and quantity of beer consumption. There are three dichotomous covariates: age (a), gender (g) and abuse (b). Define the average causal effects of

frequency ( $\delta_{bf}$ ) and quantity ( $\delta_{bq}$ ) due to abuse through (2.7) as

$$\delta_{bf} = \frac{1}{T} \sum_{j=1}^T [E(Y_{1j}|b = 1, a_j, g_j) - E(Y_{1j}|b = 0, a_j, g_j)]$$

and

$$\delta_{bq} = \frac{1}{T} \sum_{j=1}^T [E(Y_{2j}|b = 1, a_j, g_j) - E(Y_{2j}|b = 0, a_j, g_j)],$$

respectively. These measure how much more often an individual typically drinks, and how much more an individual typically drinks when they have a history of abuse, averaged over the age and gender categories in the population. The quantities are straightforward to compute from MCMC output, and the idea generalizes to the mixture model (below).

One aspect of inference that is often of interest, dating back to Pearson (1904), is the polychoric correlation. After adjusting for covariates, the polychoric correlation between  $Y_{i_1,j}$  and  $Y_{i_2,j}$  where  $1 \leq i_1, i_2 \leq m$  is simply the correlation of the underlying latent traits  $\tilde{\rho}_{i_1 i_2} = \tilde{\sigma}_{i_1 i_2} / \sqrt{\tilde{\sigma}_{i_1 i_1} \tilde{\sigma}_{i_2 i_2}}$  where  $\tilde{\sigma}_{i_1 i_2} = \text{cov}(\tilde{z}_{i_1}, \tilde{z}_{i_2})$ . For the drinking behavior data, this provides a measure of association between drinking frequency and quantity. Note that under the parametric model this association is static across all covariate vectors  $\tilde{\mathbf{X}}$  and simply equal to  $\tilde{\rho}_{i_1 i_2} = \sigma_{i_1 i_2} / \sqrt{\sigma_{i_1 i_1} \sigma_{i_2 i_2}}$ . Thus the model provides a nonparametric, covariate-adjusted estimate of the polychoric correlation, more properly termed a nonparametric multiple polychoric correlation. Note that Pearson (1904, Section 8) examines a possible way to relax the multivariate normality assumption underlying the estimation of the polychoric correlation. Upon consideration of large numbers of categories in each dimension, Pearson validates the usual Pearson correlation coefficient over the polychoric and finds “*The practical importance of this result would appear to be great, for it frees us...from the need for supposing normal frequency.*” Ekström (2011) also considers a general polychoric correlation as we do here.

## Interpretation of mixture model

Under the nonparametric mixture, the mean response is weighted according to component membership

$$E(Y_{ij}|\tilde{\mathbf{x}}_{ij}) = \sum_{r=1}^{c_i} r \sum_{k=1}^N w_k \left[ \Phi \left( \frac{\alpha_{ir} - \tilde{\mathbf{x}}_{ij} \boldsymbol{\beta}_i(k)}{\sqrt{\sigma_{ii}(k)}} \right) - \Phi \left( \frac{\alpha_{i,r-1} - \tilde{\mathbf{x}}_{ij} \boldsymbol{\beta}_i(k)}{\sqrt{\sigma_{ii}(k)}} \right) \right]. \quad (2.8)$$

One simply uses (2.8) instead of (2.7) in the definitions of the average causal effects  $\delta_{bf}$  and  $\delta_{bq}$  above.

In the parametric model, the correlation between  $z_{i1,j}$  and  $z_{i2,j}$  does not change with covariates. However, in the nonparametric model the correlation does change. For a random  $\tilde{\mathbf{z}}_j$  with design  $\tilde{\mathbf{X}}_j$  (defined in Section 2.2), let  $\boldsymbol{\Sigma} = \text{cov}(\tilde{\mathbf{z}}_j)$  and  $\tilde{\boldsymbol{\mu}}_k = \tilde{\mathbf{X}}_j \boldsymbol{\beta}(k)$  for  $k = 1, \dots, N$ . Iterated covariance provides the covariance and hence polychoric correlation:

$$\boldsymbol{\Sigma} = \sum_{k=1}^N w_k \boldsymbol{\Sigma}(k) + \sum_{k=1}^N w_k \tilde{\boldsymbol{\mu}}_k \tilde{\boldsymbol{\mu}}_k' - \left[ \sum_{k=1}^N w_k \tilde{\boldsymbol{\mu}}_k \right] \left[ \sum_{k=1}^N w_k \tilde{\boldsymbol{\mu}}_k' \right].$$

## Log pseudo marginal likelihood and maximal expected discrepancy

Geisser and Eddy (1979) define the log pseudo marginal likelihood (LPML). The predictive conditional ordinate for the  $j$ th individual is defined to be  $\text{CPO}_j = p(\tilde{\mathbf{y}}_j | \tilde{\mathbf{y}}_{-j})$ , the probability of seeing the response  $\tilde{\mathbf{y}}_j$  through the model given the remaining data  $\tilde{\mathbf{y}}_{-j} = \{\tilde{\mathbf{y}}_{j'} : j' \neq j\}$ . Gelfand and Dey (1994) provide a means for estimating  $\text{CPO}_j$  from MCMC output. Let  $\mathbf{A}_j^\alpha = (\alpha_{1,y_{1j}-1}, \alpha_{1,y_{1j}}) \times \dots \times (\alpha_{m,y_{mj}-1}, \alpha_{m,y_{mj}}) \subset \mathbb{R}^m$  be the region that the latent  $\tilde{\mathbf{z}}_j$  lives in. Then for the parametric model

$$\text{CPO}_j = \left[ \frac{1}{D} \sum_{d=1}^D \frac{1}{\Phi\{\mathbf{A}_j^\alpha | \tilde{\mathbf{X}}_j \boldsymbol{\beta}^{(d)}, \boldsymbol{\Sigma}^{(d)}\}} \right]^{-1}.$$

Under the mixture model

$$\text{CPO}_j = \left[ \frac{1}{D} \sum_{d=1}^D \frac{1}{\sum_{k=1}^N w_k^{(d)} \Phi\{\mathbf{A}_j^\alpha | \tilde{\mathbf{X}}_j \boldsymbol{\beta}^{(d)}(k), \boldsymbol{\Sigma}^{(d)}(k)\}} \right]^{-1}.$$

The LPML is the sum of the log of the conditional predictive ordinates,  $\text{LPML} = \sum_{j=1}^T \log(\text{CPO}_j)$ . Larger LPML implies better fit of the model. The pseudo Bayes factor (PBF) between two models is the exponentiated difference between the LPML values for the two models.

A referee has brought up a potential problem with Gelfand and Dey's (1994) estimate of the CPO statistics, namely that harmonic mean estimators can have very large, or infinite variance. This often happens when the harmonic mean is used to estimate the marginal likelihood of the data, as the difference between the prior and posterior distributions is typically so great that only a few MCMC atoms have non-negligible mass in the importance sampling estimate. However, with CPO computation, the distributions to consider are the full posterior and the posterior computed leaving out one of the data values – these distributions are likely to be very close and have similar behavior in the tails. We double checked our LPML estimates by repeatedly fitting models and found them to be stable within one unit.

In addition to LPML, we also considered the maximal expected discrepancy in probability (MED) to compare the models in simulation study. Let  $F(\tilde{\mathbf{x}})$  be the known probability distribution function on  $\mathcal{X}$  of the covariates and define  $\mathcal{Y} = \{1, \dots, c_1\} \times \dots \times \{1, \dots, c_m\}$  to be the set of all possible multivariate ordinal outcomes  $\tilde{\mathbf{y}}$ . Then

$$\text{MED} = \max_{\mathbf{y} \in \mathcal{Y}} \int_{\mathcal{X}} \left| P(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}} | \tilde{\mathbf{X}} = \tilde{\mathbf{x}}) - \hat{P}(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}} | \tilde{\mathbf{X}} = \tilde{\mathbf{x}}) \right| dF(\tilde{\mathbf{x}})$$

is the maximal expected discrepancy of the true and estimated cell probabilities across the contingency table. Here  $\hat{P}$  denotes estimates under the model being considered.

## 2.5 SIMULATION STUDY

We examined the fit of both parametric and nonparametric models using simulated data. The responses are bivariate ordinal variables ( $m = 2$ ) with four levels in each

dimension ( $c_1 = c_2 = 4$ ). The covariates are generated as independent normals. We also examined the models with binary covariates obtaining similar results.

First, we generated 5 datasets for each sample size  $T = 500, 1000, \text{ and } 2000$  from one distribution  $N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ . Both parametric and nonparametric ( $\alpha = 1, N = 10$ ) models were fit to each data set. Second, we generated data from two distinct clusters,  $(\boldsymbol{\beta}(1), \boldsymbol{\Sigma}(1))$  and  $(\boldsymbol{\beta}(2), \boldsymbol{\Sigma}(2))$ , with 80% from one cluster and 20% from the other. Last, we generated data from a bivariate normal whose correlation is covariate-dependent. In particular,  $x \sim N(0, 2^2)$  and then correlation  $\rho = \frac{1-e^x}{1+e^x}$ . Then, the bivariate latent variables  $\mathbf{z}$  are centered at  $(0.5, 0.5)$  with variance 0.25 and correlation  $\rho$ . Bivariate ordinal responses  $\mathbf{Y}$  were obtained using (2.1) with cutoffs  $(0, 0.5, 1)$ . For this simulation, besides the parametric and nonparametric models, we fit one more model: a model that includes the truth:  $\tilde{\mathbf{z}}_j \sim N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  where  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{R}$  and  $\mathbf{R}$  is a correlation matrix with  $\rho(x) = \frac{1-\exp(\tau_0+\tau_1 x)}{1+\exp(\tau_0+\tau_1 x)}$ . A flat prior was placed on the model parameters are  $(\mu_1, \mu_2, \log \sigma^2, \tau_0, \tau_1)$  and updating proceeded via blockwise adaptive MCMC (Haario et al., 2001). Fully 15000 iterations were generated with the last  $D = 10000$  iterations used for posterior inference. Since all chains mixed very well, thinning was not necessary. LPML and MED values are in Table 2.1. Scatter plots of simulated data from the mixture of normals and normal with covariate-dependent correlation are in Figure 2.1.

When the data come from only one component, i.e. they are truly parametric, the LPML difference between the parametric and nonparametric model ranged between  $-4$  and  $1$ , and was typically about  $-1$  indicating PBFs of  $e^1 \approx 3$  in favor of the nonparametric model. Their MED values are also very close; the average difference is less than 0.001 for the 15 cases. Parameter estimates are in Table 2.2. When the data come from two components, in all cases the LPML values grossly prefer the mixture model with differences over 200. The nonparametric MED is only about one-tenth of the parametric MED. Furthermore, the mixture model places essentially no mass in

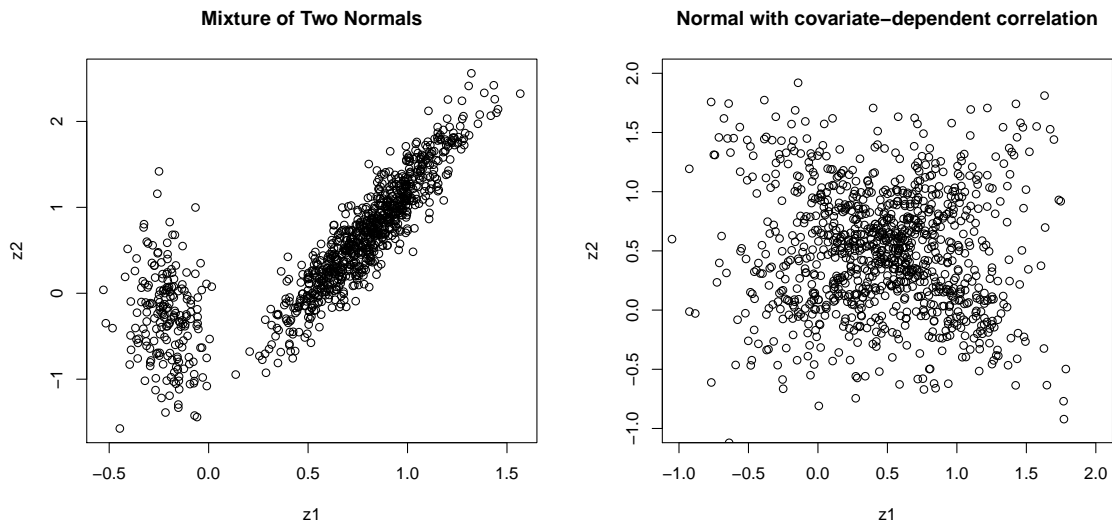


Figure 2.1 Scatter plots of  $T = 1000$  simulated latent  $\tilde{z}$ . Left: mixture of two bivariate normals. Right: bivariate normal with covariate-dependent correlation  $\rho = \frac{1-e^x}{1+e^x}$

the unnecessary eight empty components and estimates the two ‘live’ component precisely; Table 2.3. For the data with bivariate latent normal with covariate-dependent correlation, the nonparametric model fits much better than the parametric model. When sample size  $T = 500$  the LPML difference between nonparametric model and the true model is typically between 20 and 30 while the difference between the parametric model and the true one is over 100. Contour plots of the true and estimated densities for the three models are in Figure 2.2. In the posterior, a mixture of about four normals was required to approximate the true distribution.

## 2.6 DATA ANALYSIS

McMillan, Hanson, Bedrick, and Lapham (2005) consider bivariate ordinal data on drinking behavior. The Lovelace Comprehensive Screening Instrument (LCSI) was given to over 2000 driving-while-intoxicated (DWI) offenders mandated by the court to undergo screening. Among other topics, the LCSI asks offenders questions about



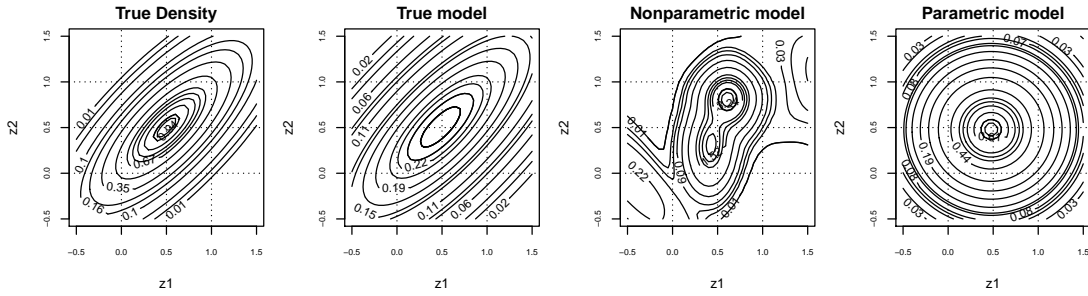


Figure 2.2 Contour plots of true and estimated densities from three models; latent variables simulated from bivariate normal with covariate-dependent correlation ( $E[(Z_1, Z_2)] = (0.5, 0.5)$ ,  $\rho = \frac{1-e^x}{1+e^x}$  where  $x = -2$ ).

psychological issues, drug and alcohol use, and sexual abuse history. The study sample includes  $T = 1964$  offenders who completed the LCSJ and were self-reported beer drinkers. Subjects were asked “*How many times each month do you drink beer?*” and could reply that they drank beer “*up to 1-2 times per month,*” “*a few times per month,*” “*a few times per week,*” or “*almost daily;*” these correspond to  $y_{1j} = 1, 2, 3, 4$  respectively. Respondents also were asked, “*How much beer do you drink?*” and specified the quantity of beer consumed per drinking occasion as “*1,*” “*2-3,*” “*4-5*” or “*6 or more*” beers, corresponding to  $y_{2j} = 1, 2, 3, 4$  respectively. In addition, subjects were asked, “*When you were growing up, were you ever sexually abused or molested by anyone?*” ( $b_j = 0$  for no and  $b_j = 1$  for yes). Because beer is the primary source of alcohol intoxication among DWI offenders, interest is in how history of sexual abuse as a child, along with gender ( $g_j = 0$  for female and  $g_j = 1$  male) and age ( $a_j = 0$  for  $\leq 30$  years old and  $a_j = 1$  for  $> 30$ ), is associated with current beer-drinking patterns. The full data are presented in Table 2.4.

We initially considered the main effects parametric and nonparametric models, with age, gender and abuse for both frequency and quantity; Examining the significance of component effects  $\{\beta(k)\}$  we fitted reduced models with age, gender, and abuse for frequency and gender and abuse for quantity; see Table 2.5 for posterior

means and standard deviations for regression effects and component weights. Following McMillan and Hanson (2005) we also tried adding a gender by abuse interaction in the model but the interactions were not significant in any component and the LPMLs were essentially the same. Among the main effects models the parametric LPML is  $-4138.5$  while the nonparametric LPML is  $-4137.0$  yielding a PBF of about 5 times in favor of the nonparametric model. Among the reduced models, the parametric LPML essentially stays constant at  $-4137.5$  whereas the nonparametric LPML drops significantly to  $-4128.7$ ; now the PBF is over 6000 in favor of the nonparametric model.

Focusing on the nonparametric reduced model with  $N = 10$ , about 55% of the population follows the first component, which has significant effects for age and gender on drinking frequency. In this component going from female to male and/or from “not older than 30-year old” to “older than 30-year old” significantly increases drinking frequency. The second component, comprising about 29% of the population, has no significant covariate effects at all. The third largest component is about 14% of the population. Within this subpopulation, abuse has significant effect on both drinking frequency and quantity. Note that you can compare the raw data in the eight subtables of Table 2.4 directly to the eight panels in Figure 2.4; the data and the model match each other quite well.

Average causal effects are in Table 2.6. For example, under the nonparametric model, going from non-abused to abused increases the mean ordinal frequency response by an estimated 0.22, and increases the mean ordinal quantity by 0.18 averaged over the population.

The polychoric correlation measures the association between drinking frequency and quantity. The posterior mean is  $\hat{\rho} = 0.52$  under the parametric model; frequency and quantity are significantly, moderately correlated. Under the nonparametric model the polychoric correlation changes with covariates. Note that a significant, positive

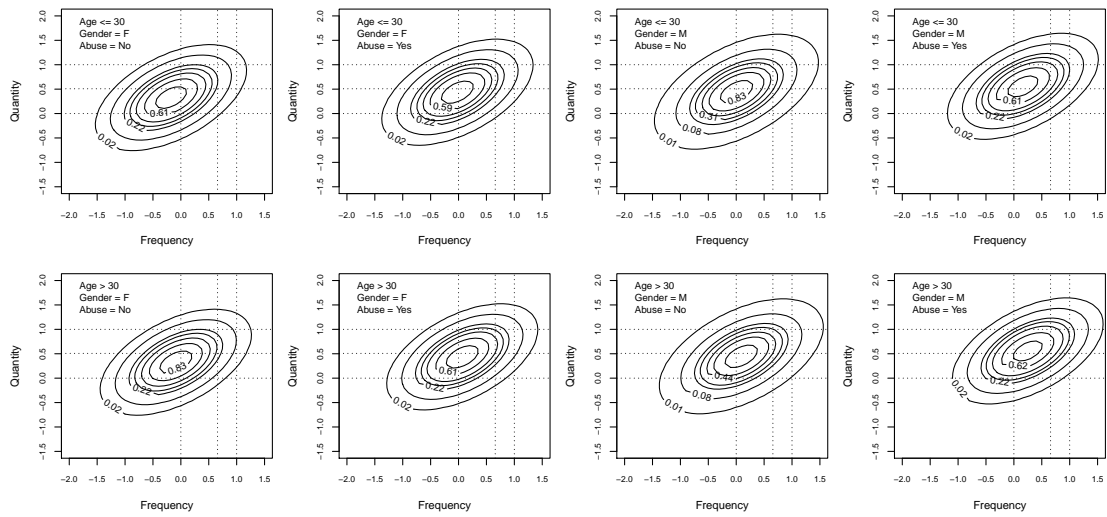


Figure 2.3 Contour plots of estimated densities; parametric reduced model; drinking data.

correlation exists for all DWI offenders.

For the data analyses in this section, Markov chains of length  $D = 100000$  were kept after burn-ins of 100000. All simulations and data analyses were carried out

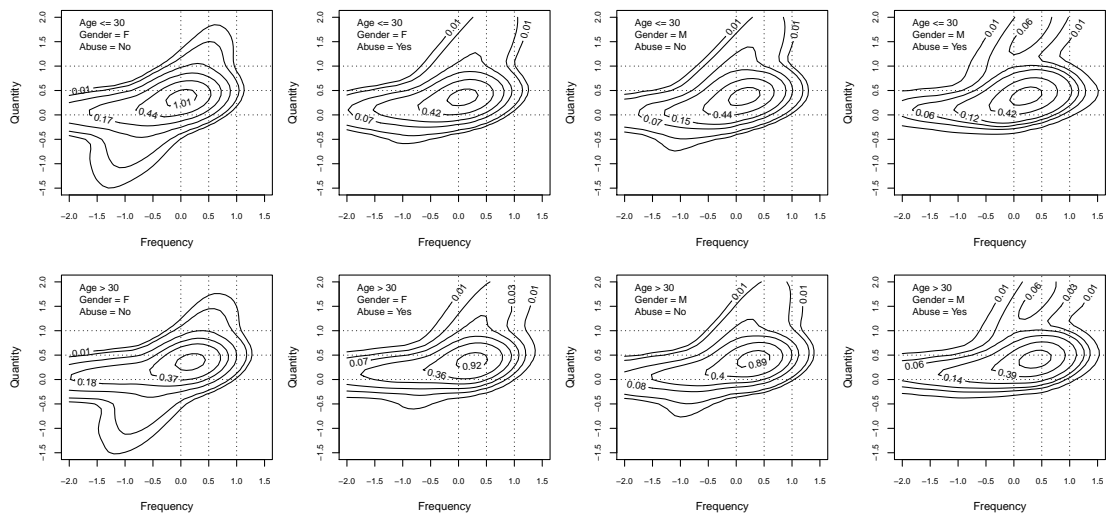


Figure 2.4 Contour plots of estimated densities; nonparametric reduced model, drinking data.

using compiled FORTRAN 90 code; code for fitting the drinking behavior data can be found in Appendix A. Similarly, a careful derivation of all full conditional distributions presented in this chapter are also available in Appendix A.

## 2.7 CONCLUSIONS

A mixture of multivariate probit models was presented along with computational details, useful summary inferential loci, and several illustrations. For the drinking quantity/frequency data, the nonparametric model provided significantly better fit according to LPML than a parametric model and interesting interpretation in terms of polychoric correlations. McMillan and Hanson (2005) found the parametric Dale (1986) model underpredicted frequency and quantity for older males with a history of abuse.

Model interpretation was advanced through consideration of regression effects, latent variable density plots, average causal effects, and polychoric correlations. However, many researchers are used to interpreting contingency table data through odds ratios. A logistic version of the proposed parametric model, marginally interpretable in terms of odds ratios, can be developed via the approach of Holmes and Held (2006). Augment the parametric model by one more layer so that

$$\tilde{\mathbf{e}}_j | \Sigma, \boldsymbol{\psi}_j \sim N_m(\mathbf{0}, \text{diag}(2\boldsymbol{\psi}_j) \Sigma \text{diag}(2\boldsymbol{\psi}_j)),$$

where  $\boldsymbol{\psi}_j = (\psi_{1j}, \dots, \psi_{mj})'$  are *iid* from the Kolmogorov- Smirnov distribution. One can then show that each  $\mathbf{y}_j$  marginally follow a proportional odds model; the polychoric correlation is no longer simple to compute, however. Nor is it clear that the especially simple updating scheme developed in this chapter will remain. A somewhat related approach was taken by O'Brien and Dunson (2004). If this augmentation is applied to the mixture model, then each  $\mathbf{y}_j$  marginally follows a weighted mixture of proportional odds models.

MacEachern’s (1999) general approach allows for component membership probabilities  $\{w_k\}$  to also change with covariates, yielding an infinite hierarchical mixture of experts (HME) model. One successful implementation of such a model is the probit-stick breaking process of Chung and Dunson (2009); a more direct, finite HME model was proposed by Villani, Kohn, and Giordani (2009). The model proposed herein can certainly be modified accordingly, for example by simply taking the stick-breaking pieces to change with individual  $j$  according to  $\log\{V_{jk}/(1 - V_{jk})\} = \boldsymbol{\theta}'_k \tilde{\boldsymbol{x}}_j$  and placing a shrinkage prior on the  $\{\boldsymbol{\theta}_k\}$ . Hanson, Branscum, and Johnson (2014) show how to pick a normal prior on  $\boldsymbol{\theta}_k$  to match  $V_k$  to a particular beta distribution. However, with just the LDDPM model developed here, we never found a case for more than  $N = 10$  components. Furthermore, as seen in the data analysis, the actual underlying joint distribution on the latent trait vector  $\tilde{\boldsymbol{z}}_j$  can change quite flexibly with covariates without the introduction of covariate-dependent weights.

Table 2.1 LPML and MED of parametric (P) and nonparametric (NP) models; simulated data.

Size (T)	Data	Model	Metric	1	2	3	4	5
500	Normal	P	LPML	-1138	-1154	-1157	-1163	-1154
			MED	0.017	0.009	0.020	0.015	0.009
		NP	LPML	-1137	-1152	-1157	-1164	-1153
			MED	0.018	0.010	0.016	0.021	0.009
	Mixture of normals	P	LPML	-737	-790	-787	-767	-767
			MED	0.141	0.153	0.151	0.147	0.143
		NP	LPML	-481	-546	-525	-530	-545
			MED	0.016	0.012	0.016	0.015	0.017
	Normal with covariate dependent correlation	P	LPML	-1299	-1323	-1322	-1313	-1324
			MED	0.136	0.135	0.135	0.135	0.134
		NP	LPML	-1199	-1207	-1227	-1203	-1264
			MED	0.030	0.030	0.029	0.031	0.037
True		LPML	-1177	-1191	-1203	-1177	-1228	
		MED	0.017	0.011	0.011	0.017	0.013	
1000	Normal	P	LPML	-2312	-2306	-2295	-2319	-2338
			MED	0.013	0.008	0.013	0.015	0.017
		NP	LPML	-2311	-2305	-2295	-2320	-2339
			MED	0.007	0.007	0.014	0.015	0.020
	Mixture of normals	P	LPML	-1502	-1531	-1518	-1563	-1570
			MED	0.145	0.150	0.144	0.158	0.151
		NP	LPML	-1037	-1004	-1028	-1045	-1031
			MED	0.017	0.001	0.007	0.009	0.017
	Normal with covariate dependent correlation	P	LPML	-2657	-2632	-2661	-2649	-2644
			MED	0.135	0.135	0.135	0.134	0.135
		NP	LPML	-2481	-2414	-2492	-2407	-2450
			MED	0.024	0.027	0.021	0.020	0.017
True		LPML	-2458	-2396	-2457	-2384	-2418	
		MED	0.095	0.011	0.010	0.009	0.009	
2000	Normal	P	LPML	-4616	-4639	-4605	-4653	-4597
			MED	0.009	0.009	0.010	0.007	0.006
		NP	LPML	-4616	-4635	-4605	-4653	-4596
			MED	0.006	0.007	0.007	0.007	0.006
	Mixture of normals	P	LPML	-3059	-2968	-3036	-3013	-3038
			MED	0.146	0.153	0.156	0.155	0.157
		NP	LPML	-2037	-2104	-2119	-2026	-2082
			MED	0.036	0.036	0.034	0.036	0.035
	Normal with covariate dependent correlation	P	LPML	-5277	-5223	-5240	-5307	-5251
			MED	0.134	0.134	0.135	0.134	0.135
		NP	LPML	-4819	-4785	-4817	-4882	-4803
			MED	0.018	0.020	0.014	0.018	0.021
True		LPML	-4776	-4748	-4778	-4833	-4779	
		MED	0.009	0.010	0.011	0.010	0.010	

Table 2.2 Posterior means and standard deviations from parametric and nonparametric models (only non-empty components); simulated bivariate normal data,  $T = 2000$ .

Parameter	True	Parametric	Nonparametric			
			$k = 1$	$k = 2$	$k=3$	
$w_k$	NA	NA	0.989 (0.014)	0.009 (0.012)	0.002 (0.003)	
$\rho$	-0.5	-0.513 (0.021)	-0.514 (0.022)	-0.050 (0.639)	-0.070(0.684)	
$Y_1$	$\beta_{10}$	0.5	0.492 (0.014)	0.500 (0.014)	0.530 (0.715)	0.473 (0.902)
	$\beta_{11}$	0.2	0.202 (0.012)	0.202 (0.008)	0.301 (0.751)	0.163 (0.886)
$Y_2$	$\beta_{20}$	0.3	0.298 (0.012)	0.300 (0.013)	0.431 (1.145)	0.238 (1.016)
	$\beta_{21}$	-0.1	-0.100 (0.006)	-0.100 (0.006)	-0.142 (0.761)	0.054 (0.937)

Table 2.3 Posterior means and standard deviations from parametric and nonparametric models (only non-empty components); simulated mixture of bivariate normal data,  $T = 2000$ .

Parameter	True		Parametric	Nonparametric	
	$k = 1$	$k = 2$		$k = 1$	$k = 2$
$w_k$	0.8	0.2	NA	0.803 (0.003)	0.197 (0.003)
$\rho$	0.5	-0.5	0.966 (0.004)	0.446 (0.076)	0.026 (0.254)
$Y_1$	$\beta_{10}$	0.8	0.511 (0.012)	0.798 (0.004)	-0.303 (0.051)
	$\beta_{11}$	0.2	0.185 (0.012)	0.201 (0.004)	-0.293 (0.049)
$Y_2$	$\beta_{20}$	0.7	0.541 (0.012)	0.696 (0.005)	-0.291 (0.020)
	$\beta_{21}$	0.6	0.602 (0.014)	0.611 (0.008)	0.486 (0.017)

Table 2.4 DWI offender alcohol quantity-frequency data by age, gender, and history of physical/sexual abuse.

Frequency		Gender							
		Females				Males			
		Beers/occasion				Beers/occasion			
		1	2-3	4-5	6+	1	2-3	4-5	6+
Age $\leq 30$									
No abuse	Up to 1-2 times / month	20	53	4	0	65	171	42	7
	A few times / month	1	30	12	1	12	130	81	19
	A few times / week	0	4	2	1	1	32	27	14
	Almost daily	0	0	0	0	0	2	2	4
Abuse	Up to 1-2 times / month	4	19	3	0	12	43	7	8
	A few times / month	0	13	7	0	4	49	25	9
	A few times / week	0	0	1	5	1	11	12	7
	Almost daily	0	0	0	0	0	3	2	1
Age $> 30$									
No abuse	Up to 1-2 times / month	24	33	3	0	62	159	27	6
	A few times / month	2	22	11	2	5	106	59	16
	A few times / week	0	8	4	1	3	50	40	8
	Almost daily	0	0	0	0	0	6	1	9
Abuse	Up to 1-2 times / month	4	25	2	1	15	53	9	1
	A few times / month	0	11	6	2	2	56	22	10
	A few times / week	0	4	3	2	1	28	21	12
	Almost daily	0	0	2	0	0	4	4	14

Table 2.5 Posterior means and standard deviations from reduced parametric and nonparametric models (only non-empty components are included); drinking data.

Parameter	Parametric	Nonparametric				
		$k = 1$	$k = 2$	$k = 3$	$k = 4$	
$w_k$	NA	0.55 (0.07)	0.29 (0.07)	0.14 (0.03)	0.02 (0.02)	
$\rho$	0.52 (0.02)	0.24 (0.08)	0.30 (0.35)	0.83 (0.16)	0.13 (0.63)	
Frequency	Intercept	-0.18 (0.03)	0.05 (0.07)	-0.97 (0.51)	-0.38 (0.34)	0.10 (0.82)
	Age	0.09 (0.02)	0.14 (0.04)	-0.52 (0.64)	0.18 (0.20)	0.15 (0.98)
	Gender	0.18 (0.03)	0.15 (0.05)	0.26 (0.64)	0.27 (0.25)	-0.03 (0.85)
	Abuse	0.16 (0.03)	0.07 (0.05)	0.10 (0.54)	0.48 (0.17)	-0.10 (0.89)
Quantity	Intercept	0.33 (0.02)	0.36 (0.03)	0.09 (0.07)	0.07 (0.35)	0.71 (0.99)
	Gender	0.12 (0.03)	0.05 (0.03)	0.02 (0.05)	0.83 (0.35)	0.31 (0.96)
	Abuse	0.11 (0.02)	0.02 (0.02)	0.07 (0.05)	1.06 (0.45)	0.11 (0.91)



Table 2.6 Posterior means and standard deviations of average causal effects from reduced parametric and nonparametric models; drinking data.

		Parametric	Nonparametric
Frequency	Age	0.128 (0.033)	0.141 (0.031)
	Gender	0.252 (0.044)	0.231 (0.042)
	Abuse	0.239 (0.041)	0.218 (0.041)
Quantity	Gender	0.210 (0.045)	0.195 (0.042)
	Abuse	0.200 (0.039)	0.181 (0.040)

Table 2.7 Posterior means and standard deviations of polychoric correlations from reduced parametric and nonparametric models; drinking data.

		Age	Gender	Abuse	$\rho$
Nonparametric	$\leq 30$	Female	No	0.54 (0.10)	
	$\leq 30$	Female	Yes	0.49 (0.10)	
	$\leq 30$	Male	No	0.47 (0.09)	
	$\leq 30$	Male	Yes	0.44 (0.09)	
	$> 30$	Female	No	0.49 (0.10)	
	$> 30$	Female	Yes	0.48 (0.09)	
	$> 30$	Male	No	0.51 (0.08)	
	$> 30$	Male	Yes	0.47 (0.08)	
Parametric				0.52 (0.02)	

## CHAPTER 3

# ASSESSMENT OF DPOAE TEST-RETEST DIFFERENCE CURVES VIA HIERARCHICAL GAUSSIAN PROCESSES<sup>1</sup>

### 3.1 INTRODUCTION

Cisplatin, a common chemotherapeutic agent used to treat a variety of cancers, can cause ototoxicity (inner ear poisoning or hearing loss) at high doses. Cisplatin induced ototoxicity typically affects both ears and is permanent. Ototoxicity is especially hurtful in children, where loss of hearing can affect speech, cognitive, and social development (Rybak et al., 2009). Serial monitoring via audiometric analysis, i.e. hearing tests, is often used to assess the presence and severity of ototoxicity; large changes from baseline hearing levels indicate ototoxicity and may result in reduction of the cisplatin dose or discontinuation of cisplatin. However, children treated with cisplatin can be too young or too ill to complete an acceptable behavioral hearing test. Traditionally, this difficulty is resolved by measuring the auditory brainstem response, which is an electrophysiological assessment of the brain's response to sound. Unfortunately, this measurement technique requires sedating the child. Distortion production otoacoustic emissions (DPOAE) testing is a promising, non-invasive alternative to behavioral hearing tests or sedation coupled with auditory brainstem response for evaluating hearing loss in pediatric cancer patients.

Children that are treated with cisplatin are seen for baseline testing before the first

---

<sup>1</sup>The content in this chapter is a reprint for “Junshu Bao, Timothy Hanson, Garnett McMillan, and Kristin Knight (2015). Assessment of DPOAE Test-Retest Difference Curves via Hierarchical Gaussian Processes. *Biometrics*. Revised and resubmitted”

cisplatin course and then right before each subsequent cisplatin chemotherapy treatment, usually the same day they are admitted for cisplatin chemotherapy. They have an end-of-therapy evaluation approximately 4-6 weeks after the last cisplatin course. After the end-of-treatment evaluation, re-evaluation is generally recommended at least every 6 months. The schedule for cisplatin chemotherapy varies according to the type of cancer and stage of disease, but on average children receive cisplatin chemotherapy every 3-6 weeks; hearing is also tested every 3-6 weeks following the chemotherapy schedule. Audiology evaluations for a few standard cancers follow. For standard risk medulloblastoma, audiology evaluations occur every 6 weeks, right before each cisplatin cycle; for osteosarcoma evaluations are administered every 6 weeks; for hepatoblastoma audiology evaluations are every 3 weeks; for germ cell tumor, hearing is tested every 4 weeks.

In terms of the measured response, sound enters the ear as a pressure wave, which is transmitted through the middle ear into the fluid-filled cochlea. In a healthy cochlea tiny hair cells vibrate in response to specific frequencies in the pressure wave spectrum and through transduction deliver that information to the brain. The vibrations generated by the outer hair cells are transmitted back through the middle ear into the ear canal as an “otoacoustic emission,” or OAE. In general, OAEs with higher amplitudes signify a healthier cochlea (i.e. more abundant and more responsive hair cells) than weaker emissions (i.e fewer, damaged hair cells). Noise exposure, aging, certain diseases, and ototoxic agents such as cisplatin will damage the cochlear outer hair cells and weaken or eliminate the OAE.

One can elicit an OAE by sealing a small speaker and a small microphone in the ear canal and playing a tone through the speaker. Pairs of tones (denoted the primary frequency ‘f2’ and secondary frequency) generate a ‘distortion product’ OAE, or DPOAE, measured in decibels of sound pressure level (dB SPL). Various clinical and research protocols exist for measuring DPOAEs, though the most common is to

play several tones at successively increasing frequencies and measuring the DPOAEs at each primary frequency. This generates a ‘DP-gram’ that an audiologist can use to evaluate the health of the cochlea. DP-grams measured over the course of a serial monitoring protocol show how the cochlea is changing, perhaps in response to treatment with ototoxic medications. Theoretically, each human has a smooth DP-gram as a function of frequency at any given time and for a given ear; however these curves do change from time to time and from left to right ear (the structure and viability of each inner ear can be quite different within a person).

Pediatric cisplatin patients are serially monitored over a one year period after the start of treatment with cisplatin. The clinical use of DPOAE testing is to assess significant changes from baseline, i.e. before the start of cisplatin therapy, as chemotherapy treatment progresses over the year. To optimize clinical practicability of DPOAE testing, the main goal of this study is to identify normal shift standards against which a particular pediatric patient’s otoacoustic emissions can be compared. That is, a (simultaneous) reference region containing routine or “normal” differences in the emissions curve from baseline is desired for different test-retest time window lengths of up to one year.

There are about half a dozen clinical DPOAE systems in widespread use. Most only measure DPOAEs at traditional audiometric test frequencies of 1, 2, 3, 4, 6, and 8 kHz, while some allow the clinician to also test in  $1/3^{rd}$  octave steps (9 total f2 primaries) or in  $1/6^{th}$  octave steps (18 total f2 primaries). Note that an octave is a doubling or halving of a frequency, so two frequencies  $f_1 < f_2$  that are one third octave apart satisfy  $\frac{1}{3} = \log_2\left(\frac{f_2}{f_1}\right)$ , i.e.  $2^{1/3} = f_2/f_1$ . One system used in this study measures up to 10 kHz in  $1/4^{th}$  octave steps, while DPOAE systems used in research can theoretically test any f2 primary. It is clear that test-retest reference standards must be suitable for any frequency list, thus any proposed modeling approach needs to handle the computation of credible bands for any and all frequencies.

DP-grams for an 18-month old male child treated with cisplatin are shown in Figure 3.1. The vertical axis shows the DPOAE amplitude, and the horizontal axis shows the  $f_2$  primary frequency in kHz. As treatment progresses over time, the patient shows considerable weakening of the DPOAE as a consequence of receiving cisplatin, ultimately resulting in the long-dashed line at 13 months after the start of treatment with cisplatin. The primary goal of pediatric ototoxicity monitoring is to determine if, during the course of treatment, changes in the cochlea are sufficiently alarming to indicate ototoxic damage and communicatively significant hearing loss, and possible need for the Pediatric Oncologist and the child's family to consider modifying treatment. Pediatric ototoxicity monitoring is also necessary to implement hearing, communication and educational services when needed.

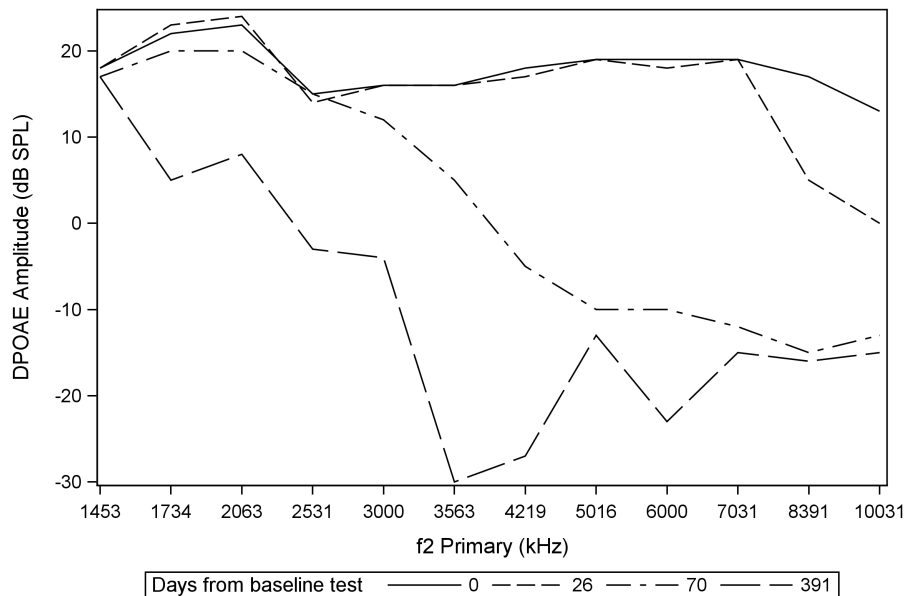


Figure 3.1 Example DP-grams for 18-month old boy treated with cisplatin at baseline and about 4, 10, and 56 weeks later.

To evaluate test-retest standards in a healthy, pediatric population we recruited 38 children age 10 years or younger from the Oregon Health and Science University Doernbecher Children's Hospital between February 2006 and July 2009. To be included

in the study subjects had to have normal hearing sensitivity, defined as pure tone thresholds 20 decibels of hearing loss (dB HL) or better between .5 and 8 kHz; have measurable DPOAEs, no history of ototoxic treatment, ear pathology, ear surgery, or tympanostomy tubes. Test sessions on an ear were excluded if there was conductive hearing loss or abnormal tympanometry or the inability to measure DPOAEs due to excessive subject noise or non-cooperation, e.g. the child was crying or pulled out the headphones. DPOAEs were measured at twelve f2 primary frequencies of 1453, 1734, 2063, 2531, 3000, 3563, 4219, 5016, 6000, 7031, 8391, 10031 Hz using a loudness ratio between the two tones of  $L2/L1 = 65/55$  dB SPL and  $f2/f1$  ratio of 1.22. For example at 1453 Hz, two tones are broadcast,  $f2=1453$  Hz at loudness 65 dB SPL and  $f1=1191$  Hz at loudness 55 dB SPL.

Generally speaking, there are three major components in growth curve modeling: a regression function for the overall trend, individual parameters modeling the differences among subjects or curves (e.g. random effects), and the variance-covariance structure for the repeated within-subject observations. Different approaches emphasize different components, but they have gradually evolved in a more flexible and nonparametric direction. The most traditional approach probably is the MANOVA (multivariate analysis of variance) model. MANOVA-type models have a long and rich history in growth curve analysis, starting with Rao (1958); Eliston and Grizzle (1962); and Pothoff and Roy (1964), and eventually leading to general mixed models (e.g. Fitzmaurice, Laird, and Ware, 2011). Our proposed Gaussian process model generalizes the mixed model framework of Laird and Ware (1982). A drawback of MANOVA is that it only models mean changes in growth curves on a fixed grid of points. To improve MANOVA, various random-coefficient methods (e.g. random-effect ANOVA) were proposed, one big branch of which is latent growth curve modeling within the structural equation modeling framework (McArdle, 1988). The latent growth modeling is comprehensive yet flexible, but it requires relatively larger

samples in addition to equispaced observations, which limit its application on more complex data.

A school of nonparametric approaches for developmental processes or curves is structural analysis (Kneip and Gasser, 1992), which later on was extended to a dynamic time warping method (Wang and Gasser, 1997). However, warping methods are more appropriate if a common structural pattern is observed among the curves. In addition, the goal of warping usually is to find an average curve or to compare two or more curves instead of doing prediction.

Wei and He (2006) proposed a semiparametric quantile regression model for growth charts. In this model the quantiles of the response are the sum of a nonparametric function of measurement time, an autoregressive function of the previous responses (whose coefficients are linear functions of measurement time distances), and a linear function of other covariates. Without any distributional assumptions, this model can identify unusual growth patterns conditioning on the previous observations of one subject. One limitation of this method is the requirement of a large sample size.

As a nonparametric approach to modeling growth curves, Gaussian process (GP) regression models are gaining more and more popularity. Gaussian process modeling was first proposed by O'Hagan (1978), motivated by an optimal design problem in a decision-making framework. However, it was until the early 1990's that GP's were recognized as a powerful tool in regression, classification, and other fields. This is likely due to the requirement of both sophisticated programming and high-speed computers in order to implement efficient Bayesian modeling. Since the late 1990's, GP modeling has been widely used in many fields including Bayesian neural networks (Neal, 1996), spatial and/or temporal modeling (Ripley, 1991), geostatistics (Diggle, Tawn and Moyeed, 1998), and longitudinal data analysis (Diggle and Verbyla, 1998). Good reviews could be found in Williams (1998) and MacKay (1998). Barry (1995)

appears to be the first to use GP's in a growth curve setting.

More recently, Shi, Murray-Smith, and Titterton (2005) proposed a Gaussian process mixture model for regression. By allowing different parameters in the GP model for different clusters, the model naturally handles heterogeneity among data. For each cluster, the dimension of the covariance matrix which needs to be inverted while implementing the model is much smaller than that of all the data, so that the computational burden is very much lightened. Shi et al. (2007) extended the GP regression model in yet another direction to a Gaussian process functional regression model, which uses a functional regression model to model the mean structure and a Gaussian process model to model the covariance structure simultaneously. In related work, Banerjee and Johnson (2006) proposed methodology for spatially varying growth curve modeling. In particular, the growth curve is modeled by a linear function of time, but the intercept and the slope are modeled as a bivariate spatial Gaussian process. Müller and Yang (2010) considered GP models for sparsely observed families of functions. Yi, Shi, and Choi (2011) proposed penalized (e.g. LASSO) GP regression and classification models for high-dimensional nonlinear data.

This chapter proposes a hierarchical GP model, which handles two sources (time and ear) of correlation in the DPOAE measurements, wherein both subject-specific random effects and variance components governing the smoothness and variability of each child's Gaussian process are coupled together. The resulting model for the intermittently observed outcomes is a Laird and Ware (1982) mixed model, but with subject-specific Gaussian process variance components correlated with subject-specific random effects. This correlation is real, seen in preliminary analyses, and requires a bit of thought and care in terms of obtaining posterior inference. The model is further generalized to accommodate subject-specific covariates that affect the variability and smoothness of each subject's GP surface.

The rest of the article is organized as follows. Section 3.2 consists of a preliminary



analysis on the DPOAE data and the hierarchical Gaussian process model. Section 3.3 shows how we obtain reference chart using a volume tube method. Section 3.4 discusses how to use an alternative method, posterior contour probability, to identify abnormal DP-grams. The hierarchical GP model is generalized to include covariates, for the DPOAE data the subject’s age and gender, in Section 3.5. An analysis of the DPOAE data is presented in Section 3.6; Section 3.7 concludes the article.

## 3.2 HIERARCHICAL GAUSSIAN PROCESS MODEL

### **Preliminary analyses**

Proceeding to the modeling of the DPOAE data, Table 3.1 shows that 38 healthy subjects provided 75 ears of data over a total of 196 measurement sessions. Subjects ranged in age from 1 month to 120 months (10 years), with a mean of 38.8 months, or about 3.2 years. Subjects were followed for varying lengths of time, ranging from less than one month to a little over one year. Table B.1 in Appendix B shows the number of valid DPOAE level measurements taken at each f2 Primary frequency. Across frequencies, well over 300 measurements were provided. Figure B.1 in Appendix B shows the followup times in months (baseline = 0 months) for each sample subject listed on the y-axis. Several features stand out from this figure: (i) Two subjects provided no valid baseline data; (ii) Despite some apparent regularity in followup times, there is quite a bit of variation in the number of followups and the followup intervals (see also Table 3.1); (iii) Despite our desire to establish DPOAE level shift standards over a 12-month followup period, most of these data only cover up to about 7 months of followup. While all of these issues are easily handled within the mixed model framework, it is important to recognize that, as a result of (iii), the precision of our DPOAE level shift standards out to 12 months will be relatively small due to the small amount of data at that time horizon.

Figures B.2 to B.5 in Appendix B reveal observed differences in DP-gram patterns among subjects. Particularly, the intercept and slope of linear trends underlying the curves are quite different. This motivates us to include individual intercepts and slopes as random effects in the DP-grams. In addition, the growth curves are non-linear but tend to have an overall common “decreasing-increasing-decreasing” shape, so a penalized B-spline function was used to model this overall nonlinear shape. Theoretically, the curves are smooth, thus a subject-specific infinitely differentiable Gaussian process is assumed on top of the subject-specific mean. Furthermore, there obviously exists strong correlation among the curves of each subject and the left- and right-ear curves are also highly correlated. This motivates the addition of another hierarchy of correlation associated with observation time and ears in the covariance structure. Finally, note that the degree of variability stays remarkably constant across all frequencies when examining DP-grams across the 38 kids; this is further supported by Figure B.10 in Appendix B showing differences in DP-grams from baseline for different followup windows.

Table 3.1 Subject characteristics

	Female	Male	Total
Number of subjects	25	13	38
Number of ears	50	25	75
Age (months)	46.6 (1–120)	23.8 (4–73)	38.8 (1–120)
Total follow-up time (months)	6.7 (3.0–12.2)	6.0 (0.9–9.2)	6.5 (0.9–12.2)
Number of valid study visits	5.1 (3–9)	5.2 (2–8)	5.2 (2–9)

## Hierarchical Gaussian process regression model

Figures B.2 to B.5 in Appendix B shows DPOAE correlation across frequencies, between ears, and across visit times; however the degree of correlation clearly varies from subject to subject with some subjects, e.g., having a large degree of correlation across visits (almost identical curves) and others having considerably more hetero-

geneity. Thus we initially posit a hierarchical Gaussian process model allowing for differing slopes, intercepts, and amounts of correlation across visits and ears for each patient. Later in Section 3.5 we generalize to a structural equation model regressing these latent patient-specific factors onto age and gender. Consider the mixed model for DPOAE response  $y_{ijkl}$

$$y_{ijkl} = \mu(f_j) + b_{i0} + b_{i1}f_j + e_{ijkl}, \quad (3.1)$$

where  $i = 1, \dots, n$  indexes the child;  $j = 1, \dots, F$  indexes the log-f2 levels  $\mathbf{f} = (f_1, \dots, f_F)'$  considered;  $k = 1, \dots, T_i$  indexes the visit times  $\mathbf{t}_i = (t_{i1}, \dots, t_{iT_i})'$  for subject  $i$ ; and  $l = 1, 2$  where  $l = 1$  indicates the left ear and  $l = 2$  indicates the right. Define the total number of observations for subject  $i$  at frequency level  $f_j$  as  $n_{ij} = \sum_{k=1}^{T_i} L_{ijk}$  where  $L_{ijk} = 0$  if neither ear was tested,  $L_{ijk} = 1$  if only one ear was tested, and  $L_{ijk} = 2$  if both ears were tested. The total number of observations for subject  $i$  is  $n_i = \sum_{j=1}^F \sum_{k=1}^{T_i} n_{ijk}$ .

The model builds upon an overall nonlinear population mean  $\mu(f)$  with a subject-specific linear term  $b_{i0} + b_{i1}f$  (e.g., Ghosh and Hanson, 2010); this provides the overall shape of the DP-gram. A mean-zero subject-specific Gaussian process is added  $e_{ijkl}$  that smoothly changes with visit time but also changes with ear. A model that includes correlation across frequency in  $e_{ijkl}$  was also considered (Section 3.6), but was found to provide poorer prediction. This is likely due to the fact that the correlation induced by the random  $b_{i0} + b_{i1}f$  among the DPOAEs at different frequencies is adequate for the data. Let  $\mathbf{y}_{ijk} = (y_{ijk1})$  if only the left ear was tested on child  $i$  at log-f2 level  $f_j$  at time  $t_{ik}$ ,  $\mathbf{y}_{ijk} = (y_{ijk2})$  if only the right ear was tested (both of these imply  $L_{ijk} = 1$ ),  $\mathbf{y}_{ijk} = (y_{ijk1}, y_{ijk2})'$  if both ears were tested ( $L_{ijk} = 2$ ), and  $\mathbf{y}_{ijk} = \emptyset$  if neither ear was tested at frequency  $f_j$  at time  $t_{ik}$ . Then  $\mathbf{y}_{ij} = (\mathbf{y}'_{ij1}, \dots, \mathbf{y}'_{ijT_i})'$  is the set of all measurements taken on child  $i$  at frequency level  $f_j$  (vacuous if the child was never measured at  $f_j$ ) and  $\mathbf{y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iF})'_{n_i \times 1}$  are all of child  $i$ 's measurements over all frequency levels.

## Population mean and linear model

The subject-specific intercepts and slopes are assumed normal:  $\mathbf{b}_i \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}_b \stackrel{iid}{\sim} N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b)$  where  $\mathbf{b}_i = (b_{i0}, b_{i1})'$  and  $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ . The overall population mean as a function of  $f$ , at any time  $t$ , for either ear is the linear portion  $\beta_0 + \beta_1 f$  plus a penalized B-spline  $\mu(f)$

$$\mu(f) = \sum_{s=1}^S \gamma_s \phi_s(f). \quad (3.2)$$

The knots are equispaced over the range of all log-frequencies in the data and  $S = 20$  basis functions used; increasing the number of basis functions did not appreciably change inference. Since  $\mu(f)$  includes constant or linear functions as special cases the mean  $\beta_0 + \beta_1 f + \mu(f)$  is overspecified unless constraints are introduced. Following Gray (1992) we set two of the B-spline coefficients to zero, in our case the first and last,  $\gamma_1 = \gamma_S = 0$ , leaving the column space of the overall mean design matrix unchanged. The parameters for the B-spline are then  $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{S-1})'$ , which is given a 2nd-order random-walk prior

$$p(\boldsymbol{\gamma}) \propto \lambda^{\frac{S-2}{2}} \exp\{-0.5\lambda \|\mathbf{D}\boldsymbol{\gamma}\|^2\}, \quad (3.3)$$

where  $\mathbf{D}$  is a  $(S-4) \times (S-2)$  matrix with  $d_{ii} = d_{i,i+2} = 1$ ,  $d_{i,i+1} = -2$ , and 0 elsewhere (e.g. Kneib, 2006, Section 4.2.2.1). Following Lang and Brezger (2004), the penalization parameter  $\lambda$  follows a gamma distribution,

$$\lambda \sim \Gamma(\alpha_1, \alpha_2), \quad (3.4)$$

with  $\alpha_1 = 1$  and  $\alpha_2 = 0.005$  or  $0.0005$ .

Let  $\mathbf{1}_a$  denote an  $a \times 1$  vector of ones. Let  $\mathbf{X}_{ijk} = \mathbf{1}_{L_{ijk}} \otimes (\phi_2(f_j), \dots, \phi_{S-1}(f_j))$  (vacuous if  $L_{ijk} = 0$ ),  $\mathbf{X}_{ij} = [\mathbf{X}'_{ij1} \cdots \mathbf{X}'_{ijT_i}]'$ , and  $\mathbf{X}_i = [\mathbf{X}'_{i1} \cdots \mathbf{X}'_{iF}]'$ ;  $\otimes$  denotes the Kronecker product. Similarly,  $\mathbf{Z}_{ijk} = \mathbf{1}_{L_{ijk}} \otimes (1, f_j)$ ,  $\mathbf{Z}_{ij} = [\mathbf{Z}'_{ij1} \cdots \mathbf{Z}'_{ijT_i}]'$ , and  $\mathbf{Z}_i = [\mathbf{Z}'_{i1} \cdots \mathbf{Z}'_{iF}]'$ . Then each child's vector of responses at frequency level  $f_j$  follows a linear model

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\gamma} + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{e}_{ij},$$

for  $i = 1, \dots, n$  and  $j = 1, \dots, F$ . However, it is not a typical Laird and Ware (1982) model as the subject-specific variance components governing the mean-zero “error” vector  $\mathbf{e}_{ij}$  are correlated with the subject-specific random effects  $\mathbf{b}_i$  as is described next.

### Child-specific deviation from the population trend

Past studies have shown a significant but very small (less than 1 dB) difference between ears in children (e.g. Keogh et al., 2001). Kemp (2002) notes that “*although OAEs can differ enormously between healthy ears, they are usually quite similar in the left and right ears.*” We do see quite marked differences in the DPOAE data analyzed here, therefore an ear effect needs to be included, as well as visit time.

Analogous to the response and frequency vectors define  $\mathbf{e}_{ijk} = (e_{ijk1})$  or  $\mathbf{e}_{ijk} = (e_{ijk2})$  when  $L_{ijk} = 1$ , or  $\mathbf{e}_{ijk} = (e_{ijk1}, e_{ijk2})'$  when  $L_{ijk} = 2$ ,  $\mathbf{e}_{ij} = (\mathbf{e}'_{ij1}, \dots, \mathbf{e}'_{ijT_i})'$  for child  $i$  at frequency level  $f_j$  and  $\mathbf{e}_i = (\mathbf{e}'_{i1}, \dots, \mathbf{e}'_{iF})'_{n_i \times 1}$ . The Gaussian process model assumes

$$\mathbf{e}_{ij} \stackrel{ind.}{\sim} N_{n_{ij}}(\mathbf{0}, \boldsymbol{\Sigma}_{ij}), \quad (3.5)$$

where  $\boldsymbol{\Sigma}_{ij}$  is the covariance matrix of  $\mathbf{e}_{ij}$  with separable covariance structure

$$\text{cov}(e_{ijkl}, e_{ijk'l'}) = \sigma_i^2 \exp\{-\theta_{ti}|t_{ik} - t_{ik'}|^2 - \theta_{ei}|l - l'|^2\}.$$

Note that if both ears are measured at the same set of frequencies each time the subject-specific covariance model reduces to

$$\mathbf{e}_{ij} \sim N_{n_{ij}}(\mathbf{0}, \sigma_i^2 \boldsymbol{\Sigma}_{ti} \otimes \boldsymbol{\Sigma}_{ei}),$$

where  $\boldsymbol{\Sigma}_{ti}$  and  $\boldsymbol{\Sigma}_{ei}$  are all correlation matrices with simple structure that are functions of  $\theta_{ti}$  and  $\theta_{ei}$  respectively. The subject-specific parameter  $\theta_{ti}$  measures the smoothness of the  $i$ th subject’s responses over time,  $\sigma_i^2$  measures the overall variability of the  $i$ th subject’s DP-gram, and  $\theta_{ei}$  measures how similar responses are in subject  $i$ ’s two ears.

The covariance components governing subject  $i$ 's DP-gram surface (DP-gram function over time and ear) is  $\mathbf{v}_i = (\log(\sigma_i^2), \log(\theta_{ti}), \log(\theta_{ei}))'$ . These are coupled with the subject-specific intercept/slope deviates  $\mathbf{r}_i = (\mathbf{b}'_i, \mathbf{v}'_i)'$  and assumed normal

$$\mathbf{r}_1, \dots, \mathbf{r}_n \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \stackrel{iid}{\sim} N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}, \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}. \quad (3.6)$$

Based on preliminary non-hierarchical individual fits in SAS' `proc mixed`, multivariate normality is reasonable. Further, the normality assumption allows ready calculation of contour probabilities in Section 3.4. Non-normal random effects can also be considered if needed; Jara, Hanson, and Lesaffre (2009) compare several Bayesian nonparametric approaches. At this point, the mixed model (3.1) specified through the penalized B-spline (3.2)–(3.3), random effects/components (3.6) and residual DP-grams (3.5) could be fit via the E-M algorithm yielding the maximum likelihood estimates  $(\hat{\boldsymbol{\mu}}_r, \hat{\boldsymbol{\Sigma}}_r, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\lambda}})$ . Instead, we adopt a Bayesian approach through the priors (3.4) and (3.7, below), primarily to obtain accurate small-sample inferences concerning the simultaneous credible region derived in the next section. The population parameters have the standard conjugate prior

$$\boldsymbol{\mu}_r \sim N_5(\mathbf{m}_0, \mathbf{M}_0), \boldsymbol{\Sigma}_r^{-1} \sim \text{Wish}_5(\mathbf{Q}, q). \quad (3.7)$$

Based on preliminary fits obtained from SAS, we took  $q = 5$ ,  $\mathbf{m}_0$  to be mean of  $\hat{r}_i, i = 1, \dots, n$ ,  $\mathbf{M}_0$  to be the sample covariance matrix of  $\hat{r}_i$  divided by  $n$ , and  $\mathbf{Q}$  to be the inverse of the sample covariance of  $\hat{r}_i$  divided by  $q$ . Quite different hyperparameter settings were used to perform a sensitivity analysis; posterior inference changed negligibly.

## Fitting the model

Markov chain Monte Carlo (MCMC) is by now a standard tool in the statistician's toolbox; a good reference is the book by Robert and Casella (2004). The MCMC

scheme we develop makes use of several closed form full conditional distributions, as well as adaptive Metropolis-Hastings (Haario, Saksman, and Tamminen, 2001 & 2005) strategies for those full conditional distributions that do not have easily-sampled closed forms. The full conditional distributions and our final MCMC scheme are presented in the Appendix B along with sample FORTRAN 90 code that calls the IMSL library (Rogue Wave Software). After an initial burn-in, samples from the augmented posterior  $(\boldsymbol{\mu}_r^m, \boldsymbol{\Sigma}_r^m, \boldsymbol{\gamma}^m, \lambda^m \mid \mathbf{y}_1, \dots, \mathbf{y}_n)$  for  $m = 1, \dots, M$  are kept and used for posterior inference, in particular a simultaneous reference chart for intermittently-observed responses on one ear at any time point for any child in the population. The method prescribed for obtaining this simultaneous reference region is described next.

### 3.3 VOLUME TUBE METHOD FOR OBTAINING REFERENCE CHART

#### One observation time

Let  $\mathbf{y}^* = (y_1^*, \dots, y_{F^*}^*)'$  be a vector of correlated responses from a random child drawn from the population at any time across the  $F^*$  frequencies  $\mathbf{f}^* = (f_1^*, \dots, f_{F^*}^*)'$ , for either ear. Unlike the actual data, these frequencies are on a fine mesh to approximate a smooth DP-gram. Let  $\mathbf{X}_j^* = (\phi_2(f_j^*), \dots, \phi_{S-1}(f_j^*))$  and  $\mathbf{X}^* = [\mathbf{X}_1^{*'} \cdots \mathbf{X}_{F^*}^{*'}]'$ . Let  $\mathbf{Z}_j^* = (1, f_j^*)$  and  $\mathbf{Z}^* = [\mathbf{Z}_1^{*'} \cdots \mathbf{Z}_{F^*}^{*'}]'$ . Let  $\mathbf{r}^* = (b_0^*, b_1^*, \log(\sigma^{2*}), \log(\theta_t^*), \log(\theta_e^*))'$  be the vector of subject-specific effects/components associated with  $\mathbf{y}^*$ ,  $\mathbf{b}^* = (b_0^*, b_1^*)'$ , and let the  $F^* \times F^*$  matrix  $\boldsymbol{\Sigma}^*$  be  $\sigma^{2*} \mathbf{I}_{F^*}$ . Due to the model's hierarchical nature, this random child's response is conveniently sampled given  $(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r, \boldsymbol{\gamma})$  by first sampling the subject-specific variables

$$\mathbf{r}^* \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \sim N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r),$$

followed by sampling the DP-gram

$$\mathbf{y}^* \mid \mathbf{r}^*, \boldsymbol{\gamma} \sim N_{F^*}(\mathbf{X}^* \boldsymbol{\gamma} + \mathbf{Z}^* \mathbf{b}^*, \boldsymbol{\Sigma}^*).$$

This sequence is carried out for all  $M$  iterates, i.e.  $\mathbf{r}^{*m} \sim N_5(\boldsymbol{\mu}_r^m, \boldsymbol{\Sigma}_r^m)$  then  $\mathbf{y}^{*m} \sim N_{F^*}(\mathbf{X}^* \boldsymbol{\gamma}^m + \mathbf{Z}^* \mathbf{b}^{*m}, \boldsymbol{\Sigma}^{*m})$ , yielding a posterior sample of random DP-grams from the population  $\mathbf{y}^{*1}, \dots, \mathbf{y}^{*M}$ .

The volume tube approach of Crainiceanu et al. (2007) and Krivobokova, Kneib, and Claeskens (2010) is modified to obtain a  $(1 - \alpha)100\%$  credible band. First note that, due to linearity, the mean of any  $\mathbf{y}^*$  is simply  $\boldsymbol{\mu}^* = \mathbf{X}^* \bar{\boldsymbol{\gamma}} + \mathbf{Z}^* \bar{\boldsymbol{\beta}}$  where  $\bar{\boldsymbol{\gamma}}$  is the posterior mean of  $\boldsymbol{\gamma}$  and  $\bar{\boldsymbol{\beta}}$  is the posterior mean of  $\boldsymbol{\beta}$ . At each frequency, the usual equal-tailed pointwise  $(1 - \alpha)100\%$  credible interval is formed yielding upper and lower pointwise interval endpoints  $u_1, \dots, u_{F^*}, l_1, \dots, l_{F^*}$ . These are well-approximated by the MCMC empirical estimates  $u_j = y_j^{*[(1-\alpha/2)M]}$  and  $l_j = y_j^{*[(\alpha/2)M]}$  where  $y_j^{*(1)}, \dots, y_j^{*(M)}$  are the order statistics of random DP-gram values at frequency  $f_j^*$  and  $[\cdot]$  denotes the ceiling function. Each pointwise interval  $(l_j, u_j)$  is adjusted by increasing  $c > 1$  to  $(\mu_j^* - c(\mu_j^* - l_j), \mu_j^* + c(u_j - \mu_j^*))$  until exactly  $(1 - \alpha)100\%$  of the  $\mathbf{y}^{*1}, \dots, \mathbf{y}^{*M}$  lie between the two adjusted bands.

The band has the interpretation that any randomly selected child's DP-gram from the population, for either ear and at any time, will completely fall inside the region  $(1 - \alpha)100\%$  of the time. Note that in actuality DPOAEs will be observed at a finite number of f2 frequencies so the band will be slightly conservative in that the probability that a finite number of DPOAEs falling within the simultaneous band will be  $\geq (1 - \alpha)$ . Alternatively, the band will be exact if  $\mathbf{f}^*$  is simply taken to coincide with the actual frequencies a child's response was observed at. In clinical practice each DPOAE measurement system would simply have a distinct volume tube over the f2 frequencies used by the manufacturer.

## Two or more observation times

A 95% reference interval corresponds to the range of DPOAE level shifts that a clinician can reasonably expect to see in a healthy population. Let  $\mathbf{y}_1^* = (y_{11}^*, \dots, y_{1F^*}^*)'$



and  $\mathbf{y}_2^* = (y_{21}^*, \dots, y_{2F^*}^*)'$  be sets of emissions recorded on the same frequencies at times  $t_1$  and  $t_2$ , often baseline and then some months later. The difference at each frequency is given by the  $F^* \times 1$  vector  $\Delta = \begin{bmatrix} \mathbf{I} & -\mathbf{I} \end{bmatrix} (\mathbf{y}_1^*, \mathbf{y}_2^*)'$ . A short calculation reveals that

$$\Delta \sim N_{F^*}(\mathbf{0}, 2(1 - \exp\{-\theta_i^* |t_1 - t_2|^2\})\Sigma^*), \quad (3.8)$$

where  $\Sigma^*$  is as defined in Section 3.3. The volume tube method immediately applies to DPOAE DP-gram shifts over time.

### 3.4 POSTERIOR CONTOUR PROBABILITIES FOR FINDING CHILDREN WITH ABNORMAL DP-GRAMS

The simultaneous credible band provides a very quick check that a child's response is normal. However, it may miss DP-grams that are unusual in ways different than very high or low responses. For example, the  $i$ th child might have a highly oscillatory DP-gram that still falls into the simultaneous credible region. A highly oscillatory DP-gram could occur if  $\sigma_i^2$  is unusually large relative to the rest of the population but  $b_{i0} \approx \beta_0$  and  $b_{i1} \approx \beta_1$ , etc.

A contour probability measures how rare or unusual an observation is in a manner similar to a p-value. For continuous  $\mathbf{y} \sim p(\cdot)$ , the contour probability for seeing an observation more unusual than  $\mathbf{y}_0$  is  $P\{p(\mathbf{y}) < p(\mathbf{y}_0)\}$ , i.e. the probability of finding other random  $\mathbf{y} \sim p(\cdot)$  in areas of even lower probability (Hanson and McMillan, 2012). Following the first part of Section 3.3 for one set of measurements  $\mathbf{y}^*$ , the contour probability for  $\mathbf{y}_0$  is estimated using ergodic averages and standard multivariate theory, e.g. Mardia, Kent, and Bibby (1979), as

$$P\{p(\mathbf{y}^*) < p(\mathbf{y}_0)\} = \frac{1}{M} \sum_{m=1}^M P\{\chi_{F^*}^2 > (\mathbf{y}_0 - \mathbf{X}^* \boldsymbol{\gamma}^m - \mathbf{Z}^* \mathbf{b}^{*m})' [\boldsymbol{\Sigma}^{*m}]^{-1} (\mathbf{y}_0 - \mathbf{X}^* \boldsymbol{\gamma}^m - \mathbf{Z}^* \mathbf{b}^{*m})\}.$$

Following the second part of Section 3.3, the contour probability for the difference

of two DP-grams taken at two different visits on the same ear, say  $\Delta_0$ , is

$$P\{p(\Delta^*) < p(\Delta_0)\} = \frac{1}{M} \sum_{m=1}^M P\{\chi_{F^*}^2 > \Delta'_0 [2(1 - e^{-\theta_t^{*m}|t_2-t_1|^2}) \Sigma^{*m}]^{-1} \Delta'_0\}. \quad (3.9)$$

This is the contour probability of seeing other random children's DP-gram differences in an area of even less probability, i.e. lower density, than  $\Delta_0$  under the model.

### 3.5 AGE-GENDER-SPECIFIC MODEL

Pediatric patients up to age 18 are monitored during treatment with cisplatin. There is a well-known physiological basis for an age effect on OAE amplitude: DPOAE amplitude decreases over the first few years of life as the ear canal gets larger and the nervous system matures. Since DPOAE levels naturally change with cochlear development, it is desirable to have age-appropriate DPOAE level shift standards as necessary. Posterior means  $\hat{r}_i$  from the model presented in Section 3.2 are plotted versus the child's baseline age in Figure B.6 in Appendix B; different plotting symbols are used for boys and girls. Note, e.g., that estimated intercepts decrease and slopes tend to roughly increase with age for both boys and girls. In general, we allow intercepts, slopes, and all four subject-specific variance components governing the smoothness and variability among DP-grams to change smoothly with age and gender, yielding a Gaussian process structural equation model.

Let  $\mathbf{a}_i$  be a  $p \times 1$  vector of baseline covariates associated with child  $i$ ; then (3.6) becomes

$$\mathbf{r}_i \mid \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r \stackrel{ind}{\sim} N_5(\boldsymbol{\mu}_r \mathbf{a}_i, \boldsymbol{\Sigma}_r), \quad \boldsymbol{\mu}_r = \begin{bmatrix} \mathbf{b}' \\ \boldsymbol{\tau}' \end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}, \quad (3.10)$$

where

$$\mathbf{b}' = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1p} \\ \beta_{21} & \cdots & \beta_{2p} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tau}' = \begin{bmatrix} \tau_{11} & \cdots & \tau_{1p} \\ \tau_{21} & \cdots & \tau_{2p} \\ \tau_{31} & \cdots & \tau_{3p} \end{bmatrix}.$$

Similar to the simpler hierarchical model, the posterior distributions of mean parameters  $\mathbf{b}$  have closed forms while those of the covariance parameters do not. The adaptive Metropolis-Hastings algorithm is adopted to generate samples from the latter. See Appendix B for derivation details.

### 3.6 DATA ANALYSIS

A simulation study was used to check the validity of the estimation methods. Data were generated based on the DPOAE data structure and modeling assumptions and the models described in Sections 3.2 and 3.5 fit. From numerous fits to replicated data sets of varying sizes, coefficient estimates are essentially unbiased and consistent. In Section 3.2, general characteristics of the DPOAE data were discussed. Now we apply both the simple hierarchical Gaussian process model followed by the age-gender-specific generalization to the DPOAE data.

We first fit the hierarchical model without considering age and gender. Fully 20,000 MCMC iterates were generated with the last 10,000 iterations used for posterior inference. All chains mixed very well so thinning was not necessary. The log-pseudo marginal likelihood (LPML) of this model (Gelfand and Dey, 1994), a leave-one-out measure of a model’s predictive ability, is -11786.9. During the last 10,000 iterations, a random child’s DP-gram was predicted, consisting of responses corresponding to 31 log(f2 primary) levels. Based on these samples, both the pointwise and simultaneous 95% credible bands were generated for DP-grams of a randomly selected healthy child; see Figure 3.2. In addition, test-retest differences were predicted according to (3.8). Using the method discussed in Section 3.2, pointwise and simultaneous credible bands for test-retest differences are also estimated. Figure 3.3 shows 10 DP-grams of test-retest differences and the 95% credible region when the followup time is one month. The numbers along the curves are posterior contour probabilities calculated by Equation (3.9). Note that the contour probability of one

of ID 6’s DP-grams is only about 0.0206, though it is still within the credible region. The two DP-grams of ID 9 have almost the same mean but their contour probabilities are quite different (0.9953 versus 0.4935). The more oscillatory one has a much smaller contour probability. This shows how important it is to combine the credible region and the contour probability to identify abnormal DP-grams. To illustrate this point in more detail, we designed three “extreme” cases of DP-grams of test-retest differences and calculated the posterior contour probabilities. Note that although they are quite extreme, they are still within the 95% credible bands; see Figures B.7, B.8, and B.9 in Appendix B. Although all within the credible band, these figures illustrate highly unusual DP-grams relative to what was seen in the actual data. In most cases, the contour probabilities are very small, signaling significant changes in DP-grams over time, alerting medical practitioners to possible hearing damage if the current level of dosing is continued.

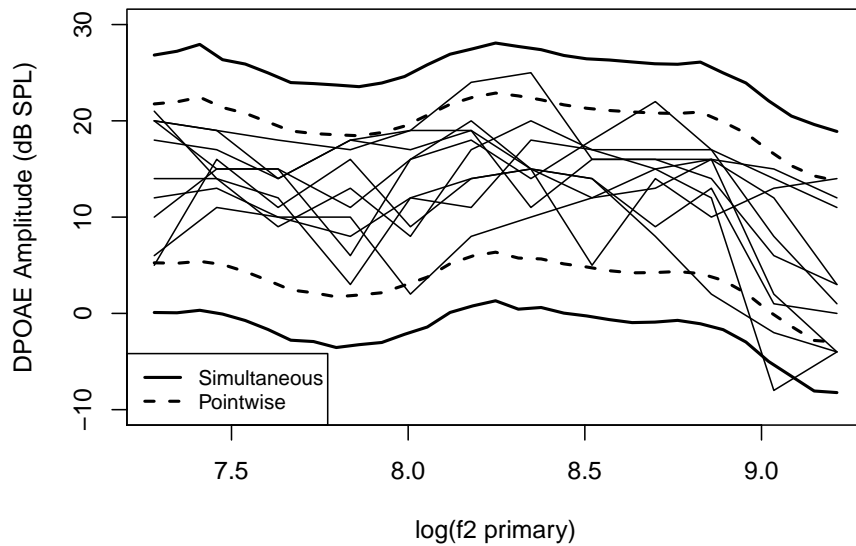


Figure 3.2 95% credible bands and 10 sample DP-grams

The age-gender-specific model was also fit to the DPOAE data. By allowing

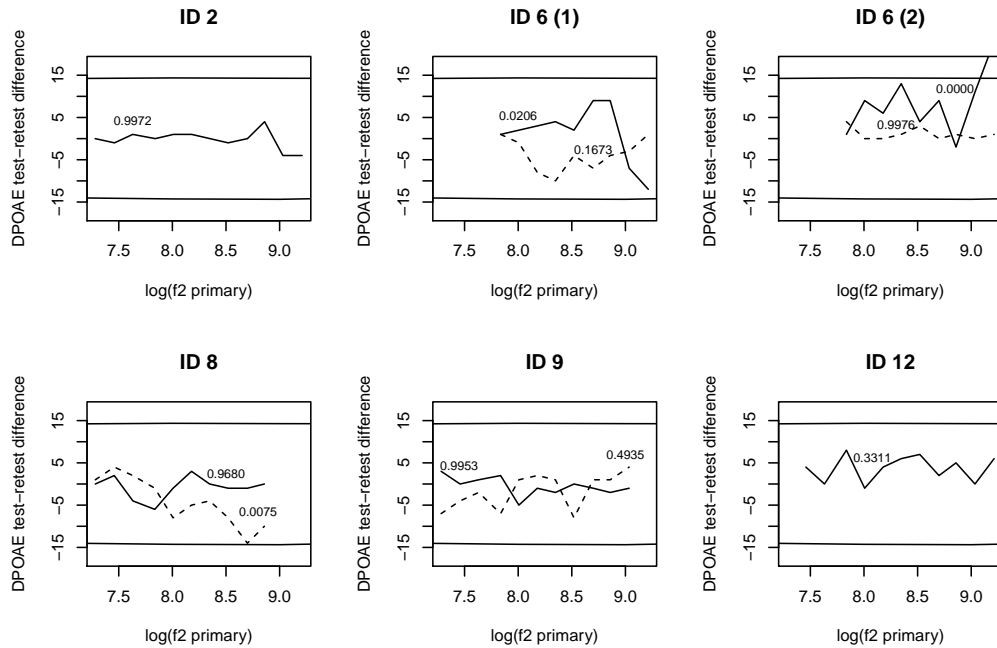


Figure 3.3 10 sample DP-grams of test-retest differences of 5 children and 95% simultaneous credible band; followup time = 1 month.

subject-specific intercept-slope and Gaussian process variance components to be co-variate dependent, the structural equation model may have better predictive power than the hierarchical one, provided that baseline covariate information is available, which of course is almost always the case. As before, 10,000 posterior samples were kept after burn-in. The LPML for the age-gender model -11785.5 is a bit better than the hierarchical model; the pseudo Bayes factor (exponentiated difference in LPML) is about 4 in favor of the age-gender model. Figures 3.4 and 3.5 show that as followup time increases, the credible band tends to be wider; the width of the credible band increases quickly as followup time goes from half a month to two months. After two months, the curve is essentially static, i.e. temporal correlation dies down to almost zero. Also note that as the children get older, the credible band tends to be wider, reflecting more variability in DPOAE response, and boys have wider credible bands than girls at the same age with the same followup time. In a meta-analysis of ten

earlier studies, Reavis et al. (2015) also found that increased variability in DPOAE response as monitoring interval length increases.

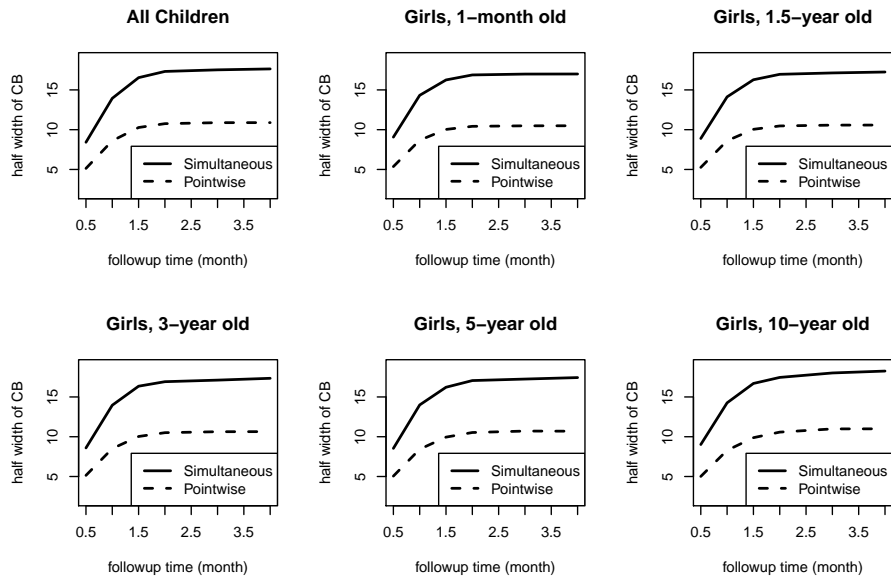


Figure 3.4 Half widths of credible bands of test-retest differences for all children and for girls.

As for the real out-of-sample 18-month-old male cancer patient’s DP-grams seen in Figure 3.1, the posterior contour probabilities at 26, 70, and 391 days after baseline are 0.19, 0.00, and 0.00. Figure 3.6 shows simultaneous bands constructed from the half-widths obtained from Figure 3.5 (upper right) along with the test-retest difference curves for the cancer patient. This subject’s test-retest difference first goes outside the simultaneous bands at 70 days and is almost completely outside the bands at 391 days. For this patient, even the simultaneous bands for DP-grams observed only once (Figure 3.6 lower right), the subject’s bands are “normal” at baseline and 26 days, but become abnormal at 70 and 391 days. All three inferential loci: simultaneous bands for one DP-gram, simultaneous bands for test-retest differences, and contour probabilities imply the same finding: A marked shift the DP-gram is observed at 70 days, signaling an alert to the child’s oncology specialists.

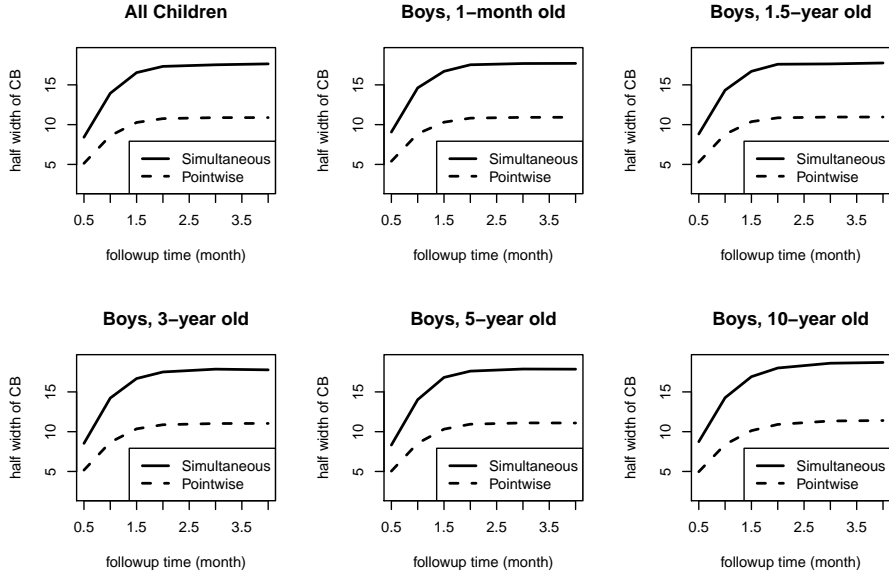


Figure 3.5 Half widths of credible bands of test-retest differences for all children and for boys.

Eventually, it is hoped that doctors will use the half-widths in Figure 3.4 and Figure 3.5 to serially monitor their pediatric cancer patients' hearing, e.g. test-retest differences will be automatically displayed along with the appropriate simultaneous credible region, i.e. the horizontal bands. For example, one 5-year old girl with followup time of one month gives half-width of 14.0, so if a test-retest difference function has a separation greater than 14.0 in magnitude the subject is flagged as having a significantly abnormal hearing change. Alternatively, these figures can be used as is.

Several other models were considered. Originally, we considered a model that also allowed for subject-specific correlation in  $f_2$  as well as time and ear, i.e.

$$\text{cov}(e_{ijkl}, e_{ij'kl}) = \sigma_i^2 \exp(-\theta_{if} |f_j - f_{j'}|^2).$$

The hierarchical and age-gender versions of this model have LPML values of -11846.6 and -11841.0 respectively; this is significantly worse than the models developed thus far without residual correlation in frequency. Thus additional correlation in frequency

simply adds unnecessary noise to the model. This is likely due to two reasons: (i) the random subject-specific  $b_{i0} + b_{i1}f$  on top of the population  $\mu(f)$  induces correlation across frequencies already and so residual correlation was unnecessary, and (ii) the log- primary f2 values are reasonably sparse, enough so that the residual correlation dies down to zero in between them. Simpler models with no residual correlation at all were also considered,  $\Sigma_{ij} = \mathbf{I}_{n_{ij}}\sigma_i^2$  (subject-specific variance only) and  $\Sigma_{ij} = \mathbf{I}_{n_{ij}}\sigma^2$  (usual Laird and Ware model); here  $\mathbf{I}_a$  is an  $a \times a$  identity matrix. These models had LPML -14288.1 and -14723.3 respectively, much smaller than the other four models; residual correlation in both time and ear significantly improve prediction for these data. The models are described in Sections B.4 to B.7.

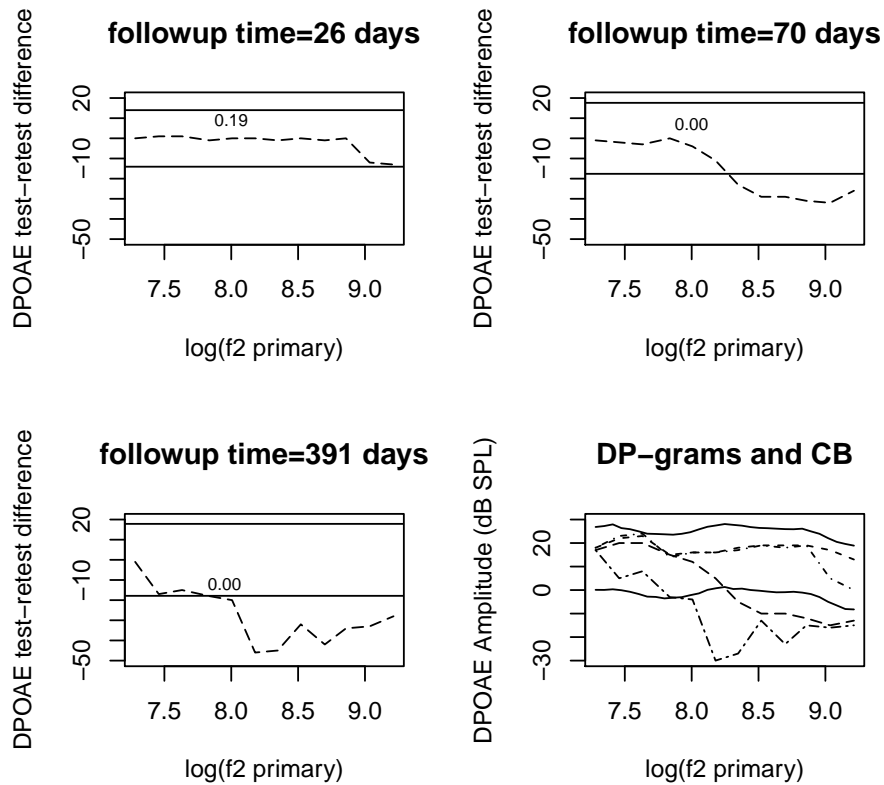


Figure 3.6 Test-retest difference simultaneous bands at 4, 10, and 56 weeks (upper left, upper right, lower left) along with actual difference for 18-month male cancer patient; credible band for one DP-gram with corresponding DP-grams (lower right).



### 3.7 CONCLUDING REMARKS

Hierarchical Gaussian processes are suggested for the modeling of DPOAE in children. DP-grams are theoretically smooth, but only sparsely sampled. Gaussian processes assume little in the way of structure and DP-gram characteristics such as correlation and variability are governed by subject-specific variance components, allowing for a great deal of flexibility from child-to-child in DP-gram shape. Parsimony is achieved through subject-specific slopes and intercepts atop an overall smooth population trend modeled with a penalized B-spline. Subject-specific intercept, slope, and variance components are further regressed onto a child's age and gender, improving both prediction and interpretation.

Fortunately the children's DP-grams tend to closely follow one overall mean for both ears, which we postulated as  $\mu(f) + b_{i0} + b_{i1}f$ . Ear-to-ear and time-to-time variability was ably captured by the Gaussian process surface for each child. Had the DP-grams been highly variable within many of the children, or variable across ears, a more natural model would allow separate intercept/slope combinations at each observation time and/or each ear.

The goal of this research to provide DPOAE system manufacturers accurate simultaneous regions and contour probabilities that are reported directly by the software. The current plan is to increase the sample size to improve the precision with which the bands are estimated and disseminate this information through subject-matter journal publications and presentations. Also, we intend to extend these results to children with cystic fibrosis, who are commonly treated with ototoxic antibiotics.

## CHAPTER 4

### A MEAN-CONSTRAINED FINITE MIXTURE OF NORMALS<sup>1</sup>

#### 4.1 INTRODUCTION

This chapter proposes a mean-constrained finite mixture of multivariate Gaussian densities. The model proposed is constructive, i.e. the mean-constraint is built into the model, versus approaches which impose mean constraints after fitting (e.g. Li, Lin, and Müller, 2010; Jara, Hanson, and Lesaffre, 2009; Yang and Dunson, 2010) or approaches that impose the constraints during nonparametric estimation (e.g. Hall and Presnell, 1999; Eloyan and Ghosh, 2011; Laurence, Pignol, and Tabak, 2014). Mean constraints are necessary in many inferential situations, including generalized linear mixed models (e.g. Jara, Hanson, and Lesaffre, 2009), structural equation models (e.g. Yang and Dunson, 2010), and in the modeling of extreme value distributions (e.g. Boldi and Davison, 2007), to name a few.

Linear mixed models (LMM) are widely applied on the analysis of longitudinal and other types of repeated measures data. An open question in LMM is how to best model the random effects. Classical approaches assume that the random effects follow a mean-zero Gaussian distribution (Laird and Ware, 1982). However, it has been found that this assumption is often violated, affecting prediction for subjects not in the data set (Claeskens and Hart, 2009). To relax this assumption, novel approaches have been proposed to model the random effects more flexibly including the Dirichlet process prior (Kleinman and Ibrahim, 1998), Hermite expansions (Zhang

---

<sup>1</sup>The content in this chapter is a reprint for “Junshu Bao and Timothy Hanson (2015). A mean-constrained finite mixture of normals. *Statistics and Probability Letters*. In revision”.

and Davidian, 2001), penalized Gaussians over a grid (Ghidey, Lesaffre and Eilers, 2004), and mixtures of multivariate Polya trees (Jara, Hanson, and Lesaffre, 2009). For identifiability, ideally a random effects distribution should be centered at zero, enhancing interpretation of fixed effects in terms of population averages. This is a simple constraint for parametric approaches but becomes challenging for semi- and non-parametric approaches. In Section 4.2, we introduce the mean-constrained finite mixture (MCFM) for multivariate density estimation and, for illustration, show how the MCFM can be used in mixed models. In Section 4.3, we evaluate the performance of the proposed MCFM model using a simple simulation study. Section 4.4 describes one application of the model on the Framingham cholesterol data and Section 4.5 concludes the chapter.

## 4.2 MODEL

Consider  $p$ -dimensional data arising from a finite mixture model with  $J$  components. Let  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_J)$ ,  $\boldsymbol{\Sigma} = (\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_J)$ , and  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)'$  be component means, covariance matrices, and weights respectively. Assume for now that  $\boldsymbol{\pi}$  is given. Then

$$\mathbf{y}_1, \dots, \mathbf{y}_n | \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma} \stackrel{iid}{\sim} G = \sum_{j=1}^J \pi_j N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j).$$

The mean-zero constraint

$$E(\mathbf{y}) = \sum_{j=1}^J \pi_j \boldsymbol{\mu}_j = \mathbf{0}_p, \tag{4.1}$$

which can be written as

$$[\boldsymbol{\pi}' \otimes \mathbf{I}_p] \text{Vec}(\boldsymbol{\mu}) = \mathbf{0}_p,$$

forces  $\text{Vec}(\boldsymbol{\mu}) = (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_J)'$  to live in a  $(J-1) \times p$ -dimensional hyperplane in  $\mathbb{R}^{J \times p}$ . Let  $\boldsymbol{\theta}_j = \pi_j \boldsymbol{\mu}_j$  and  $\boldsymbol{\theta} = (\pi_1 \boldsymbol{\mu}_1, \dots, \pi_J \boldsymbol{\mu}_J)'$ . The constraint (4.1) is satisfied when

$\mathbf{1}'_J \boldsymbol{\theta} = \mathbf{0}$  where  $\mathbf{1}_J$  is a vector of  $J$  ones. Define

$$\boldsymbol{\theta} = \mathbf{M}_{J \times (J-1)} \mathbf{Z}_{(J-1) \times p} = \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_J \end{bmatrix} \begin{bmatrix} \mathbf{z}'_1 \\ \vdots \\ \mathbf{z}'_{J-1} \end{bmatrix}, \quad (4.2)$$

where

$$\mathbf{z}_1, \dots, \mathbf{z}_{J-1} | \boldsymbol{\Omega} \stackrel{ind.}{\sim} N_p(\mathbf{0}, \boldsymbol{\Omega})$$

and  $\mathbf{m}_j$ 's are  $1 \times (J-1)$  vectors. Note that from Equation (4.2), we have  $\boldsymbol{\theta}_j = (\mathbf{m}_j \mathbf{Z})'$ . The columns of matrix  $\mathbf{M}$ , of dimension  $J \times (J-1)$ , span the space orthogonal to the vector of all ones  $\mathbf{1}_J$ ,  $\mathcal{C}(\mathbf{1}_J)$ . That is,  $\mathcal{C}(\mathbf{M}) = \mathcal{C}(\mathbf{1}_J)^\perp$ . As a concrete example,  $\mathbf{M}$  could have  $m_{jj} = 1$ ,  $m_{j+1,j} = -1$  for  $j = 1, \dots, J$ , and zeros elsewhere. Note then  $\mathbf{1}'_J \boldsymbol{\theta} = \mathbf{0}_{1 \times p}$  a.s., i.e.  $E(\mathbf{y}) = \mathbf{0}$ . Let  $s_i = j$  if  $\mathbf{y}_i$  comes from component  $j$  and define  $\mathbf{s} = (s_1, \dots, s_n)'$ . The data model conditional on the  $\mathbf{s} = (s_1, \dots, s_n)'$  is

$$\mathbf{y}_i | \mathbf{Z}, \boldsymbol{\Sigma}, \mathbf{s} \stackrel{ind.}{\sim} N_p \left( \pi_{s_i}^{-1} (\mathbf{m}_{s_i} \mathbf{Z})', \boldsymbol{\Sigma}_{s_i} \right), \quad P(s_i = j) = \pi_j.$$

The full conditional for  $\mathbf{Z}$  is proportional to

$$p(\mathbf{Z} | \text{else}) \propto \prod_{j=1}^{J-1} \exp \left\{ -0.5 \mathbf{z}'_j \boldsymbol{\Omega}^{-1} \mathbf{z}_j \right\} \prod_{i=1}^n \exp \left\{ -0.5 [\mathbf{y}_i - \pi_{s_i}^{-1} (\mathbf{m}_{s_i} \mathbf{Z})']' \boldsymbol{\Sigma}_{s_i}^{-1} [\mathbf{y}_i - \pi_{s_i}^{-1} (\mathbf{m}_{s_i} \mathbf{Z})'] \right\}.$$

Note that

$$\pi_{s_i}^{-1} \mathbf{Z}' \mathbf{m}'_{s_i} = \pi_{s_i}^{-1} [\mathbf{z}_1 \cdots \mathbf{z}_{J-1}] \mathbf{m}'_{s_i} = \pi_{s_i}^{-1} \sum_{j=1}^{J-1} m_{s_i, j} \mathbf{z}_j.$$

Thus,

$$\mathbf{z}_j | \text{else} \sim N_p \left\{ \mathbf{V}_j \sum_{i=1}^n \frac{m_{s_i, j} \boldsymbol{\Sigma}_{s_i}^{-1}}{\pi_{s_i}} \left( \mathbf{y}_i - \sum_{k \neq j} \frac{m_{s_i, k} \mathbf{z}_k}{\pi_{s_i}} \right), \mathbf{V}_j \right\}, \quad \mathbf{V}_j = \left[ \boldsymbol{\Omega}^{-1} + \sum_{i=1}^n \frac{m_{s_i, j}^2 \boldsymbol{\Sigma}_{s_i}^{-1}}{\pi_{s_i}^2} \right]^{-1}.$$

Under the prior

$$\boldsymbol{\Omega}^{-1} \sim \text{Wish}_p(df_{\boldsymbol{\Omega}}, \mathbf{O}),$$

the full conditional for  $\boldsymbol{\Omega}$  is

$$\boldsymbol{\Omega}^{-1} | \text{else} \sim \text{Wish}_p \left\{ df_{\boldsymbol{\Omega}} + J - 1, \left[ \mathbf{O}^{-1} + \sum_{j=1}^{J-1} \mathbf{z}_j \mathbf{z}'_j \right]^{-1} \right\}.$$

Note that  $\sum_{j=1}^{J-1} \mathbf{z}_j \mathbf{z}_j' = \mathbf{Z}'\mathbf{Z}$ . Assuming the hierarchical model for the precision matrices

$$\Sigma_1^{-1}, \dots, \Sigma_J^{-1} | \mathbf{S}, df_{\Sigma}, \sim \text{Wish}_p(df_{\Sigma}, \mathbf{S}), \quad p(\mathbf{S}^{-1}) \propto |\mathbf{S}^{-1}|^{-\frac{p+1}{2}},$$

we have

$$\mathbf{S}^{-1} | \text{else} \sim \text{Wish}_p \left\{ J df_{\Sigma}, \left[ \sum_{j=1}^J \Sigma_j^{-1} \right]^{-1} \right\},$$

and

$$\Sigma_j^{-1} | \text{else} \sim \text{Wish}_p \left\{ df_{\Sigma} + \sum_{i=1}^n I\{s_i = j\}, \left[ \mathbf{S}^{-1} + \sum_{i:s_i=j} \mathbf{E}_i(j) \mathbf{E}_i'(j) \right]^{-1} \right\}.$$

where  $\mathbf{E}_i(j) = \mathbf{y}_i - \pi_j^{-1}(\mathbf{m}_j \mathbf{Z})'$ . The component indicators  $\mathbf{s} = (s_1, \dots, s_n)'$  are sampled one at a time via

$$P(s_i = j | \text{else}) \propto \pi_j \phi_p(\mathbf{y}_i | \pi_j^{-1}(\mathbf{m}_j \mathbf{Z})', \Sigma_j), \quad j = 1, \dots, J,$$

where  $\phi_p(\cdot | \boldsymbol{\mu}, \Sigma)$  is the usual  $p$ -dimensional normal density. A Gibbs sampler samples from the full conditional distributions of  $\{\mathbf{z}_j\}_{j=1}^{J-1}$ ,  $\{s_i\}_{i=1}^n$ ,  $\{\Sigma_j\}_{j=1}^J$ ,  $\boldsymbol{\Omega}$ , and  $\mathbf{S}$  in turn. Note that  $\boldsymbol{\mu}_j$  is indirectly sampled as  $\pi_j^{-1}(\mathbf{m}_j \mathbf{Z})'$ , for  $j = 1, \dots, J$ .

## Prior on $\boldsymbol{\pi}$

A Dirichlet prior on  $\boldsymbol{\pi}$  results in the so-called ‘label switching’ problem that plagues discrete mixture models. Roughly speaking, allocations of observations to components are arbitrary in that components can be relabeled without the likelihood changing value; this is the very definition of non-identifiability. Many approaches that ensure identifiability order parameters of the components in some manner. The easiest ordering to consider for the multivariate model considered here is to order the probabilities  $\pi_1 > \pi_2 > \dots > \pi_J$ . To this end, we consider an expected stick-breaking prior, which yields geometric probabilities for  $\boldsymbol{\pi}$  with the ordering  $\pi_1 > \pi_2 > \dots > \pi_J$ .

A truncated Dirichlet process (Ishwaran and James, 2001) induces a prior on  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_J)$  as

$$\pi_j = v_j \prod_{k=1}^{j-1} (1 - v_k), \quad v_1, \dots, v_{J-1} | M \stackrel{iid}{\sim} \text{beta}(1, M), \quad v_J = 1.$$

The geometric stick-breaking process (Mena, Ruggiero, and Walker, 2011; Gutiérrez, Gutiérrez-Peña, and Mena, 2014) simply replaces each piece  $v_j$  by its expectation  $E(v_j) = q = (1 + M)^{-1}$ . Thus we replace  $(J - 1)$  free parameters in  $\boldsymbol{\pi}$  with just one parameter  $0 < q < 1$  yielding truncated geometric probabilities

$$\pi_j = \frac{q(1 - q)^{j-1}}{1 - (1 - q)^J}, \quad j = 1, \dots, J.$$

This leads to the full conditional distribution

$$p(q|\text{else}) \propto \frac{q^n (1 - q)^{\sum_{i=1}^n (s_i - 1)}}{[1 - (1 - q)^J]^n} \prod_{i=1}^n \phi_p \left( \mathbf{y}_i \mid \frac{(\mathbf{m}_{s_i} \mathbf{Z})' [1 - (1 - q)^J]}{q(1 - q)^{s_i - 1}}, \boldsymbol{\Sigma}_{s_i} \right).$$

The parameter  $q$  is updated using an adaptive Metropolis-Hastings (AM) step (Haario, Saksman, and Tamminen, 2005).

## Linear mixed model

Assume the Laird and Ware (1982) model for  $n_i$  repeated continuous measures on subject  $i$ , for  $i = 1, \dots, n$ . Let  $\mathbf{X}_i = [\mathbf{x}_1 \cdots \mathbf{x}_{n_i}]'$  be the fixed effects design for subject  $i$  and  $\mathbf{U}_i = [\mathbf{u}_1 \cdots \mathbf{u}_{n_i}]'$  be the random effects design. The model is given by

$$y_{ik} = \mathbf{x}'_{ik} \boldsymbol{\beta} + \mathbf{u}'_{ik} \boldsymbol{\gamma}_i + \epsilon_{ik}; \quad \boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_n | G \stackrel{iid}{\sim} G \text{ indep. } \epsilon_{ik} \stackrel{iid}{\sim} N(0, \sigma^2).$$

Subject  $i$ 's repeated measures are written in terms of matrices  $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{U}_i \boldsymbol{\gamma}_i + \boldsymbol{\epsilon}_i$ . Assume *a priori*  $\sigma^{-2} \sim \Gamma(a, b)$ . Standard calculations reveal the full conditional distributions for the population effects  $\boldsymbol{\beta}$

$$\boldsymbol{\beta} | \text{else} \sim N_q \left\{ \mathbf{V}_\beta \left[ \boldsymbol{\Sigma}_\beta^{-1} \boldsymbol{\mu}_\beta + \sigma^{-2} \sum_{i=1}^n \mathbf{X}'_i (\mathbf{y}_i - \mathbf{U}'_i \boldsymbol{\gamma}_i) \right], \mathbf{V}_\beta \right\},$$

where

$$\mathbf{V}_\beta = \left[ \boldsymbol{\Sigma}_\beta^{-1} + \sigma^{-2} \sum_{i=1}^n \mathbf{X}_i' \mathbf{X}_i \right]^{-1},$$

and the individual random effects  $\boldsymbol{\gamma}_i, i = 1, \dots, n$ ,

$$\boldsymbol{\gamma}_i | \text{else} \sim N_p \left\{ \mathbf{V}_{\boldsymbol{\gamma}_i} \left[ \sigma^{-2} \mathbf{U}_i' (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}) + \boldsymbol{\Sigma}_{s_i}^{-1} \boldsymbol{\pi}_{s_i}^{-1} (\mathbf{m}_{s_i} \mathbf{Z})' \right], \mathbf{V}_{\boldsymbol{\gamma}_i} \right\}$$

where

$$\mathbf{V}_{\boldsymbol{\gamma}_i} = \left[ \sigma^{-2} \mathbf{U}_i' \mathbf{U}_i + \boldsymbol{\Sigma}_{s_i}^{-1} \right]^{-1}.$$

$\mathbf{z}_1, \dots, \mathbf{z}_{J-1}$  are sampled one after another by

$$\mathbf{z}_j | \text{else} \sim N_p \left\{ \mathbf{V}_{\mathbf{z}_j} \left[ \sum_{i=1}^n \frac{m_{s_i, j} \boldsymbol{\Sigma}_{s_i}^{-1}}{\pi_{s_i}} \left( \boldsymbol{\gamma}_i - \sum_{j' \neq j} \frac{m_{s_i, j'} \mathbf{z}_{j'}}{\pi_{s_i}} \right) \right], \mathbf{V}_{\mathbf{z}_j} \right\},$$

where

$$\mathbf{V}_{\mathbf{z}_j} = \left[ \sum_{i=1}^n \frac{m_{s_i, j}^2 \boldsymbol{\Sigma}_{s_i}^{-1}}{\pi_{s_i}^2} + \boldsymbol{\Omega}^{-1} \right]^{-1}.$$

The error precision is updated as

$$\sigma^{-2} | \text{else} \sim \Gamma \left( a + 0.5n, b + 0.5 \sum_{i=1}^n \sum_{k=1}^{n_i} (y_{ik} - \mathbf{x}'_{ik} \boldsymbol{\beta} - \mathbf{u}'_{ik} \boldsymbol{\gamma}_i)^2 \right).$$

Finally, the full conditional of  $q$  is proportional to

$$p(q | \text{else}) \propto \frac{q^n (1-q)^{\sum_{i=1}^n (s_i - 1)}}{[1 - (1-q)^J]^n} \prod_{i=1}^n \phi_p \left( \boldsymbol{\gamma}_i \mid \frac{(\mathbf{m}_{s_i} \mathbf{Z})' [1 - (1-q)^J]}{q(1-q)^{s_i - 1}}, \boldsymbol{\Sigma}_{s_i} \right).$$

### 4.3 A SIMULATION STUDY

A simulation study to evaluate the performance of the proposed MCFM LMM is considered here. The random effects follow a mixture of two bivariate Gaussian distributions. We used two settings of component weights: (0.7, 0.3) and (0.5, 0.5), and two sample sizes:  $n = 50$  and  $n = 200$ . For weights=(0.7, 0.3), the random effects,  $\boldsymbol{\gamma}_i, i = 1, \dots, n$ , follow:

$$\boldsymbol{\gamma}_i \sim 0.7 \times N \left\{ \left[ \begin{array}{c} 0.3 \\ -0.3 \end{array} \right], 0.1^2 \left[ \begin{array}{cc} 1.0 & -0.5 \\ -0.5 & 1.0 \end{array} \right] \right\} + 0.3 \times N \left\{ \left[ \begin{array}{c} -0.7 \\ 0.7 \end{array} \right], 0.1^2 \left[ \begin{array}{cc} 1.0 & 0.5 \\ 0.5 & 1.0 \end{array} \right] \right\}$$

For weights= (0.5, 0.5),

$$\gamma_i \sim 0.5 \times N \left\{ \begin{bmatrix} 0.5 \\ -0.5 \end{bmatrix}, 0.1^2 \begin{bmatrix} 1.0 & -0.5 \\ -0.5 & 1.0 \end{bmatrix} \right\} + 0.5 \times N \left\{ \begin{bmatrix} -0.5 \\ 0.5 \end{bmatrix}, 0.1^2 \begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix} \right\}$$

The covariate for both random and fixed effects is time,  $t_{ik} = [(k - 1) - 2.5]/5$ ,  $i = 1, \dots, n$  and  $k = 1, \dots, 5$ . The fixed effects intercept and slope are  $(\beta_0, \beta_1) = (-2.5, 1.8)$ . The standard deviation of the error is  $\sigma = 0.01$ . Given the random effects, the response for each individual  $i$  at time  $k$  is

$$y_{ik} | \beta, \gamma_i, \sigma^2 \sim N(\beta_0 + \beta_1 t_{ik} + \gamma_{i0} + \gamma_{i1} t_{ik}, \sigma^2)$$

We initially used  $J = 5$  components because there are only two clusters in the simulated data. Table 4.1 and Table 4.2 show the true and estimated weights, fixed effects and random effects. Although the random effects distribution is assigned  $J = 5$  components, there are still 70% (51% + 13% + 6%) with mean close to (0.3, -0.3) while 30% (27% + 3%) with mean close to (-0.7, 0.7). The fixed effects are also estimated precisely. Similar results are found in Table 4.2. Figure 4.1 presents the true and estimated distributions of the random intercept and slope for the two weight settings and two sample sizes.

Table 4.1 Posterior means and standard errors (in brackets) from MCFM LMM; simulated data,  $n = 200$  subjects, weights=(0.7, 0.3)

Parameter	True		Estimates				
	$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
weight	0.70	0.30	0.51(0.03)	0.27(0.02)	0.13(0.02)	0.06(0.02)	0.03(0.02)
$\mu_{j1}$	0.30	-0.70	0.28(0.02)	-0.71(0.02)	0.36(0.06)	0.28(0.10)	-0.83(0.48)
$\mu_{j2}$	-0.30	0.70	-0.28(0.02)	0.70(0.02)	-0.35(0.06)	-0.29(0.09)	0.82(0.47)
$\rho_j$	-0.50	0.50	-0.40(0.13)	0.46(0.13)	-0.55(0.33)	-0.37(0.48)	0.08(0.64)
$\beta_0$	-2.50		-2.50(0.01)				
$\beta_1$	1.80		1.80(0.01)				



Table 4.2 Posterior means and standard errors (in brackets) from the MCFM LMM; simulated data,  $n = 200$  subjects, weights=(0.5, 0.5)

Parameter	True		Estimates				
	$j = 1$	$j = 2$	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
weight	0.50	0.50	0.32(0.03)	0.25(0.03)	0.18(0.03)	0.14(0.02)	0.11(0.02)
$\mu_{j1}$	0.50	-0.50	-0.50(0.02)	0.51(0.04)	-0.52(0.03)	0.48(0.08)	0.52(0.10)
$\mu_{j2}$	-0.50	0.50	0.48(0.02)	-0.49(0.03)	0.50(0.03)	-0.46(0.08)	-0.50(0.09)
$\rho_j$	-0.50	0.50	0.52(0.15)	-0.12(0.35)	0.48(0.23)	-0.21(0.51)	0.14(0.60)
$\beta_0$	-2.50		-2.51(0.01)				
$\beta_1$	1.80		1.80(0.01)				

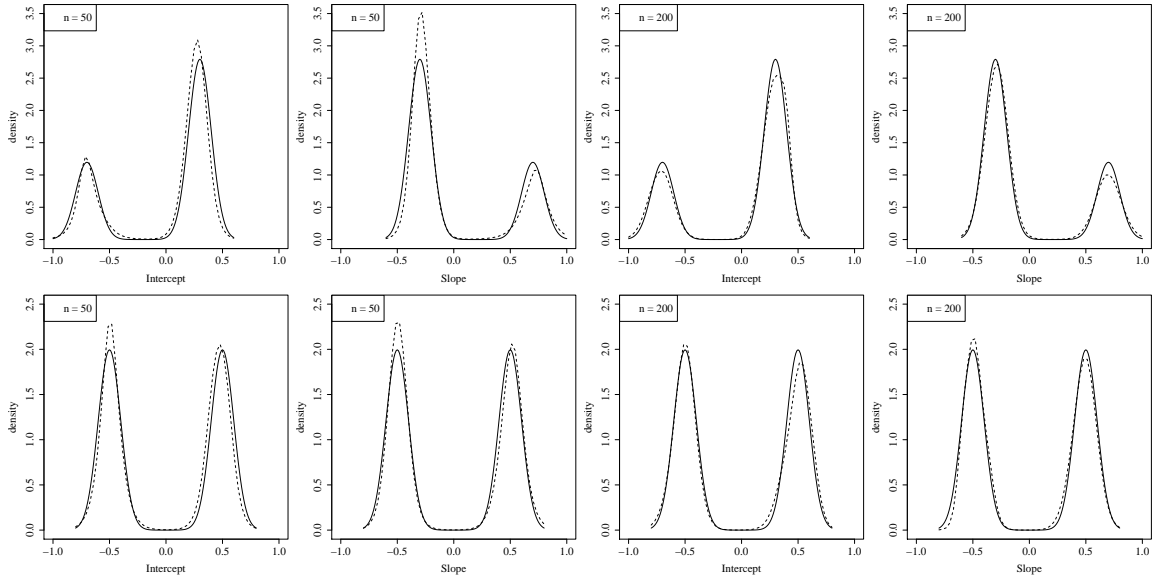


Figure 4.1 MCFM LMM estimates of random intercepts and slopes, with sample size  $n = 50$  and  $200$ , weights=(0.7, 0.3) (top four panels) and (0.5, 0.5) (bottom four panels). Estimated distributions (dashed line) are superimposed to the true distributions (solid line).

#### 4.4 DATA ANALYSIS

We employ the longitudinal cholesterol data from the Framingham heart study to illustrate our approach. The data set consists of a random sample of 200 subjects from the study. For each subject, cholesterol levels were measured every other year for ten years. Control factors such as gender and baseline age were also included. The data are not balanced, with a total of 1044 observations. With this data set,

Zhang and Davidian (2001), Ghidry, Lesaffre, and Eilers (2004), and Jara, Hanson, and Lesaffre (2009) fitted linear mixed models following different assumptions on the random effects as follows:

$$y_{ik} = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{sex}_i + \beta_3 t_{ik} + \gamma_{i0} + \gamma_{i1} t_{ik} + \epsilon_{ik},$$

where  $i = 1, \dots, 200$ ,  $k = 1, \dots, n_i$ , and  $\epsilon_{ik} \sim N(0, \sigma^2)$ . Initially, we fitted the MCFM LMM to the cholesterol data with 5 components. The prior means and the starting values of the parameters are set to be the estimates from the Gaussian linear mixed model (R package: lme4). The results are shown in Table 4.3. The estimates from the MCFM LMM and those of the classical Gaussian LMM are very close. We also fitted a 10-component model but found that the posterior inference is almost the same.

Table 4.3 Posterior means and standard errors (in brackets) by the Gaussian LMM and MCFM LMM; Framingham cholesterol data.

Parameter	Models	
	Gaussian LMM	MCFM LMM
$\beta_0$ (intercept)	1.5969(0.1503)	1.5582(0.0880)
$\beta_1$ (age)	0.0184(0.0035)	0.0186(0.0021)
$\beta_2$ (sex)	-0.0630(0.0554)	-0.0618(0.0379)
$\beta_3$ (time)	0.2817(0.0241)	0.2831(0.0170)
$\sigma$	0.2084(0.0057)	0.1947(0.0049)

Figure 4.2 is a visual demonstration of the model fitting results: the first panel is the random effect distribution surface; the second panel is the contour plot with the crosses being the posterior mean random effects of the 200 subjects; the last two plots show the marginal densities of the random intercept and slope. Note that they are both centered at zero, which indicates that our constraint works.

Being able to provide similar posterior inference, the proposed MCFM model is more concise compared with other semi- or non-parametric models. If the dimension of random effects is greater than two, it will be more difficult to use the penalized

Gaussians model because the number of Gaussian basis functions grows exponentially with dimension (Ghidey, Lesaffre, and Eilers, 2004). The proposed MCFM model is able to model higher dimensional data without much extra effort. The LPML (Geisser and Eddy, 1979) of the MCFM model is 6.00, which is slightly higher than the LPML of the classical Dirichlet process mixtures model (4.22) and the mixture of Dirichlet process model (5.59) but lower than the three mixtures of multivariate Polya trees models (8.39, 11.49, and 11.54) in Jara, Hanson, and Lasaffre (2009). Thus, the proposed model has competitive predictive power with similar non-parametric approaches.

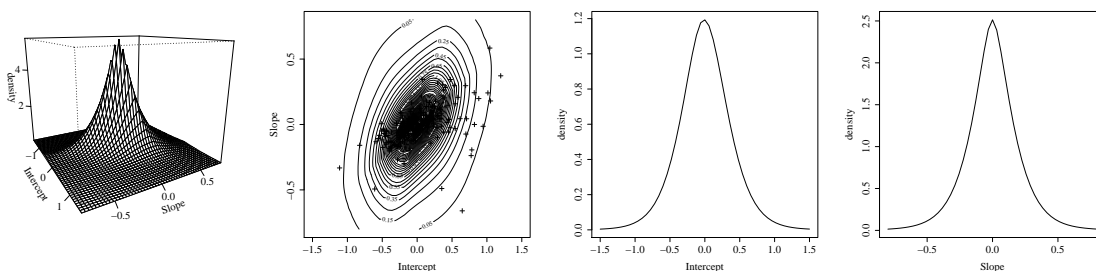


Figure 4.2 Estimated random effects distribution from the MCFM LMM fitted to the cholesterol data.

## 4.5 CONCLUSIONS

The LMM is readily generalized to generalized linear mixed models via the approach of Gamerman (1997); for example, Komárek and Lesaffre (2008) generalized the LMM of Ghidey, Lesaffre, and Eilers (2004) via this approach; also see Jara, Hanson, and Lesaffre (2009).

In principle, a variance constraint can also be imposed. Iterated expectation gives  $\text{Cov}(\mathbf{y}) = \sum_{j=1}^J \pi_j (\boldsymbol{\Sigma}_j + \boldsymbol{\mu}_j \boldsymbol{\mu}_j')$ . Setting this equal to a known  $\boldsymbol{\Sigma}_0$  implies several additional constraints on both  $\{\boldsymbol{\mu}_j\}_{j=1}^J$  and  $\{\boldsymbol{\Sigma}_j\}_{j=1}^J$ . Another extension is to allow the shape of  $G$  to change with predictors. This can be accomplished through a linear

dependent Dirichlet process (e.g. De Iorio et al., 2009). R code to fit the MCFM LMM is available from the first author.

## BIBLIOGRAPHY

- [1] Agresti, A. (2013). *Categorical Data Analysis*. John Wiley & Sons: New Jersey.
- [2] Banerjee, S., and Johnson, G. A. (2006). Coregionalized Single and Multiresolution Spatially Varying Growth Curve Modeling with Application to Weed Growth. *Biometrics*, 62(3), 864–876.
- [3] Barrientos, A., Jara, A., and Quintana, F. A. (2012). On the Support of MacEachern’s Dependent Dirichlet Process and Extensions. *Bayesian Analysis*, 7, 277–310.
- [4] Barry, D. (1995). A Bayesian Model for Growth Curve Analysis. *Biometrics*, 51, 639–655.
- [5] Biswas, A. and Das, K. (2002). A Bayesian Analysis of Bivariate Ordinal Data: Wisconsin Epidemiologic Study of Diabetic Retinopathy Revisited. *Statistics in Medicine*, 21, 549–559.
- [6] Blackwell, D. and MacQueen, J. B. (1973). Ferguson Distributions via Polya Urn Schemes. *Annals of Statistics*, 1, 353–355.
- [7] Boldi, M-O., and Davison, A. C. (2007). A Mixture Model for Multivariate Extremes. *Journal of the Royal Statistical Society: Series B*, 69(2), 217–229.
- [8] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer-Verlag: New York.
- [9] Chib, S. and Greenberg, E. (1998). Analysis of Multivariate Probit Models. *Biometrika*, 85, 347–361.
- [10] Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2010). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. CRC Press, Boca Raton.
- [11] Chu, W. and Ghahramani, Z. (2005). Gaussian Processes for Ordinal Regression. *Journal of Machine Learning Research*, 6, 1019–1041.

- [12] Chung, Y. and Dunson, D. B. (2009). Nonparametric Bayes Conditional Distribution Modeling With Variable Selection. *Journal of the American Statistical Association*, 104, 1646–1660.
- [13] Claeskens, G. and Hart, J. D. (2009). *Goodness-of-fit Tests in Mixed Models*. Test, 18, 213–239.
- [14] Crainiceanu, C. M., Ruppert, D., Carroll, R. J., Joshi, A., and Goodner, B. (2007). Spatially Adaptive Bayesian Penalized Splines with Heteroscedastic Errors. *Journal of Computational and Graphical Statistics*, 16(2), 265–288.
- [15] Dale, J. (1986). Global Cross-ratio Models for Bivariate, Discrete, Ordered Responses. *Biometrics*, 42, 909–917.
- [16] de Carvalho, V., Jara, A., Hanson, T., and de Carvalho, M. (2013). Bayesian Nonparametric ROC Regression Modeling. *Bayesian Analysis*, 8, 623–646.
- [17] De Iorio, M., Johnson, W. O., Müller, P., and Rosner, G.L. (2009). Bayesian Nonparametric Nonproportional Hazards Survival Modeling. *Biometrics*, 65, 762–771.
- [18] De Iorio, M., Müller, P., Rosner, G. L., and MacEachern, S. N. (2004). An ANOVA Model for Dependent Random Measures. *Journal of the American Statistical Association*, 99 (465), 205–215.
- [19] De Yoreo, M. and Kottas, A. (2014). Bayesian Nonparametric Modeling for Multivariate Ordinal Regression. arXiv:1408.1027.
- [20] Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based Geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3), 299–350.
- [21] Diggle, P. J., and Verbyla, A. P. (1998). Nonparametric Estimation of Covariance Structure in Longitudinal Data. *Biometrics*, 401–415.
- [22] Dunson, D. B. and Herring, A. H. (2006). Semiparametric Bayesian Latent Trajectory Models. Technical report, ISDS Discussion Paper 16, Duke University.
- [23] Ekström, J. (2011). A Generalized Definition of the Polychoric Correlation Coefficient. Technical Report, Department of Statistics, UCLA. UCLA: Department of Statistics, UCLA.

- [24] Eloyan, A. and Ghosh, S. K. (2011). Smooth Density Estimation with Moment Constraints using Mixture Distributions. *Journal of Nonparametric Statistics*, 23(2), 513–531.
- [25] Escobar, M. D. and West, M., (1995). Bayesian Density Estimation and Inference using Mixtures. *Journal of the American Statistical Association*, 90(430), pp.577-588.
- [26] Fuentes-García, R., Mena, R. H. and Walker, S. G. (2009). A Nonparametric Dependent Process for Bayesian Regression. *Statistics and Probability Letters*, 79, 1112–1119.
- [27] Gamerman, D. (1997). Sampling from the Posterior Distribution in Generalized Linear Mixed Models. *Statistics and Computing*, 7(1), 57–68.
- [28] Geisser, S. (1965). Bayesian Estimation in Multivariate Analysis. *The Annals of Mathematical Statistics*, 36, 150–159.
- [29] Geisser, S. and Eddy, W. (1979). A Predictive Approach to Model Selection. *Journal of the American Statistical Association*, 74, 153–160.
- [30] Gelfand, A. E., and Dey, D. K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society, Series B*, 56, 501–514.
- [31] Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005). Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing. *Journal of the American Statistical Association*, 100, pp. 1021-1035.
- [32] Gelman, A., Carlin, J., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian Data Analysis*. Chapman & Hall.
- [33] Geweke, J. (1991). Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints and the Evaluation of Constraint Probabilities. In Keramidas, E. M., editor, *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578, Fairfax Station, VA. Interface Foundation of North America, Inc.
- [34] Ghidry, W., Lesaffre, E., and Eilers, P. (2004). Smooth Random Effects Distribution in a Linear Mixed Model. *Biometrics*, 60(4), 945–953.

- [35] Ghosh, P. and Hanson, T. (2010). A Semiparametric Bayesian Approach to Multivariate Longitudinal Data. *Australian and New Zealand Journal of Statistics*, 52, 275–288.
- [36] Griffin, J. E. (2011). Bayesian Multivariate Density Estimation for Observables and Random Effects. Technical report, School of Mathematics, Statistics and Actuarial Science, University of Kent.
- [37] Griffin, J. E., and Steel, M. F. J (2006). Order-Based Dependent Dirichlet Processes. *Journal of the American Statistical Association*, 101, Issue 473, 179-194.
- [38] Gray, R. J. (1992). Flexible Methods for Analyzing Survival Data Using Splines, with Applications to Breast Cancer Prognosis. *Journal of the American Statistical Association*, 87(420), 942–951.
- [39] Gutiérrez, L., Gutiérrez-Peña, E., and Mena, R. H. (2014). Bayesian Nonparametric Classification for Spectroscopy Data. *Computational Statistics & Data Analysis*, 78, 56–68.
- [40] Haario, H., Saksman, E., and Tamminen, J. (2001). An Adaptive Metropolis Algorithm, *Bernoulli*, 7, 223–242.
- [41] Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise Adaptation for High Dimensional MCMC. *Computational Statistics*, 20(2), 265-273.
- [42] Hall, P. and Presnell, B. (1999). Density Estimation under Constraints. *Journal of Computational and Graphical Statistics*, 8(2), 259–277.
- [43] Hanson, T. (2006). Modeling Censored Lifetime Data Using A Mixture of Gammas Baseline. *Bayesian Analysis*, 1, 575–594.
- [44] Hanson, T., Branscum, A., and Johnson, W. (2014). Informative  $g$ -priors for Logistic Regression. *Bayesian Analysis*, 9, 597–612.
- [45] Hanson, T. and McMillan, G. (2012). Scheffe Style Simultaneous Credible Bands for Regression Surfaces with Application to Ache Honey Gathering. *Journal of Data Science*, 10, 175-193.
- [46] Hanson, T. and Jara, A. (2013). Surviving fully Bayesian Nonparametric Regression Models. *Bayesian Theory and Applications*, pp. 593-615. P. Damien, P. Dellaportas, N. Polson, and D. Stephens, eds. Oxford University Press: Oxford.



- [47] Hanson, T. and Pearson, O. (2007). Fitting MANOVA Models with Missing Continuous or Ordinal Data using Reference Priors. *Communications in Statistics: Simulation and Computation*, 36, 621-630.
- [48] Hastie, T. J., Botha, J. L., and Schnitzler, C. M. (1989). Regression with An Ordered Categorical Response. *Statistics in Medicine*, 8, 785-794.
- [49] Holmes, C. C. and Held, L. (2006). Bayesian Auxiliary Variable Models for Binary and Multinomial Regression. *Bayesian Analysis*, 1, 145-168.
- [50] Haario, H., Saksman, E., and Tamminen, J. (2005). Componentwise Adaptation for High Dimensional MCMC. *Computational Statistics*, 20(2), 265-273.
- [51] Ishwaran, H. and Zarepour, M. (2000). Markov Chain Monte Carlo in Approximate Dirichlet and Beta Two-parameter Process Hierarchical Models. *Biometrika*, 87, 371-390.
- [52] Ishwaran, H. and Zarepour, M. (2002). Exact and Approximate Sum Representations for the Dirichlet Process. *Canadian Journal of Statistics*, 30(2), 269-283.
- [53] Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for Stick-breaking Priors. *Journal of the American Statistical Association*, 96(453), 161-173.
- [54] Jara, A., Hanson, T., and Lesaffre, E. (2009). Robustifying Generalized Linear Mixed Models Using A New Class of Mixtures of Multivariate Polya trees. *Journal of Computational and Graphical Statistics*, 18, 838-860.
- [55] Jara, A., Hanson, T., Quintana, F. A., Müller, P., and Rosner, G. L. (2011). DP-package: Bayesian Semi- and Nonparametric Modeling in R. *Journal of Statistical Software*, 40, 1-30.
- [56] Johnson, V. E. and Albert, J. H. (1999). *Ordinal Data Modeling*. Springer-Verlag: New York.
- [57] Kalli, M., Griffin, J. E., and Walker, S. G. (2011). Slice Sampling Mixture Models. *Statistics in Computing*, 21, 93-105.
- [58] Kemp, D. T. (2002). Otoacoustic Emissions, Their Origin in Cochlear Function, and Use. *British Medical Bulletin*, 63, 223-241.

- [59] Keogh, T., Kei, J., Driscoll, C., and Smyth, V. (2001). Distortion-Product Otoacoustic Emissions in Schoolchildren: Effects of Ear Asymmetry, Handedness, and Gender. *Journal-American Academy of Audiology*, 12(10), 506-513.
- [60] Kleinman, K. P. and Ibrahim, J. G. (1998). A Semiparametric Bayesian Approach to the Random Effects Model. *Biometrics*, 54, 921-938.
- [61] Kneib, T. (2006). *Mixed Model Based Inference in Structured Additive Regression*. Dissertation, Ludwig-Maximilians-Universität München.
- [62] Komárek, A. and Lesaffre, E. (2008). Generalized Linear Mixed Model with a Penalized Gaussian Mixture as a Random-effects Distribution. *Computational Statistics & Data Analysis*, 52(7), 3441-3458.
- [63] Kottas, A., Müller, P., and Quintana, F. (2005). Nonparametric Bayesian Modeling for Multivariate Ordinal Data. *Journal of Computational and Graphical Statistics*, 14, 610-625.
- [64] Krivobokova, T., Kneib, T. and Claeskens, G. (2010). Simultaneous Confidence Bands for Penalized Spline Estimators. *Journal of the American Statistical Association*, 105(490), 852-863.
- [65] Laird, N. M. and Ware, J. H. (1982). Random-effects Models for Longitudinal Data. *Biometrics*, 38, 963-974.
- [66] Lang, S. and Brezger, A. (2004). Bayesian P-splines. *Journal of Computational and Graphical Statistics*, 13(1), 183-212.
- [67] Laurence, P. M., Pignol, R. J., and Tabak, E. G. (2014). Constrained Density Estimation. In Proceedings of the 2011 Wolfgang Pauli Institute conference on energy and commodity trading, pp. 259-284. Springer Verlag.
- [68] Lawrence, E., Bingham, D., Liuc, C., and Naird, V. N. (2008). Bayesian Inference for Multivariate Ordinal Data Using Parameter Expansion, *Technometrics*, 50, 182-191.
- [69] Lesaffre E. and Molenberghs, G. (1991). Multivariate Probit Analysis: A Neglected Procedure in Medical Statistics. *Statistics in Medicine*, 10, 1391-1403.
- [70] Li, L., Lin, X., and Müller, P. (2010). Bayesian Inference in Semiparametric Mixed Models for Longitudinal Data. *Biometrics*, 66(1), 70-78.

- [71] MacEachern, S. N. (1999), Dependent Nonparametric Processes, in *In A.S.A. Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association, pp. 50–55.
- [72] MacKay, D. J. (1998). Introduction to Gaussian Processes. *NATO ASI Series F Computer and Systems Sciences*, 168, 133–166.
- [73] Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic press: New York.
- [74] Mena, R. H., Ruggiero, M. and Walker, S. G. (2011). Geometric Stick-breaking Processes for Continuous-time Bayesian Nonparametric Modeling. *Journal of Statistical Planning and Inference*, 141, 3217–3230.
- [75] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley & Sons: New York.
- [76] McMillan, G., Hanson, T., Bedrick, E., and Lapham, S. (2005). Using the Bivariate Dale Model to Jointly Estimate Predictors of Frequency and Quantity of Alcohol Use. *Journal of Studies on Alcohol*, 65, 643–650.
- [77] McMillan, G. and Hanson, T. (2005). SAS Macro BDM for Fitting the Dale Regression Model to Bivariate Ordinal Response Data. *Journal of Statistical Software*, 14, 1–12.
- [78] Müller, H.-G., and Yang, W. (2010). Dynamic Relations for Sparsely Sampled Gaussian Processes. *Test*, 19, 1–29.
- [79] Müller, P., Erkanli, A., and West, M. (1996). Bayesian Curve Fitting Using Multivariate Normal Mixtures. *Biometrika*, 83, 1, pp. 67-79.
- [80] Neal, R. M. (1996). *Bayesian Learning for Neural Networks*, Springer-Verlag: New York.
- [81] O’Brien, S. M. and Dunson, D. B. (2004). Bayesian Multivariate Logistic Regression. *Biometrics*, 60, 739–746.
- [82] O’Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 40, 1–42.

- [83] Pearson, K. (1904). Mathematical Contributions to the Theory of Evolution XIII: On the Theory of Contingency and Its Relation to Association and Normal Correlation. *Draper's Co. Research Memiors, Biometric Series*, no. 1. (Reprinted in *Karl Pearson's Early Papers*, ed. E.S. Pearson, Cambridge, UK: Cambridge University Press, 1948).
- [84] Reavis, K. M., McMillan, G. P., Dille, M.F., Konrad-Martin, D. (2015). Meta-analysis of Distortion Product Otoacoustic Emission Retest Variability for Serial Monitoring of Cochlear Function in Adults. *Ear and Hearing*, 36, e251-260.
- [85] Ripley, B. D. (1991). *Statistical Inference for Spatial Processes*. Cambridge University press.
- [86] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, Second Edition, Springer Science: New York.
- [87] Rybak, L. P., Mukherjea, D., Jajoo, S., and Ramkumar, V. (2009). Cisplatin Ototoxicity and Protection: Clinical and Experimental Studies. *Tohoku Journal of Experimental Medicine*, 219, 177–186.
- [88] Sethuraman, J. (1994). A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, 4, 639–650.
- [89] Shi, J. Q., Murray-Smith, R., and Titterington, D. M. (2005). Hierarchical Gaussian Process Mixtures for Regression. *Statistics and Computing*, 15(1), 31–41.
- [90] Shi, J. Q., Wang, B., Murray-Smith, R., and Titterington, D. M. (2007). Gaussian Process Functional Regression Modeling for Batch Data. *Biometrics*, 63(3), 714–723.
- [91] Srijith, P. K., Shevade, S., and Sundararajan, S. (2013). Semi-supervised Gaussian Process Ordinal Regression. *Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science Volume 8190*, pp. 144-159. Blockeel, H., Kersting, K., Nijssen, S. and Železný, F. Eds. Springer Berlin Heidelberg.
- [92] Tiao, G. C. and Zellner, A. (1964). On the Bayesian Estimation of Multivariate Regression. *Journal of the Royal Statistical Society, Series B*, 26, 277–285.
- [93] Viele, K. and Tong, B. (2002). Modeling with Mixtures of Linear Regressions. *Statistics and Computing*, 12, 315-330.

- [94] Villani, M., Kohn, R., and Giordani, P. (2009). Regression Density Estimation using Smooth Adaptive Gaussian Mixtures. *Journal of Econometrics*, 153, 155–173.
- [95] Williams, C. K. (1998). Prediction with Gaussian Processes: From Linear Regression to Linear Prediction and Beyond. In *Learning in Graphical Models*, pp. 599–621. Springer Netherlands.
- [96] Yang, M. (2006). Applications of Mixture of Polya Trees to Multivariate Survival and Ordinal Data. Unpublished Ph.D. Thesis. Department of Mathematics and Statistics. University of New Mexico.
- [97] Yang, M. and Dunson, D.B. (2010). Bayesian Semiparametric Structural Equation Models with Latent Variables. *Psychometrika*, 75(4), 675–693.
- [98] Yi, G., Shi, J. Q., and Choi, T. (2011). Penalized Gaussian Process Regression and Classification for High-dimensional Nonlinear Data. *Biometrics*, 67(4), 1285–1294.
- [99] Yule, G. U. (1900). On the Association of Attributes in Statistics. *Philosophical Transactions of the Royal Society of London Series A*, 194, 257–319.
- [100] Zhang, D. and Davidian, M. (2001). Linear Mixed Models with Flexible Distributions of Random Effects for Longitudinal Data. *Biometrics*, 57(3), 795–802.

## APPENDIX A

### SUPPLEMENT TO CHAPTER 2

In the derivation of models, we used notations slightly different from Chapter 2 due to simplicity. We put tilde ( $\tilde{\cdot}$ ) on the letter which represent the vector of subject  $j$  in Chapter 2. For example,  $\tilde{\mathbf{y}}_j$  is the observed ordinal response vector for subject  $j$ . In the derivation, we used  $\mathbf{y}_j$ ,  $\mathbf{z}_j$ , and  $\mathbf{x}_j$  to denote  $\tilde{\mathbf{y}}_j$ ,  $\tilde{\mathbf{z}}_j$ , and  $\tilde{\mathbf{x}}_j$ , respectively.

#### A.1 DERIVATION OF THE PARAMETRIC MODEL

##### Joint Posterior Distribution

$$\begin{aligned}
 & p(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \boldsymbol{\alpha} | \mathbf{Y}) \\
 & \propto p(\mathbf{Y} | \mathbf{Z}, \boldsymbol{\alpha}) p(\mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\Sigma}) p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) p(\boldsymbol{\alpha}) \\
 & \propto \prod_{j=1}^T \underbrace{p(\mathbf{y}_j | \mathbf{z}_j, \boldsymbol{\alpha})}_{(1)} \prod_{j=1}^T \underbrace{p(\mathbf{z}_j | \boldsymbol{\beta}, \boldsymbol{\Sigma})}_{(2)} \underbrace{p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1})}_{(3)} \underbrace{p(\boldsymbol{\alpha})}_{(4)}
 \end{aligned}$$

for  $j = 1, 2, \dots, T$ ;  $i = 1, 2, \dots, m$ .

$j$  is the index for individuals.  $i$  is the index for dimensions.

(1)  $p(\mathbf{y}_j | \mathbf{z}_j, \boldsymbol{\alpha}) = \prod_{i=1}^m I\{\alpha_{i, y_{ij}-1} < z_{ij} < \alpha_{i, y_{ij}}\}$ , where  $I$  stands for the indicator variable.

(2)  $p(\mathbf{z}_j | \boldsymbol{\beta}, \boldsymbol{\Sigma}) = \phi_m(\mathbf{z}_j | \boldsymbol{\beta}, \boldsymbol{\Sigma})$ , where  $\phi_m$  stands for the pdf of  $m$ -dimensional normal.

(3)  $p(\boldsymbol{\beta}, \Sigma | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) = \phi_{p+}(\boldsymbol{\beta} | \mathbf{b}, \mathbf{M}) \text{Wish}^{-1}(\Sigma | \mathbf{S}^{-1}, \nu)$ , where  $p+ = \sum_{i=1}^m p_i$  and  $\text{Wish}^{-1}$  stands for the inverse-Wishart distribution.

(4)  $p(\boldsymbol{\alpha}) \propto \prod_{i=1}^m I\{\alpha_{i,1} < \alpha_{i,2} < \dots < \alpha_{i,c_i-1}\}$ .

Note that  $\alpha_{i,0} = -\infty$  and  $\alpha_{i,c_i} = \infty$ .

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mT} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_T \end{bmatrix}, \text{ where } \mathbf{y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{mj} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{m1} \\ z_{21} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mT} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_T \end{bmatrix}, \text{ where } \mathbf{z}_j = \begin{bmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{mj} \end{bmatrix}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_m \end{bmatrix} \quad \boldsymbol{\alpha}_i = \begin{bmatrix} \alpha_{i,1} \\ \alpha_{i,2} \\ \vdots \\ \alpha_{i,c_i-1} \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_m \end{bmatrix}_{p^+ \times c} \quad \text{where } \boldsymbol{\beta}_i = \begin{bmatrix} \beta_{i,1} \\ \beta_{i,2} \\ \vdots \\ \beta_{i,p_i} \end{bmatrix}_{p_i \times 1}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1m} \\ \Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{m1} & \Sigma_{m2} & \cdots & \Sigma_{mm} \end{bmatrix}_{m \times m}$$



## Full Conditionals

$\Sigma^{-1}$ |else

$$\begin{aligned}
& p(\Sigma^{-1}|\text{else}) \\
& \propto \prod_{j=1}^T p(\mathbf{z}_j|\boldsymbol{\beta}, \Sigma^{-1})p(\Sigma^{-1}|\mathbf{S}, \nu) \\
& \propto \prod_{j=1}^T \phi_m(\mathbf{z}_j|\mathbf{x}_j\boldsymbol{\beta}, \Sigma)\text{Wish}(\Sigma^{-1}|\mathbf{S}, \nu) \\
& \propto \prod_{j=1}^T \frac{1}{\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})'\Sigma^{-1}(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})\right\} \\
& \quad \times \left[\frac{1}{\sqrt{|\Sigma|}}\right]^{\nu+m+1} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{S}^{-1}\Sigma^{-1})\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma|}}\right]^{(T+\nu)+m+1} \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^T(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})'\Sigma^{-1}(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}) + \text{tr}(\mathbf{S}^{-1}\Sigma^{-1})\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma|}}\right]^{(T+\nu)+m+1} \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{j=1}^T(\tilde{\mathbf{z}}_j - \tilde{\mathbf{x}}_j\boldsymbol{\beta})(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})'\Sigma^{-1} + \mathbf{S}^{-1}\Sigma^{-1}\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma|}}\right]^{(T+\nu)+m+1} \exp\left\{-\frac{1}{2}\text{tr}\left[\left(\sum_{j=1}^T(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})' + \mathbf{S}^{-1}\right)\Sigma^{-1}\right]\right\}
\end{aligned}$$

So,

$$\Sigma^{-1}|\text{else} \sim \text{Wish}\left(\left[\sum_{j=1}^T(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta})' + \mathbf{S}^{-1}\right]^{-1}, T + \nu\right)$$

$\beta$ |else

$$\begin{aligned}
& p(\beta|\text{else}) \\
& \propto \prod_{j=1}^T p(\mathbf{z}_j|\beta, \Sigma)p(\beta|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j=1}^T \phi_m(\mathbf{z}_j|\mathbf{x}_j\beta, \Sigma)\phi_{p^+}(\beta|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j=1}^T \frac{1}{\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\beta)' \Sigma^{-1}(\mathbf{z}_j - \mathbf{x}_j\beta)\right\} \exp\left\{-\frac{1}{2}[\beta - \mathbf{b}]' \mathbf{M}^{-1}[\beta - \mathbf{b}]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\sum_{j=1}^T (\beta' \mathbf{x}'_j \Sigma^{-1} \mathbf{x}_j \beta - \beta' \mathbf{x}'_j \Sigma^{-1} \mathbf{z}_j - \mathbf{z}'_j \Sigma^{-1} \mathbf{x}_j \beta)\right.\right. \\
& \quad \left.\left.+ (\beta' \mathbf{M}^{-1} \beta - \beta' \mathbf{M}^{-1} \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \beta)\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta' \underbrace{\left(\sum_{j=1}^T \mathbf{x}'_j \Sigma^{-1} \mathbf{x}_j + \mathbf{M}^{-1}\right)}_{\mathbf{W}} \beta - \underbrace{\left(\sum_{j=1}^T \mathbf{z}'_j \Sigma^{-1} \mathbf{x}_j + \mathbf{b}' \mathbf{M}^{-1}\right)}_{\Delta_1} \beta\right.\right. \\
& \quad \left.\left.+ \beta' \underbrace{\left(\sum_{j=1}^T \mathbf{x}'_j \Sigma^{-1} \mathbf{z}_j + \mathbf{M}^{-1} \mathbf{b}\right)}_{\Delta'_1}\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta' \underbrace{(\mathbf{W} + \mathbf{M}^{-1})}_{\mathbf{D}^{-1}} \beta - \underbrace{(\Delta_1 + \mathbf{b}' \mathbf{M}^{-1})}_{\mathbf{K}} \beta - \beta' (\Delta'_1 + \mathbf{M}^{-1} \mathbf{b})\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta' \mathbf{D}^{-1} - \mathbf{K}) \beta - \beta' \mathbf{K}'\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta - \mathbf{D} \mathbf{K}')' \mathbf{D}^{-1} (\beta - \mathbf{D} \mathbf{K}')\right]\right\}
\end{aligned}$$

Thus,

$$\beta|\text{else} \sim N_{p^+}(\mathbf{D} \mathbf{K}', \mathbf{D})$$

where

$$\mathbf{D} = [\mathbf{W} + \mathbf{M}^{-1}]^{-1}$$

$$\mathbf{K} = \Delta_1 + \mathbf{b}' \mathbf{M}^{-1}$$

$z_{ij}$ |else

Let  $\mathbf{x}_{ij} = (x_{i,j,1}, x_{i,j,2}, \dots, x_{i,j,p_i})$ , where  $p_i$  is the number of covariates of the  $i^{\text{th}}$  dimension, and  $\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_i$ . Then,

$$\begin{aligned}
& p(z_{ij}|\text{else}) \\
& \propto p(y_{ij}|z_{ij}, \boldsymbol{\alpha}_i)p(\mathbf{z}_j|\boldsymbol{\beta}, \Sigma) \\
& \propto \underbrace{\prod_{i=1}^m I\{\alpha_{i,y_{ij}-1} < z_{ij} \leq \alpha_{i,y_{ij}}\}}_{I^*} \exp\left\{-\frac{1}{2}[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}]'\Sigma^{-1}[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(z_{ij} - \mu_{ij})^2\Sigma_{ii}^{-1} + 2\left(\sum_{l \neq i} (z_{lj} - \mu_{lj})\Sigma_{il}^{-1}z_{ij}\right)\right]\right\} I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}\left[z_{ij}^2 - 2\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}}\sum_{l \neq i} \Sigma_{il}^{-1}(z_{lj} - \mu_{lj})\right)z_{ij}\right]\right\} I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}\left[z_{ij} - \left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}}\sum_{l \neq i} \Sigma_{il}^{-1}(z_{lj} - \mu_{lj})\right)\right]^2\right\} I^*
\end{aligned}$$

Therefore,

$$z_{ij} \sim N\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}}\sum_{l \neq i} \Sigma_{il}^{-1}(z_{lj} - \mu_{lj}), \frac{1}{\Sigma_{ii}^{-1}}\right)$$

truncated to  $(\alpha_{i,y_{ij}-1}, \alpha_{i,y_{ij}})$ .

$\boldsymbol{\alpha}_i$ |else

Let  $c_i$  be the number of levels of the  $i$ -th ordinal response. Then, for  $r = 2, \dots, c_i - 2$ ,

$$\alpha_{i,r}|\text{else} \sim \text{Unif}\left(\max_{j \in (1, \dots, T)} \{z_{ij} : y_{ij} = r\}, \min_{j \in (1, \dots, T)} \{z_{ij} : y_{ij} = r + 1\}\right)$$

for  $i = 1, 2, \dots, m$ .

The two end-cutoffs are fixed, i.e.,  $\alpha_{i,1} = 0$  and  $\alpha_{i,c_i-1} = 1$ .

## Joint Posterior Distribution

$$\begin{aligned}
& p(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{V}, \mathbf{s} | \mathbf{Y}) \\
& \propto p(\mathbf{Y} | \mathbf{Z}) p(\mathbf{Z} | \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{s}) p(\boldsymbol{\beta}, \boldsymbol{\Sigma} | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) p(\mathbf{s} | \mathbf{V}) p(\mathbf{V}) \\
& \propto \prod_{j=1}^T \underbrace{p(\mathbf{y}_j | \mathbf{z}_j)}_{(1)} \prod_{j=1}^T \underbrace{p(\mathbf{z}_j | \boldsymbol{\beta}(s_j), \boldsymbol{\Sigma}(s_j))}_{(2)} \prod_{k=1}^N \underbrace{\left[ p(\boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k) | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) \right]}_{(3)} \\
& \quad \prod_{j=1}^T \underbrace{p(s_j | \mathbf{V})}_{(4)} \prod_{k=1}^{N-1} \underbrace{p(V_k)}_{(5)}
\end{aligned}$$

for  $j = 1, 2, \dots, T$ ;  $i = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, N$ .

$j$  is the index for individuals.  $i$  is the index for dimensions.  $k$  is the index for components in the mixture of normals.

$$(1) \quad p(\mathbf{y}_j | \mathbf{z}_j) = \prod_{i=1}^m I\{\alpha_{i, y_{ij}-1} < z_{ij} < \alpha_{i, y_{ij}}\},$$

where  $\alpha_{i,r} = \frac{r-1}{c_i-2}$ ,  $r = 1, \dots, c_i - 1$ , are the fixed cutoffs.

$$(2) \quad p(\mathbf{z}_j | \boldsymbol{\beta}(s_j), \boldsymbol{\Sigma}(s_j)) = \phi_m(\mathbf{z}_j | \boldsymbol{\beta}(s_j), \boldsymbol{\Sigma}(s_j))$$

$$(3) \quad p(\boldsymbol{\beta}(k), \boldsymbol{\Sigma}(k) | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) = \phi_{p^+}(\boldsymbol{\beta}(k) | \mathbf{b}, \mathbf{M}) \text{Wish}^{-1}(\boldsymbol{\Sigma}(k) | \mathbf{S}^{-1}, \nu), \text{ where } p^+ = \sum_{i=1}^m p_i$$

$$(4) \quad p(s_j = k | \mathbf{V}) = w_k = V_k \prod_{l < k} (1 - V_l), \text{ where } \mathbf{V} = (V_1, V_2, \dots, V_N)$$

$$(5) \quad p(V_k) \propto (1 - V_k)^{1-\alpha}, \text{ since } V_k \sim \text{Beta}(1, \alpha)$$

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1T} \\ y_{21} & y_{22} & \cdots & y_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mT} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_T \end{bmatrix}, \text{ where } \mathbf{y}_j = \begin{bmatrix} y_{1j} \\ y_{2j} \\ \vdots \\ y_{mj} \end{bmatrix}$$

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{m1} \\ z_{21} & z_{22} & \cdots & z_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ z_{m1} & z_{m2} & \cdots & z_{mT} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \cdots & \mathbf{z}_T \end{bmatrix}, \text{ where } \mathbf{z}_j = \begin{bmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{mj} \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_N \end{bmatrix} \quad \mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_T \end{bmatrix}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1(1) & \beta_1(2) & \cdots & \beta_1(N) \\ \beta_2(1) & \beta_2(2) & \cdots & \beta_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_m(1) & \beta_m(2) & \cdots & \beta_m(N) \end{bmatrix}_{p^+ \times c} = \begin{bmatrix} \boldsymbol{\beta}(1) & \boldsymbol{\beta}(2) & \cdots & \boldsymbol{\beta}(N) \end{bmatrix}$$

where

$$\boldsymbol{\beta}(k) = \begin{bmatrix} \beta_1(k) \\ \beta_2(k) \\ \vdots \\ \beta_m(k) \end{bmatrix}_{p^+ \times 1} \quad \boldsymbol{\beta}_i(k) = \begin{bmatrix} \beta_{i,1}(k) \\ \beta_{i,2}(k) \\ \vdots \\ \beta_{i,p_i}(k) \end{bmatrix}_{p_i \times 1}$$

## Full Conditionals

$V_k$ |else

Let  $n_k$  be the number of observations that follow the  $k^{th}$  component, i.e.

$$n_k = \sum_{j=1}^T I\{s_j = k\}$$

Then,

$$\begin{aligned} p(V_k|\text{else}) &\propto \prod_{j=1}^T p(s_j|V_1, V_2, \dots, V_N)p(V_k) \\ &\propto \prod_{j:s_j \geq k} p(s_j|V_1, V_2, \dots, V_N)p(V_k) \\ &\propto \prod_{j:s_j \geq k} \left\{ V_{s_j} \prod_{l < s_j} (1 - V_l) \right\} p(V_k) \\ &\propto \prod_{j:s_j = k} V_k \prod_{j:s_j > k} (1 - V_k) p(V_k) \\ &\propto V_k^{n_k} (1 - V_k)^{\sum_{l > k} n_l} (1 - V_k)^{\alpha - 1} \\ &\propto V_k^{(n_k + 1) - 1} (1 - V_k)^{(\sum_{l > k} n_l + \alpha) - 1} \end{aligned}$$

So,

$$V_k|\text{else} \sim \text{Beta}(n_k + 1, \sum_{l > k} n_l + \alpha)$$

$s_j$ |else

$$\begin{aligned} p(s_j = k|\text{else}) \\ &\propto p(\mathbf{z}_j|\boldsymbol{\beta}(k), \Sigma(k))p(s_j = k|\mathbf{V}) \\ &\propto \phi_m(\mathbf{z}_j|\mathbf{x}_j\boldsymbol{\beta}(k), \Sigma(k))\left\{V_k \prod_{l < k} (1 - V_l)\right\} \end{aligned}$$

Define

$$\begin{aligned} o_j &= \sum_{k=1}^N p(s_j = k|\text{else}) \triangleq \sum_{k=1}^N p_{j,k}, \\ \boldsymbol{\pi}_j &= \left( \frac{p_{j,1}}{o_j}, \frac{p_{j,2}}{o_j}, \dots, \frac{p_{j,N}}{o_j} \right) \end{aligned}$$

then,

$$s_j|\text{else} \sim \text{Multinomial}(1, \boldsymbol{\pi}_j)$$

$\Sigma^{-1}(\mathbf{k})|\text{else}$

$$\begin{aligned}
& p(\Sigma^{-1}(k)|\text{else}) \\
& \propto \prod_{j:s_j=k} p(\mathbf{z}_j|\boldsymbol{\beta}(k), \Sigma^{-1}(k))p(\Sigma^{-1}(k)|\mathbf{S}, \nu) \\
& \propto \prod_{j:s_j=k} \phi_m(\mathbf{z}_j|\mathbf{x}_j\boldsymbol{\beta}(k), \Sigma^{-1}(k))\text{Wish}(\Sigma^{-1}(k)|\mathbf{S}, \nu) \\
& \propto \prod_{j:s_j=k} \frac{1}{\sqrt{|\Sigma(k)|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))\right\} \\
& \quad \times \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{\nu+m+1} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{S}^{-1}\Sigma^{-1}(k))\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k)) + \text{tr}(\mathbf{S}^{-1}\Sigma^{-1}(k))\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k) + \mathbf{S}^{-1}\Sigma^{-1}(k)\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\underbrace{\left(\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\right)}_{\Delta_2} + \mathbf{S}^{-1}\right)\Sigma^{-1}(k)\right]\right\}
\end{aligned}$$

So,

$$\Sigma^{-1}(k)|\text{else} \sim \text{Wish}\left(\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\right]^{-1}, n_k + \nu\right)$$

Let  $X_i(k)$  be an  $n_k \times p_i$  matrix with rows comprising only those individuals  $j$  with  $s_j = k$ , where  $i = 1, 2, \dots, m$ . Similarly, let  $\mathbf{Z}(i)$  be an  $n_k \times m$  matrix defined as



follows

$$\mathbf{Z}(k) = \begin{bmatrix} \mathbf{z}_1(k) & \mathbf{z}_2(k) & \dots & \mathbf{z}_m(k) \end{bmatrix}_{n_k \times m}$$

where  $\mathbf{z}_i(k)$  are those elements of  $\mathbf{z}_i$  such that  $s_j = k$ . Define

$$\mathbf{\Lambda}(k) = \begin{bmatrix} \mathbf{X}_1(k)\boldsymbol{\beta}_1(k) & \mathbf{X}_2(k)\boldsymbol{\beta}_2(k) & \dots & \mathbf{X}_m(k)\boldsymbol{\beta}_m(k) \end{bmatrix}_{n_k \times m}$$

Then it could be shown that

$$\Delta_2 = [\mathbf{Z}(k) - \mathbf{\Lambda}(k)]' [\mathbf{Z}(k) - \mathbf{\Lambda}(k)]$$

$\beta(k)|\text{else}$

$$\begin{aligned}
& p(\beta(k)|\text{else}) \\
& \propto \prod_{j:s_j=k} p(\mathbf{z}_j|\beta(k), \Sigma(k))p(\beta(k)|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j:s_j=k} \phi_m(\mathbf{z}_j|\mathbf{x}_j\beta(k), \Sigma(k))\phi_{p+}(\beta(k)|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j:s_j=k} \frac{1}{\sqrt{|\Sigma(k)|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\beta(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\beta(k))\right\} \\
& \quad \times \exp\left\{-\frac{1}{2}[\beta(k) - \mathbf{b}]'\mathbf{M}^{-1}[\beta(k) - \mathbf{b}]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\sum_{j:s_j=k} (\beta'(k)\mathbf{x}'_j\Sigma^{-1}(k)\mathbf{x}_j\beta(k) - \beta'(k)\mathbf{x}'_j\Sigma^{-1}(k)\mathbf{z}_j - \mathbf{z}'_j\Sigma^{-1}(k)\mathbf{x}_j\beta(k))\right.\right. \\
& \quad \left.\left.(\beta'(k)\mathbf{M}^{-1}\beta(k) - \beta'(k)\mathbf{M}^{-1}\mathbf{b} - \mathbf{b}'\mathbf{M}^{-1}\beta(k))\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta'(k)\left(\underbrace{\sum_{j:s_j=k} \mathbf{x}'_j\Sigma^{-1}(k)\mathbf{x}_j + \mathbf{M}^{-1}}_{\mathbf{W}(k)}\right)\beta(k) - \left(\underbrace{\sum_{j:s_j=k} \mathbf{z}'_j\Sigma^{-1}(k)\mathbf{x}_j}_{\Delta'_3}\right.\right.\right. \\
& \quad \left.\left.\left. + \mathbf{b}'\mathbf{M}^{-1}\right)\beta(k)\beta'(k)\left(\underbrace{\sum_{j:s_j=k} \mathbf{x}'_j\Sigma^{-1}(k)\mathbf{z}_j + \mathbf{M}^{-1}\mathbf{b}}_{\Delta_3}\right)\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta'(k)\left(\underbrace{\mathbf{W}(k) + \mathbf{M}^{-1}}_{\mathbf{D}^{-1}}\right)\beta(k) - \left(\underbrace{\Delta'_3 + \mathbf{b}'\mathbf{M}^{-1}}_{\mathbf{K}}\right)\beta(k)\right.\right. \\
& \quad \left.\left.- \beta'(k)(\Delta_3 + \mathbf{M}^{-1}\mathbf{b})\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta'(k)\mathbf{D}^{-1} - \mathbf{K})\beta(k) - \beta'(k)\mathbf{K}'\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta(k) - \mathbf{DK}')'\mathbf{D}^{-1}(\beta(k) - \mathbf{DK}')\right]\right\}
\end{aligned}$$

Thus,

$$\beta(k)|\text{else} \sim N_{p+}(\mathbf{DK}', \mathbf{D})$$

where

$$\mathbf{D} = [\mathbf{W}(k) + \mathbf{M}^{-1}]^{-1}$$

$$\mathbf{K} = \Delta'_3 + \mathbf{b}'\mathbf{M}^{-1}$$

Now, let us take a closer look at  $\mathbf{W}(k)$  and  $\Delta_3$ . Recall that  $X_i(k)$  is an  $n_k \times p_i$  matrix with rows comprising only those individuals  $j$  with  $s_j = k$ . Let  $\mathbf{B}(k)$  be the block-diagonal matrix defined as

$$\mathbf{B}(k) = \begin{bmatrix} X_1(k) & & & \\ & X_2(k) & & \\ & & \ddots & \\ & & & X_m(k) \end{bmatrix}$$

Then, it is not difficult to verify that

$$\mathbf{W}(k) = \mathbf{B}'(k) \left( \Sigma^{-1}(k) \otimes \mathbf{I}_{n_k} \right) \mathbf{B}(k)$$

Let  $\mathbf{z}(k) = \left( \mathbf{z}'_1(k), \dots, \mathbf{z}'_m(k) \right)'$ , where  $\mathbf{z}_i(k)$  are those elements of  $\mathbf{z}_i$  such that  $s_j = k$ . Then one can show that

$$\Delta_3 = \mathbf{B}'(k) \left( \Sigma^{-1}(k) \otimes \mathbf{I}_{n_k} \right) \mathbf{z}(k)$$

$z_{ij}|\text{else}$

Let  $\mathbf{x}_{ij} = (x_{i,j,1}, x_{i,j,2}, \dots, x_{i,j,p_i})$ , where  $p_i$  is the number of covariates of the  $i^{\text{th}}$  dimension, and  $\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_i(s_j)$ . Then,

$$\begin{aligned}
& p(z_{ij}|\text{else}) \\
& \propto p(y_{ij}|z_{ij})p(\mathbf{z}_j|\boldsymbol{\beta}(s_j), \Sigma(s_j)) \\
& \propto \underbrace{\prod_{i=1}^m I\{\alpha_{i,y_{ij}-1} < z_{ij} < \alpha_{i,y_{ij}}\}}_{I^*} \\
& \quad \times \exp\left\{-\frac{1}{2}\left[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(s_j)\right]'\Sigma^{-1}(s_j)\left[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(s_j)\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(z_{ij} - \mu_{ij})^2\Sigma_{ii}^{-1}(s_j) + 2\left(\sum_{l \neq i} (z_{lj} - \mu_{lj})\Sigma_{il}^{-1}(s_j)z_{ij}\right)\right]\right\} I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}(s_j)\left[z_{ij}^2 - 2\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj})\right)z_{ij}\right]\right\} I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}(s_j)\left[z_{ij} - \left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj})\right)\right]^2\right\} I^*
\end{aligned}$$

Therefore,

$$z_{ij} \sim \text{N}\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj}), \frac{1}{\Sigma_{ii}^{-1}(s_j)}\right)$$

truncated to  $(\alpha_{i,y_{ij}-1}, \alpha_{i,y_{ij}})$ .

### A.3 DERIVATION OF THE NONPARAMETRIC MODEL WITH HYPERPRIORS

#### Joint Posterior Distribution

$$\begin{aligned}
& p(\mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{V}, \mathbf{s}|\mathbf{Y}) \\
& \propto p(\mathbf{Y}|\mathbf{Z})p(\mathbf{Z}|\boldsymbol{\beta}, \boldsymbol{\Sigma}, \mathbf{s})p(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{b}, \mathbf{M}, \mathbf{S}^{-1})p(\mathbf{s}|\mathbf{V})p(\mathbf{V})p(\mathbf{b}, \mathbf{M})p(\mathbf{S}^{-1}) \\
& \propto \prod_{j=1}^T \underbrace{p(\mathbf{y}_j|\mathbf{z}_j)}_{(1)} \prod_{j=1}^T \underbrace{p(\mathbf{z}_j|\boldsymbol{\beta}(s_j), \Sigma(s_j))}_{(2)} \prod_{k=1}^N \underbrace{\left[p(\boldsymbol{\beta}(k), \Sigma(k)|\mathbf{b}, \mathbf{M}, \mathbf{S}^{-1})\right]}_{(3)} \\
& \quad \prod_{j=1}^T \underbrace{p(s_j|\mathbf{V})}_{(4)} \prod_{k=1}^{N-1} \underbrace{p(V_k)}_{(5)} \underbrace{p(\mathbf{b}, \mathbf{M})}_{(6)} \underbrace{p(\mathbf{S}^{-1})}_{(7)}
\end{aligned}$$

for  $j = 1, 2, \dots, T$ ;  $i = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, N$ .

$j$  is the index for individuals.  $i$  is the index for dimensions.  $k$  is the index for components in the mixture of normals.

$$(1) p(\mathbf{y}_j | \mathbf{z}_j) = \prod_{i=1}^m I\{\alpha_{i, y_{ij}-1} < z_{ij} < \alpha_{i, y_{ij}}\},$$

where  $\alpha_{i,r} = \frac{r-1}{c_i-2}$ ,  $r = 1, \dots, c_i - 1$ , are the fixed cutoffs.

$$(2) p(\mathbf{z}_j | \boldsymbol{\beta}(s_j), \Sigma(s_j)) = \phi_m(\mathbf{z}_j | \boldsymbol{\beta}(s_j), \Sigma(s_j))$$

$$(3) p(\boldsymbol{\beta}(k), \Sigma(k) | \mathbf{b}, \mathbf{M}, \mathbf{S}^{-1}) = \phi_{p^+}(\boldsymbol{\beta}(k) | \mathbf{b}, \mathbf{M}) \text{Wish}^{-1}(\Sigma(k) | \mathbf{S}^{-1}, \nu),$$

where  $p^+ = \sum_{i=1}^m p_i$

$$(4) p(s_j = k | \mathbf{V}) = w_k = V_k \prod_{l < k} (1 - V_l), \text{ where } \mathbf{V} = (V_1, V_2, \dots, V_N)$$

$$(5) p(V_k) \propto (1 - V_k)^{1-\alpha}, \text{ since } V_k \sim \text{Beta}(1, \alpha)$$

$$(6) p(\mathbf{b}, \mathbf{M}) = \text{Wish}_{p^+}^{-1}(\mathbf{M} | \Psi_1^{-1}, \nu_1)$$

$$(7) p(\mathbf{S}^{-1}) = \text{Wish}_m(\mathbf{S}^{-1} | \Psi_2, \nu_2)$$

## Full Conditionals

$V_k$ |else

Let  $n_k$  be the number of observations that fall in the  $k^{th}$  component, i.e.

$$n_k = \sum_{j=1}^T I\{s_j = k\}$$

Then,

$$\begin{aligned} p(V_k|\text{else}) &\propto \prod_{j=1}^T p(s_j|V_1, V_2, \dots, V_N)p(V_k) \\ &\propto \prod_{j:s_j \geq k} p(s_j|V_1, V_2, \dots, V_N)p(V_k) \\ &\propto \prod_{j:s_j \geq k} \left\{ V_{s_j} \prod_{l < s_j} (1 - V_l) \right\} p(V_k) \\ &\propto \prod_{j:s_j=k} V_k \prod_{j:s_j > k} (1 - V_k) p(V_k) \\ &\propto V_k^{n_k} (1 - V_k)^{\sum_{l>k} n_l} (1 - V_k)^{\alpha-1} \\ &\propto V_k^{(n_k+1)-1} (1 - V_k)^{(\sum_{l>k} n_l + \alpha)-1} \end{aligned}$$

So,

$$V_k|\text{else} \sim \text{Beta}(n_k + 1, \sum_{l>k} n_l + \alpha)$$

$s_j$ |else

$$\begin{aligned} p(s_j = k|\text{else}) \\ &\propto p(\mathbf{z}_j|\boldsymbol{\beta}(k), \Sigma(k))p(s_j = k|\mathbf{V}) \\ &\propto \phi_m(\mathbf{z}_j|\boldsymbol{x}_j\boldsymbol{\beta}(k), \Sigma(k))\left\{V_k \prod_{l < k} (1 - V_l)\right\} \end{aligned}$$

Define

$$\begin{aligned} o_j &= \sum_{k=1}^N p(s_j = k|\text{else}) \triangleq \sum_{k=1}^N p_{j,k}, \\ \boldsymbol{\pi}_j &= \left( \frac{p_{j,1}}{o_j}, \frac{p_{j,2}}{o_j}, \dots, \frac{p_{j,N}}{o_j} \right) \end{aligned}$$

then,

$$s_j|\text{else} \sim \text{Multinomial}(1, \boldsymbol{\pi}_j)$$

$\Sigma^{-1}(\mathbf{k})|\text{else}$

$$\begin{aligned}
& p(\Sigma(k)|\text{else}) \\
& \propto \prod_{j:s_j=k} p(\mathbf{z}_j|\boldsymbol{\beta}(k), \Sigma(k))p(\Sigma(k)|\mathbf{S}^{-1}, \nu) \\
& \propto \prod_{j:s_j=k} \phi_m(\mathbf{z}_j|\mathbf{x}_j\boldsymbol{\beta}(k), \Sigma(k))\text{Wish}(\Sigma^{-1}(k)|\mathbf{S}, \nu) \\
& \propto \prod_{j:s_j=k} \frac{1}{\sqrt{|\Sigma(k)|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))\right\} \\
& \quad \times \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{\nu+m+1} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{S}^{-1}\Sigma^{-1}(k))\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k)) + \text{tr}(\mathbf{S}^{-1}\Sigma^{-1}(k))\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\Sigma^{-1}(k) + \mathbf{S}^{-1}\Sigma^{-1}(k)\right]\right\} \\
& \propto \left[\frac{1}{\sqrt{|\Sigma(k)|}}\right]^{(n_k+\nu)+m+1} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\underbrace{\left(\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\right)}_{\Delta_2} + \mathbf{S}^{-1}\right)\Sigma^{-1}(k)\right]\right\}
\end{aligned}$$

So,

$$\Sigma^{-1}(k)|\text{else} \sim \text{Wish}\left(\left[\sum_{j:s_j=k} (\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))(\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(k))'\right]^{-1}, n_k + \nu\right)$$

Please refer to Subsection 2.2.3 for calculation detail of  $\Delta_2$ .



$\mathbf{S}^{-1}|\text{else}$

$$\begin{aligned}
& p(\mathbf{S}^{-1}|\text{else}) \\
& \propto \prod_{k=1}^N p(\boldsymbol{\Sigma}(k)|\mathbf{S}^{-1})p(\mathbf{S}^{-1}) \\
& \propto \left[ \prod_{k=1}^N |\mathbf{S}^{-1}|^{\frac{\nu}{2}} \exp\left\{ -\frac{1}{2}\text{tr}[\mathbf{S}^{-1}\boldsymbol{\Sigma}^{-1}(k)] \right\} \right] |\mathbf{S}^{-1}|^{\frac{\nu_2-m-1}{2}} \exp\left\{ -\frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1}\mathbf{S}^{-1}] \right\} \\
& \propto |\mathbf{S}^{-1}|^{\frac{(N\nu+\nu_2)-m-1}{2}} \exp\left\{ -\frac{1}{2}\sum_{k=1}^N \text{tr}[\mathbf{S}^{-1}\boldsymbol{\Sigma}^{-1}(k)] - \frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1}\mathbf{S}^{-1}] \right\} \\
& \propto |\mathbf{S}^{-1}|^{\frac{(N\nu+\nu_2)-m-1}{2}} \exp\left\{ -\frac{1}{2}\text{tr}\left[\mathbf{S}^{-1}\left(\sum_{k=1}^N \boldsymbol{\Sigma}^{-1}(k) + \boldsymbol{\Psi}_1^{-1}\right)\right] \right\} \\
& \propto |\mathbf{S}^{-1}|^{\frac{(N\nu+\nu_2)-m-1}{2}} \exp\left\{ -\frac{1}{2}\text{tr}\left[\left(\sum_{k=1}^N \boldsymbol{\Sigma}^{-1}(k) + \boldsymbol{\Psi}_1^{-1}\right)\mathbf{S}^{-1}\right] \right\}
\end{aligned}$$

Thus,

$$\mathbf{S}^{-1}|\text{else} \sim \text{Wish}_m\left(\left[\sum_{k=1}^N \boldsymbol{\Sigma}^{-1}(k) + \boldsymbol{\Psi}_1^{-1}\right]^{-1}, N\nu + \nu_2\right)$$

$\beta(k)|\text{else}$

$$\begin{aligned}
& p(\beta(k)|\text{else}) \\
& \propto \prod_{j:s_j=k} p(\mathbf{z}_j|\beta(k), \Sigma(k))p(\beta(k)|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j:s_j=k} \phi_m(\mathbf{z}_j|\mathbf{x}_j\beta(k), \Sigma(k))\phi_{p+}(\beta(k)|\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{j:s_j=k} \frac{1}{\sqrt{|\Sigma(k)|}} \exp\left\{-\frac{1}{2}(\mathbf{z}_j - \mathbf{x}_j\beta(k))'\Sigma^{-1}(k)(\mathbf{z}_j - \mathbf{x}_j\beta(k))\right\} \\
& \quad \times \exp\left\{-\frac{1}{2}[\beta(k) - \mathbf{b}]'\mathbf{M}^{-1}[\beta(k) - \mathbf{b}]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\sum_{j:s_j=k} (\beta'(k)\mathbf{x}'_j\Sigma^{-1}(k)\mathbf{x}_j\beta(k) - \beta'(k)\mathbf{x}'_j\Sigma^{-1}(k)\mathbf{z}_j \right. \right. \\
& \quad \left. \left. - \mathbf{z}'_j\Sigma^{-1}(k)\mathbf{x}_j\beta(k)) + (\beta'(k)\mathbf{M}^{-1}\beta(k) - \beta'(k)\mathbf{M}^{-1}\mathbf{b} - \mathbf{b}'\mathbf{M}^{-1}\beta(k))\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta'(k)\left(\underbrace{\sum_{j:s_j=k} \mathbf{x}'_j\Sigma^{-1}(k)\mathbf{x}_j}_{\mathbf{W}(k)} + \mathbf{M}^{-1}\right)\beta(k) \right. \right. \\
& \quad \left. \left. - \left(\underbrace{\sum_{j:s_j=k} \mathbf{z}'_j\Sigma^{-1}(k)\mathbf{x}_j}_{\Delta'_3} + \mathbf{b}'\mathbf{M}^{-1}\right)\beta(k) \right. \right. \\
& \quad \left. \left. + \beta'(k)\left(\underbrace{\sum_{j:s_j=k} \mathbf{x}'_j\Sigma^{-1}(k)\mathbf{z}_j}_{\Delta_3} + \mathbf{M}^{-1}\mathbf{b}\right)\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\beta'(k)\left(\underbrace{\mathbf{W}(k) + \mathbf{M}^{-1}}_{\mathbf{D}^{-1}}\right)\beta(k) \right. \right. \\
& \quad \left. \left. - \underbrace{(\Delta'_3 + \mathbf{b}'\mathbf{M}^{-1})}_{\mathbf{K}}\beta(k) - \beta'(k)(\Delta_3 + \mathbf{M}^{-1}\mathbf{b})\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta'(k)\mathbf{D}^{-1} - \mathbf{K})\beta(k) - \beta'(k)\mathbf{K}'\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(\beta(k) - \mathbf{DK}')'\mathbf{D}^{-1}(\beta(k) - \mathbf{DK}')\right]\right\}
\end{aligned}$$

Thus,

$$\boldsymbol{\beta}(k)|\text{else} \sim N_{p+}(\mathbf{DK}', \mathbf{D})$$

where

$$\mathbf{D} = [\mathbf{W}(k) + \mathbf{M}^{-1}]^{-1}$$

$$\mathbf{K} = \Delta'_3 + \mathbf{b}' \mathbf{M}^{-1}$$

Please refer to Subsection 2.2.4 for calculation details of  $\mathbf{W}(k)$  and  $\Delta_3$ .

$\mathbf{M}^{-1}|\text{else}$

$$\begin{aligned}
& p(\mathbf{M}^{-1}|\text{else}) \\
& \propto \prod_{k=1}^N p(\boldsymbol{\beta}(k)|\mathbf{b}, \mathbf{M}^{-1})p(\mathbf{b}, \mathbf{M}^{-1}) \\
& \propto \prod_{k=1}^N \frac{1}{\sqrt{|\mathbf{M}|}} \exp\left\{-\frac{1}{2}[\mathbf{b} - \boldsymbol{\beta}(k)]' \mathbf{M}^{-1}[\mathbf{b} - \boldsymbol{\beta}(k)]\right\} |\mathbf{M}^{-1}|^{\frac{\nu_1 - p^+ - 1}{2}} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1} \mathbf{M}^{-1}]\right\} \\
& \propto |\mathbf{M}^{-1}|^{\frac{N}{2}} \exp\left\{-\frac{1}{2} \sum_{k=1}^N [\mathbf{b} - \boldsymbol{\beta}(k)]' \mathbf{M}^{-1}[\mathbf{b} - \boldsymbol{\beta}(k)]\right\} |\mathbf{M}^{-1}|^{\frac{\nu_1 - p^+ - 1}{2}} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1} \mathbf{M}^{-1}]\right\} \\
& \propto |\mathbf{M}^{-1}|^{\frac{(N+\nu_1) - p^+ - 1}{2}} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{k=1}^N (\mathbf{b} - \boldsymbol{\beta}(k))' \mathbf{M}^{-1}(\mathbf{b} - \boldsymbol{\beta}(k))\right] - \frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1} \mathbf{M}^{-1}]\right\} \\
& \propto |\mathbf{M}^{-1}|^{\frac{(N+\nu_1) - p^+ - 1}{2}} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\sum_{k=1}^N (\mathbf{b} - \boldsymbol{\beta}(k))(\mathbf{b} - \boldsymbol{\beta}(k))' \mathbf{M}^{-1}\right] - \frac{1}{2}\text{tr}[\boldsymbol{\Psi}_1^{-1} \mathbf{M}^{-1}]\right\} \\
& \propto |\mathbf{M}^{-1}|^{\frac{(N+\nu_1) - p^+ - 1}{2}} \\
& \quad \times \exp\left\{-\frac{1}{2}\text{tr}\left[\left[\sum_{k=1}^N (\mathbf{b} - \boldsymbol{\beta}(k))(\mathbf{b} - \boldsymbol{\beta}(k))' + \boldsymbol{\Psi}_1^{-1}\right] \mathbf{M}^{-1}\right]\right\}
\end{aligned}$$

So,

$$\mathbf{M}^{-1}|\text{else} \sim \text{Wish}_{p^+} \left( \left[ \sum_{k=1}^N (\mathbf{b} - \boldsymbol{\beta}(k))(\mathbf{b} - \boldsymbol{\beta}(k))' + \boldsymbol{\Psi}_1^{-1} \right]^{-1}, N + \nu_1 \right)$$

$\mathbf{b}|\text{else}$

$$\begin{aligned}
& p(\mathbf{b}|\text{else}) \\
& \propto \prod_{k=1}^N p(\boldsymbol{\beta}(k)|\mathbf{b}, \mathbf{M})p(\mathbf{b}, \mathbf{M}) \\
& \propto \prod_{k=1}^N \exp\left\{-\frac{1}{2}[\mathbf{b} - \boldsymbol{\beta}(k)]' \mathbf{M}^{-1}[\mathbf{b} - \boldsymbol{\beta}(k)]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \sum_{k=1}^N [\mathbf{b} - \boldsymbol{\beta}(k)]' \mathbf{M}^{-1}[\mathbf{b} - \boldsymbol{\beta}(k)]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \sum_{k=1}^N [\mathbf{b}' \mathbf{M}^{-1} \mathbf{b} - \boldsymbol{\beta}'(k) \mathbf{M}^{-1} \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \boldsymbol{\beta}(k)]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \left[ N \mathbf{b}' \mathbf{M}^{-1} \mathbf{b} - \sum_{k=1}^N \boldsymbol{\beta}'(k) \mathbf{M}^{-1} \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \sum_{k=1}^N \boldsymbol{\beta}(k) \right]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \left[ \left( \mathbf{b}' (N \mathbf{M}^{-1}) - \sum_{k=1}^N \boldsymbol{\beta}'(k) \mathbf{M}^{-1} \right) \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \sum_{k=1}^N \boldsymbol{\beta}(k) \right]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \left[ \left( \mathbf{b}' - \sum_{k=1}^N \boldsymbol{\beta}'(k) \mathbf{M}^{-1} N^{-1} \mathbf{M} \right) (N \mathbf{M}^{-1}) \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \sum_{k=1}^N \boldsymbol{\beta}(k) \right]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \left[ \left( \mathbf{b}' - \frac{1}{N} \sum_{k=1}^N \boldsymbol{\beta}'(k) \right) (N \mathbf{M}^{-1}) \mathbf{b} - \mathbf{b}' \mathbf{M}^{-1} \sum_{k=1}^N \boldsymbol{\beta}(k) \right]\right\} \\
& \propto \exp\left\{-\frac{1}{2} \left[ \left( \mathbf{b} - \frac{1}{N} \sum_{k=1}^N \boldsymbol{\beta}(k) \right)' (N \mathbf{M}^{-1}) \left( \mathbf{b} - \frac{1}{N} \sum_{k=1}^N \boldsymbol{\beta}(k) \right) \right]\right\}
\end{aligned}$$

Therefore,

$$\mathbf{b}|\text{else} \sim N_{p+} \left( \frac{1}{N} \sum_{k=1}^N \boldsymbol{\beta}(k), \frac{1}{N} \mathbf{M} \right)$$

$z_{ij}$ |else

Let  $\mathbf{x}_{ij} = (x_{i,j,1}, x_{i,j,2}, \dots, x_{i,j,p_i})$ , where  $p_i$  is the number of covariates of the  $i^{\text{th}}$  dimension, and  $\mu_{ij} = \mathbf{x}_{ij}\boldsymbol{\beta}_i(s_j)$ . Then,

$$\begin{aligned}
& p(z_{ij}|\text{else}) \\
& \propto p(y_{ij}|z_{ij})p(\mathbf{z}_j|\boldsymbol{\beta}(s_j), \Sigma(s_j)) \\
& \propto \underbrace{\prod_{i=1}^m I\{\alpha_{i,y_{ij}-1} < z_{ij} < \alpha_{i,y_{ij}}\}}_{I^*} \\
& \quad \times \exp\left\{-\frac{1}{2}\left[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(s_j)\right]'\Sigma^{-1}(s_j)\left[\mathbf{z}_j - \mathbf{x}_j\boldsymbol{\beta}(s_j)\right]\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[(z_{ij} - \mu_{ij})^2\Sigma_{ii}^{-1}(s_j) + 2\left(\sum_{l \neq i} (z_{lj} - \mu_{lj})\Sigma_{il}^{-1}(s_j)z_{ij}\right)\right]\right\}I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}(s_j)\left[z_{ij}^2 - 2\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj})\right)z_{ij}\right]\right\}I^* \\
& \propto \exp\left\{-\frac{1}{2}\Sigma_{ii}^{-1}(s_j)\left[z_{ij} - \left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj})\right)\right]^2\right\}I^*
\end{aligned}$$

Therefore,

$$z_{ij} \sim \text{N}\left(\mu_{ij} - \frac{1}{\Sigma_{ii}^{-1}(s_j)}\sum_{l \neq i} \Sigma_{il}^{-1}(s_j)(z_{lj} - \mu_{lj}), \frac{1}{\Sigma_{ii}^{-1}(s_j)}\right)$$

truncated to  $(\alpha_{i,y_{ij}-1}, \alpha_{i,y_{ij}})$ .

## A.4 SAMPLE FORTRAN 90 CODE

We posted the code for the nonparametric model with drinking data here. Code for the parametric model and simulation study can be obtained from the author by request.

### Main Program: the Nonparametric Model with Drinking

#### Data

```
!Bayesian Bivariate Ordinal Regression
!Nonparametric model
!Drinking data
!Reduced model: Y1= age gender abuse; Y2= gender abuse;
!Fixed cutoffs
!LPML is calculated for the thinned tail chain

program drinking_npr
  use numerical_libraries
  use module_1 !modules contain some functions/subroutines
  use module_2

  implicit none

  integer,parameter::rd=100000 !Number of iterations in the Gibbs sampling
  double precision,parameter::p0=0.5 !Proportion of burn-in
  !***** Data Input *****
  integer::count=0,i,j,k,l
  double precision::it !iteration number
  integer,parameter::d1=4,d2=3 !dimensions of covariates for Y1 and Y2.
  integer,parameter::dd=d1+d2
  integer,parameter::T=1964,m=2 !T=number of subjects, m=dimension of Y
  integer,parameter::N=10 !total number of components
  integer,parameter::alpha=1 !V~beta(1,alpha)

  double precision,dimension(T,5)::W !Data matrix
  double precision,dimension(T,2)::Y !bivariate responses
  double precision,dimension(T,2)::Z !latent variables
  double precision,dimension(T,d1)::X1 !design matrix for dimension 1
  double precision,dimension(T,d2)::X2 !design matrix for dimension 2
```

```

double precision,dimension(dd,1)::bb !prior mean for beta; beta~N(bb,MM)
double precision,dimension(dd,dd)::MM !prior covariance matrix of beta
double precision,dimension(dd,dd)::Minv !inverse matrix of MM

double precision,dimension(2,2)::Sigma0 !starting value of Sigma
double precision,dimension(2,2)::Prec0,SS,SS_inv
integer,parameter::s=2 !Sigma~IWish(SS,s)

double precision,parameter::a11=0.0,a12=0.5,a13=1.0,a21=0.0,a22=0.5,a23=1.0
integer,dimension(rd,N)::cs !component sizes
integer,dimension(N+1)::cc

!*** Define the chain of variables and statistics ***
double precision,dimension(rd*N,dd)::bbeta
double precision,dimension(rd*N,3)::Prec
double precision,dimension(rd,N)::rho

!*** Some media parameters in Gibbs ***
integer::mn,rr,pp
double precision::vv
double precision,dimension(N-1)::V !Vk~beta(1,alpha), k=1,...,N-1
double precision,dimension(N)::wt !weights of components
integer,dimension(T,N)::ccmi
double precision,dimension(T,N)::P
double precision::sp !sum(P(j,:))

double precision,dimension(2)::mu !mean of Z
double precision,dimension(2,2)::sigma,precision !Sigma and precision of Z
double precision,dimension(2,2)::D,Precision_i, Sigma_i

double precision,dimension(N,dd)::beta_it !beta vector at iteration "it"
double precision,dimension(3)::prec_it !precision for a component
double precision,dimension(d1,1)::beta_hat1 !starting value for beta1
double precision,dimension(d2,1)::beta_hat2 !starting value for beta2
double precision,dimension(dd,1)::beta_hat

double precision,dimension(dd,dd)::V_i,VM
double precision,dimension(dd,1)::mu_beta,beta_avg
double precision,dimension(5)::Q1,Q2
double precision::mu1,s1,lower1,upper1,mu2,s2,lower2,upper2,u1,u2
double precision,dimension(N)::ss1,ss2

double precision,dimension(T,dd+6)::allmat
double precision,dimension(dd+6)::media
integer,dimension(T)::ind

double precision,dimension(T*2,dd)::B !block-diagonal matrix

```



```

double precision,dimension(T,d1)::zeo1 !block of zeros in B
double precision,dimension(T,d2)::zeo2 !block of zeros in B

double precision,dimension(dd,1)::XZ
double precision,dimension(4)::U !vector of uniform random variables

double precision,dimension(N)::a1,b1,a2,b2 !standardized cutoffs
double precision,dimension(T,N)::prob,m1,m2 !LPML calculation
double precision,dimension(T)::prob_wt,pinv !LPML calculation

double precision,dimension(2,5)::cutoff !cutoffs (including the end points)
double precision::bound !a large number

!*** Define some allocatable matrices for Gibbs sampling ***
double precision, dimension(:,),allocatable::Bii,Bi_1,Bi_2,Yi,Zi
double precision, dimension(:,),allocatable::zero1,zero2,X1i,X2i
double precision,dimension(:,),allocatable::diagMat
double precision,dimension(:,),allocatable::SD

double precision::time0,time1,time !system time

bb=reshape((/0.5,0.0,0.0,0.0,0.5,0.0,0.0/),(/dd,1/))
call diag_matrix(dd,MM)
Minv=inv(size(MM,dim=1),MM)
call diag_matrix(s,SS)
SS=SS*12.5
Sigma0=reshape((/1.0,0.0,0.0,1.0/),(/2,2/))
Prec0=inv(size(Sigma0,dim=1),Sigma0)

bound=10**9
cutoff(:,1)=(/-bound,-bound/)
cutoff(:,2)=(/dble(0),dble(0)/)
cutoff(:,3)=(/dble(0.5),dble(0.5)/)
cutoff(:,4)=(/dble(1),dble(1)/)
cutoff(:,5)=(/bound,bound/)

!Import drinking data
open(unit=1,file="C:\Ordinal Regression\drinking\data\drinking.txt")
do i=1,T
read(1,*)W(i,:)
end do
close(unit=1)

Y=W(1:T,1:2)

X1=colbind(VecConstant(T,dble(1)),W(:,3:5),T,1,T,3)

```

```

X2=colbind(VecConstant(T,db1e(1)),W(:,4:5),T,1,T,2)

!Define the block data matrix B
zeo1=MatConstant(T,d1,db1e(0))
zeo2=MatConstant(T,d2,db1e(0))
B(1:T,1:d1)=X1
B(1:T,d1+1:dd)=zeo2
B(T+1:2*T,1:d1)=zeo1
B(T+1:2*T,d1+1:dd)=X2

!*** Set up the initial values of the chains ***

do j=1,T
  do k=1,2
    U=VecConstant(4,db1e(0))
    select case (int(Y(j,k)))
      case(1)
        call random_number(U(1))
        Z(j,k)=-1.0+U(1)
      case(2)
        call random_number(U(2))
        Z(j,k)=0.5*U(2)
      case(3)
        call random_number(U(3))
        Z(j,k)=0.5+0.5*U(3)
      case(4)
        call random_number(U(4))
        Z(j,k)=1.0+U(4)
    end select
  end do
end do

do i=1,N-1
  cs(1,i)=int(T/N)
end do
cs(1,N)=T-sum(cs(1,1:N-1))

beta_hat1=reshape(matmul(inv(d1,matmul(transpose(X1),X1)),
matmul(transpose(X1),Z(:,1))),(/d1,1/))
beta_hat2=reshape(matmul(inv(d2,matmul(transpose(X2),X2)),
matmul(transpose(X2),Z(:,2))),(/d2,1/))
beta_hat=rowbind(beta_hat1,beta_hat2,d1,1,d2,1)

do i=1,N
  Prec((i-1)*rd+1,:)=(/Prec0(1,1),Prec0(1,2),Prec0(2,2)/)
  do k=1,dd
    bbeta((i-1)*rd+1,k)=beta_hat(k,1)
  end do
end do

```

```

    end do
end do

allmat(:,dd+6)=VecConstant(T,dbl(0))

!Initial time;
time0=dbl(0)

!*** Start Gibbs Sampling ***
do it=2,rd

10 vv=1.0
nn=0
V=VecConstant(N-1,dbl(0))
wt=VecConstant(N,dbl(0))
  do i=1,N-1
    nn=nn+cs(it-1,i)
    call drnbet(1,dbl(cs(it-1,i)+1),dbl(alpha+(T-nn)),V(i))
    wt(i)=V(i)*vv
    vv=vv*(1-V(i))
  end do
  wt(N)=vv

  do i=1,N-1
if (wt(i)<wt(i+1)) then
go to 10
end if
  enddo

  do j=1,T
    do i=1,N
      prec_it=Prec((i-1)*rd+(it-1),:)
      precision=reshape((/prec_it(1),prec_it(2),prec_it(2),prec_it(3)/),(/2,2/))
      sigma=inv(2,precision)
      mu(1)=VecMulSum(dd,B(j,:),bbeta((i-1)*rd+(it-1),:))
      mu(2)=VecMulSum(dd,B(j+T,:),bbeta((i-1)*rd+(it-1),:))
      P(j,i)=wt(i)*pmvn(2,Z(j,1:2),mu,sigma)
    end do

    sp=sum(P(j,:))
    P(j,:)=P(j,+)/sp
    call mult(N,P(j,:),ccmi(j,:))

    do i=1,N
      if (ccmi(j,i)==1) then
        ind(j)=i
      end if
    end do
  end do
end do

```

```

    end do
end do

do i=1,N
    cs(it,i)=sum(ccmi(:,i))
end do

allmat(:,1)=ind
allmat(:,2:d1+1)=X1
allmat(:,d1+2:dd+1)=X2
allmat(:,dd+2:dd+3)=Z
allmat(:,dd+4:dd+5)=Y

cc=VecConstant(N+1,db1e(0))
do i=1,N
cc(i+1)=cc(i)+cs(it,i)
end do

do j=1,T-1
do k=j+1,T
    if (allmat(j,1)>allmat(k,1))then
        media=allmat(k,:)
        allmat(k,:)=allmat(j,:)
        allmat(j,:)=media
    end if
end do
end do

do i=1,N
    pp=cs(it,i)
    if (pp>0) then
allocate(Bii(2*pp,dd))
allocate(Bi_1(pp,dd))
allocate(Bi_2(pp,dd))
allocate(Yi(pp,2))
allocate(Zi(pp,2))
allocate(SD(pp,2))
allocate(zero1(pp,d1))
allocate(zero2(pp,d2))
allocate(X1i(pp,d1))
allocate(X2i(pp,d2))

X1i=allmat(cc(i)+1:cc(i+1),2:d1+1)
X2i=allmat(cc(i)+1:cc(i+1),d1+2:dd+1)
zero1=MatConstant(pp,d1,db1e(0))
zero2=MatConstant(pp,d2,db1e(0))

```

```

Bi_1(:,1:d1)=X1i
Bi_1(:,d1+1:dd)=zero2
Bi_2(:,1:d1)=zero1
Bi_2(:,d1+1:dd)=X2i

Bii=rowbind(Bi_1,Bi_2,pp,dd,pp,dd)
Zi=allmat(cc(i)+1:cc(i+1),dd+2:dd+3)
Yi=allmat(cc(i)+1:cc(i+1),dd+4:dd+5)

SD=Zi-colbind(matmul(X1i,bbeta((i-1)*rd+it-1,1:d1)),
matmul(X2i,bbeta((i-1)*rd+it-1,d1+1:dd)),pp,1,pp,1)
D=matmul(transpose(SD),SD)

Precision_i=wishart(2,pp+s,inv(2,D+inv(2,SS)))
Prec((i-1)*rd+it,:)=(/Precision_i(1,1),Precision_i(1,2),Precision_i(2,2)/)
Precision(i) is updated

V_i(1:d1,1:d1)=matmul(transpose(X1i),X1i)*Precision_i(1,1)
V_i(1:d1,d1+1:dd)=matmul(transpose(X1i),X2i)*Precision_i(1,2)
V_i(d1+1:dd,1:d1)=matmul(transpose(X2i),X1i)*Precision_i(2,1)
V_i(d1+1:dd,d1+1:dd)=matmul(transpose(X2i),X2i)*Precision_i(2,2)

VM=inv(dd,V_i+Minv)
XZ(1:d1,1)=matmul(transpose(X1i),Zi(:,1))*Precision_i(1,1)
+matmul(transpose(X1i),Zi(:,2))*Precision_i(1,2)
XZ(d1+1:dd,1)=matmul(transpose(X2i),Zi(:,1))*Precision_i(1,2)
+matmul(transpose(X2i),Zi(:,2))*Precision_i(2,2)

mu_beta=matmul(VM,XZ+matmul(Minv,reshape(bb,(/dd,1/))))
bbeta((i-1)*rd+it,:)=rmvn(dd,mu_beta,VM) !Beta(i) is updated

s1=dsqrt(1/Prec((i-1)*rd+it,1))
s2=dsqrt(1/Prec((i-1)*rd+it,3))

Sigma_i=inv(2,Precision_i)
ss1(i)=dsqrt(Sigma_i(1,1))
ss2(i)=dsqrt(Sigma_i(2,2))
rho(it,i)=Sigma_i(1,2)/(ss1(i)*ss2(i))

do rr=1,pp
  Q1=(/0.0,0.0,0.0,0.0,1.0/)
  Q2=(/0.0,0.0,0.0,0.0,1.0/)

mu1=VecMulSum(dd,Bi_1(rr,:),bbeta((i-1)*rd+it,:))
-Prec((i-1)*rd+it,2)/Prec((i-1)*rd+it,1)*(Zi(rr,2)
-VecMulSum(dd,Bi_2(rr,:),bbeta((i-1)*rd+it,:))
Q1(2)=dnordf((a11-mu1)/s1);

```

```

Q1(3)=dnordf((a12-mu1)/s1);
Q1(4)=dnordf((a13-mu1)/s1)
lower1=Q1(Yi(rr,1));upper1=Q1(Yi(rr,1)+1)
call random_number(u1)
u1=u1*(upper1-lower1)+lower1
Zi(rr,1)=dnorin(u1)*s1+mu1

mu2=VecMulSum(dd,Bi_2(rr,:),bbeta((i-1)*rd+it,:))
-Prec((i-1)*rd+it,2)/Prec((i-1)*rd+it,3)*(Zi(rr,1)
-VecMulSum(dd,Bi_1(rr,:),bbeta((i-1)*rd+it,:))
Q2(2)=dnordf((a21-mu2)/s2);
Q2(3)=dnordf((a22-mu2)/s2);
Q2(4)=dnordf((a23-mu2)/s2)
lower2=Q2(Yi(rr,2));upper2=Q2(Yi(rr,2)+1)
call random_number(u2)
u2=u2*(upper2-lower2)+lower2
Zi(rr,2)=dnorin(u2)*s2+mu2

end do

allmat(cc(i)+1:cc(i+1),dd+2:dd+3)=Zi !Z is updated

deallocate(Bii)
deallocate(Bi_1)
deallocate(Bi_2)
deallocate(Yi)
deallocate(Zi)
deallocate(SD)
deallocate(zero1)
deallocate(X1i)
deallocate(zero2)
deallocate(X2i)

else
  Precision_i=wishart(2,s,SS)
  Prec((i-1)*rd+it,:)
=(/Precision_i(1,1),Precision_i(1,2),Precision_i(2,2)/)
  bbeta((i-1)*rd+it,:)=rmvn(dd,bb,MM)
end if
end do

X1=allmat(:,2:d1+1);
X2=allmat(:,d1+2:dd+1)
Z=allmat(:,dd+2:dd+3)
Y=allmat(:,dd+4:dd+5)
B(1:T,1:d1)=X1
B(1:T,d1+1:dd)=zeo2

```

```

B(T+1:2*T,d1+1:dd)=X2
B(T+1:2*T,1:d1)=zeo1

!Start to calculate CPD for the thinned tail chain
if (it>rd*p0 .and. it/50==int(it/50)) then
do i=1,N
beta_it(i,:)=bbeta((i-1)*rd+it,:)
end do

m1=matmul(X1,transpose(beta_it(:,1:d1)))
m2=matmul(X2,transpose(beta_it(:,d1+1:dd)))

do j=1,T
do i=1,N
a1(i)=(cutoff(1,Y(j,1))-m1(j,i))/ss1(i)
b1(i)=(cutoff(1,Y(j,1)+1)-m1(j,i))/ss1(i)
a2(i)=(cutoff(2,Y(j,2))-m2(j,i))/ss2(i)
b2(i)=(cutoff(2,Y(j,2)+1)-m2(j,i))/ss2(i)
prob(j,i)=DBNRDF(b1(i),b2(i),rho(it,i))-DBNRDF(a1(i),b2(i),rho(it,i))
-DBNRDF(b1(i),a2(i),rho(it,i))+DBNRDF(a1(i),a2(i),rho(it,i))
enddo
prob_wt(j)=VecMulSum(N,prob(j,:),wt)
pinv(j)=1/prob_wt(j)
enddo
allmat(:,dd+6)=allmat(:,dd+6)+pinv

endif

call cpu_time(time1)
time=time1-time0
time0=time1
print*,it,time,time1

end do

!Write the results to text files

open(unit=11,file="C:\Ordinal Regression\MCMC\drinknpr\beta_12.txt")
do it=1,N*rd
write(11,*)bbeta(it,1:2)
end do
close(unit=11)

open(unit=12,file="C:\Ordinal Regression\MCMC\drinknpr\beta_34.txt")
do it=1,N*rd
write(12,*)bbeta(it,3:4)
end do

```

```

close(unit=12)

open(unit=13,file="C:\Ordinal Regression\MCMC\drinknpr\beta_567.txt")
do it=1,N*rd
write(13,*)bbeta(it,5:7)
end do
close(unit=13)

open(unit=20,file="C:\Ordinal Regression\MCMC\drinknpr\precision.txt")
do it=1,N*rd
write(20,*)Prec(it,1:3)
end do
close(unit=20)

open(unit=30,file="C:\Ordinal Regression\MCMC\drinknpr\size1.txt")
do it=1,rd
write(30,*)cs(it,1:N/2)
end do
close(unit=30)

open(unit=31,file="C:\Ordinal Regression\MCMC\drinknpr\size2.txt")
do it=1,rd
write(31,*)cs(it,N/2+1:N)
end do
close(unit=31)

open(unit=40,file="C:\Ordinal Regression\MCMC\drinknpr\rho.txt")
do k=1,N
do it=1,rd
write(40,*)rho(it,k)
end do
enddo
close(unit=40)

open(unit=100,file="C:\Ordinal Regression\MCMC\drinknpr\X.txt")
do i=1,T
write(100,*)X1(i,2:4)
end do
close(unit=100)

open(unit=110,file="C:\Ordinal Regression\MCMC\drinknpr\Y.txt")
do i=1,T
write(110,*)Y(i,1:2)
end do
close(unit=110)

open(unit=111,file="C:\Ordinal Regression\MCMC\drinknpr\Z.txt")

```



```

do i=1,T
write(111,*)Z(i,1:2)
end do
close(unit=111)

open(unit=120,file="C:\Ordinal Regression\MCMC\drinknpr\sump.txt")
do i=1,T
write(120,*)allmat(i,dd+6)
end do
close(unit=120)

end program drinking_npr

```

## Modules

```

MODULE module_1
use numerical_libraries
IMPLICIT NONE
double precision, parameter :: pi=dbl(3.14159265358979)
!*****
contains
!*****

!*****
! pdf of MVN_p(mu,Sigma), p(y|mu,Sigma)
!*****
function pmvn(p,y,mu,Sigma)
integer, intent(in) :: p; integer :: j
double precision :: pmvn; double precision :: det1, det2
integer, dimension(p) :: ipvt
double precision, intent(in), dimension(p) :: mu, y
double precision, intent(in), dimension(p,p) :: Sigma
double precision, dimension(p) :: rj
double precision, dimension(p,p) :: factor, Sinv
! Sigma=U*D*U' (real symmetric)
rj=dbl(0); call dlftsf(p,Sigma,p,factor,p,ipvt)
! first compute inverse Sinv
do j=1, p;
rj(j)=dbl(1);
call dlfsf(p,factor,p,ipvt,rj,Sinv(1:p,j));
rj(j)=dbl(0);
end do
call dlfsf(p,factor,p,ipvt,det1,det2) ! then get determinant
! print *, det1," ",det2
pmvn=exp(-0.5*dot_product(y-mu,matmul(Sinv,y-mu)))

```

```

/dsqrt(dble(2*Pi)**dble(p)
*det1*dble(10)**det2)
end function pmvn

!*****
! pdf of N(mu,1/tau), p(y|mu,1/tau)
!*****
function pn(y,mu,tau)
  double precision :: pn; double precision, intent(in) :: y,mu,tau
  pn=dsqrt(tau)*exp(-0.5*tau*(y-mu)**2)/dsqrt(2*Pi)
end function pn

!*****
! sample y ~ MVN_p(mu,Sigma)
!*****
function rmvn(p,mu,Sigma)
  integer, intent(in) :: p; integer :: i, j
  double precision, dimension(1:p) :: rmvn
  integer, dimension(p) :: ipvt
  double precision, intent(in), dimension(p) :: mu
  double precision, intent(in), dimension(p,p) :: Sigma
  double precision, dimension(p) :: r
  double precision, dimension(p,p) :: cd ! Cholesky decomposition
  call dlchrg(p,Sigma,p,.False.,ipvt,cd,p); call drnoa(p,r)
  do i=2, p; do j=1,i-1; cd(i,j)=0; end do; end do
  rmvn=mu+matmul(r,cd)
end function rmvn

!*****
! sample M-wish_dimen(df, Sigma)
!*****
function wishart(p,df,Sigma)
  integer, intent(in) :: p, df
  double precision, intent(in), dimension(p,p) :: Sigma
  double precision, dimension(p,p) :: Ssq, m, E, L
  double precision, dimension(1:p,1:p) :: wishart
  double precision, dimension(p) :: la
  integer :: i, j
  m=dble(0); do i=1, p;
m(i,i)=dsqrt(rgamma(dble(df-i+1)/dble(2),dble(0.5)));
end do
  do i=2, p; do j=1, i-1; m(j,i)=rnnof(); end do; end do
  call devcsf(p,Sigma,p,la,E,p) ! need symmetric square root of Sigma
  L=0; do i=1, p; L(i,i)=sqrt(la(i)); end do;
  Ssq=matmul(E,matmul(L,transpose(E)))
  wishart=matmul(transpose(matmul(m,Ssq)),matmul(m,Ssq))
end function wishart

```

```

!*****
! random draw from discrete distribution
! with probabilities prob(1),...prob(length)
!*****
integer function disran(length,probs)
  integer, intent(in) :: length
  double precision, intent(in), dimension(length) :: probs
  integer, dimension(1) :: ir
  integer, dimension(length) :: iwk
  double precision, dimension(length) :: wk
  integer :: i
  wk(1)=probs(1)
  do i=2, length; wk(i)=wk(i-1)+probs(i); end do
  wk=wk/wk(length)
  call drngdt(1,1,length,wk,ir)
  disran=ir(1)
end function disran

!*****
! y ~ gamma(a,b)
!*****
function rgamma(a,b)
  double precision :: rgamma; double precision :: u
  double precision, intent(in) :: a, b; double precision, dimension(1) :: r
  call drngam(1,a,r); u=r(1)/b; if(u<1.0E-045) then; rgamma=1.0E-045; else;
  rgamma=u; end if
end function rgamma

!*****
! y ~ N(mu,sigma) ! note sigma is standard deviation not variance
!*****
function rn(mu,sigma)
  double precision :: rn; double precision, intent(in) :: mu, sigma
  double precision, dimension(1) :: r
  call drnnoa(1,r)
  rn=mu+r(1)*sigma
end function rn

END MODULE module_1

module module_2
  implicit none
  contains
!*****
! simple program-specific functions...
!*****

```

```

!*****
! Computes matrix x*x' where x is p by 1 vector
!*****
function xxt(p,x)
  integer, intent(in) :: p; double precision, dimension(p), intent(in) :: x
  double precision, dimension(1:p,1:p) :: xxt; integer :: i, j
  do i=1,p; do j=1,p; xxt(i,j)=x(i)*x(j); end do; end do
end function xxt

!*****
! Inverse of real, symmetric matrix
!*****
function inv(p,Sigma)
  integer, intent(in) :: p;
  double precision, dimension(p,p), intent(in) :: Sigma
  double precision, dimension(1:p,1:p) :: inv;
  integer, dimension(p) :: ipvt
  double precision, dimension(p) :: rj; integer :: j
  double precision, dimension(p,p) :: factor ! Sigma=U*D*U' (real symmetric)
  rj=dbl(0); call dlftsf(p,Sigma,p,factor,p,ipvt)
  do j=1, p; rj(j)=dbl(1);
  call dlfssf(p,factor,p,ipvt,rj,inv(1:p,j)); rj(j)=dbl(0);
  end do
end function inv

!*****
! Vector Multiplication(form a matrix)
!*****
function VecMul(p1,p2,a,b)
  integer,intent(in)::p1,p2
  double precision,dimension(p1),intent(in)::a
  double precision,dimension(p2),intent(in)::b
  double precision,dimension(1:p1,1:p2)::VecMul
  integer::i,j
  do i=1,p1
    do j=1,p2
      VecMul(i,j)=a(i)*b(j)
    end do
  end do
end function VecMul

!*****
! Vector Multiplication(form a scalar)
!*****
function VecMulSum(p,a,b)
  integer,intent(in)::p

```

```

double precision,dimension(p),intent(in)::a
double precision,dimension(p),intent(in)::b
double precision::VecMulSum
integer::i
VecMulSum=0.0
do i=1,p
VecMulSum=VecMulSum+a(i)*b(i)
end do
end function VecMulSum

!*****
! Matrix Multiplication(A*B)
!*****
function MatMulSelf(r1,m,c2,A,B)
integer,intent(in)::r1,m,c2
double precision,dimension(r1,m),intent(in)::A
double precision,dimension(m,c2),intent(in)::B
double precision,dimension(r1,c2)::MatMulSelf
integer::i,j
do i=1,r1
do j=1,c2
MatMulSelf(i,j)=VecMulSum(m,A(i,:),B(:,j))
end do
end do
end function MatMulSelf

!*****
! Create a matrix of constants
!*****
function MatConstant(p1,p2,c)
integer,intent(in)::p1,p2
double precision,intent(in)::c
double precision,dimension(1:p1,1:p2)::MatConstant
integer::i,j
do i=1,p1
do j=1,p2
MatConstant(i,j)=c
end do
end do
end function MatConstant

!*****
! Create a vector of constants
!*****
function VecConstant(p,c)
integer,intent(in)::p

```

```

    double precision,intent(in)::c
    double precision,dimension(p)::VecConstant
    integer::i
    do i=1,p
    VecConstant(i)=c
    end do
end function VecConstant

!*****
! row-bind two matrices
!*****
function rowbind(A,B,r1,c1,r2,c2)
    integer,intent(in)::r1,c1,r2,c2
    double precision,dimension(r1,c1),intent(in)::A
    double precision,dimension(r2,c2),intent(in)::B
    double precision,dimension(r1+r2,c1)::rowbind
    if (c1==c2) then
rowbind(1:r1,1:c1)=A
rowbind(r1+1:r1+r2,1:c1)=B
    else
        print *,"Cannot rowbind"
    end if
end function rowbind

!*****
! column-bind two matrices
!*****
function colbind(A,B,r1,c1,r2,c2)
    integer,intent(in)::r1,c1,r2,c2
    double precision,dimension(r1,c1),intent(in)::A
    double precision,dimension(r2,c2),intent(in)::B
    double precision,dimension(r1,c1+c2)::colbind
    if (r1==r2) then
colbind(1:r1,1:c1)=A
colbind(1:r1,c1+1:c1+c2)=B
    else
        print *,"Cannot colbind"
    end if
end function colbind

!*****
! Kronecker product of two matrices
!*****
    subroutine Kron_Prod(A1,A2,r1,c1,r2,c2,B)
implicit none
integer, intent(in)::r1,c1,r2,c2
double precision, dimension(r1,c1),intent(in) :: A1

```

```

double precision, dimension(r2,c2),intent(in) :: A2
double precision, dimension(r1*r2,c1*c2),intent(out)::B
integer::ii,jj
  do ii=1,r1
do jj=1,c1
  B(((ii-1)*r2+1):ii*r2,((jj-1)*c2+1):jj*c2)=A1(ii,jj)*A2
end do
  end do
end subroutine Kron_Prod

!*****
! Identity matrix
!*****
subroutine diag_matrix(p,M)
implicit none
integer,intent(in)::p
double precision,dimension(p,p),intent(out)::M
integer::i,j
  do i=1,p
do j=1,p
  if (i==j) then
M(i,j)=dble(1)
  else
M(i,j)=dble(0)
  end if
  end do
  end do
end subroutine diag_matrix

!*****
! Generate a multinomial random variable
!*****
subroutine mult(K,p,y)
integer,intent(in)::K
double precision,dimension(K),intent(in)::p
integer,dimension(K),intent(out)::y
integer::m,i
double precision::u
double precision,dimension(K)::pstar
call random_number(u)
m=1;y=0
do i=1,K
  pstar(i)=sum(p(1:i))
  if(u>pstar(i))then
    m=m+1
  end if
end do
end do

```

```
    y(m)=1  
end subroutine  
end module module_2
```



## APPENDIX B

### SUPPLEMENT TO CHAPTER 3

#### B.1 SUPPLEMENTARY TABLE AND FIGURES

Table B.1 Number of measurements by f2 Primary

f2 Primary	Total Measurement
1453	321
1734	341
2063	346
2531	352
3000	353
3563	354
4219	354
5016	356
6000	355
7031	352
8391	337
10031	327
total	4148

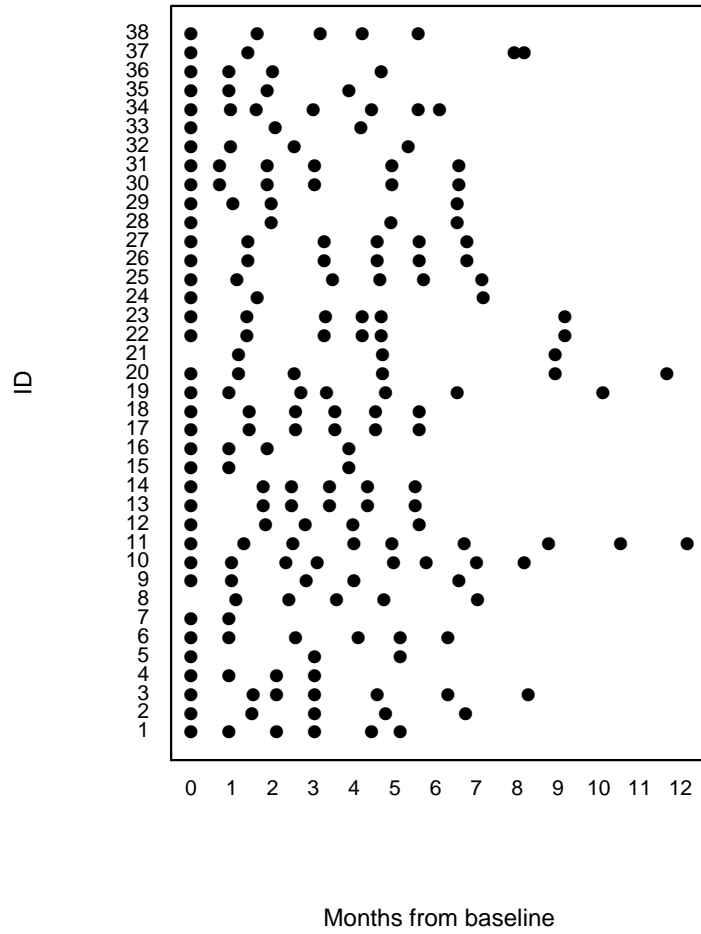


Figure B.1 Followup time (months).

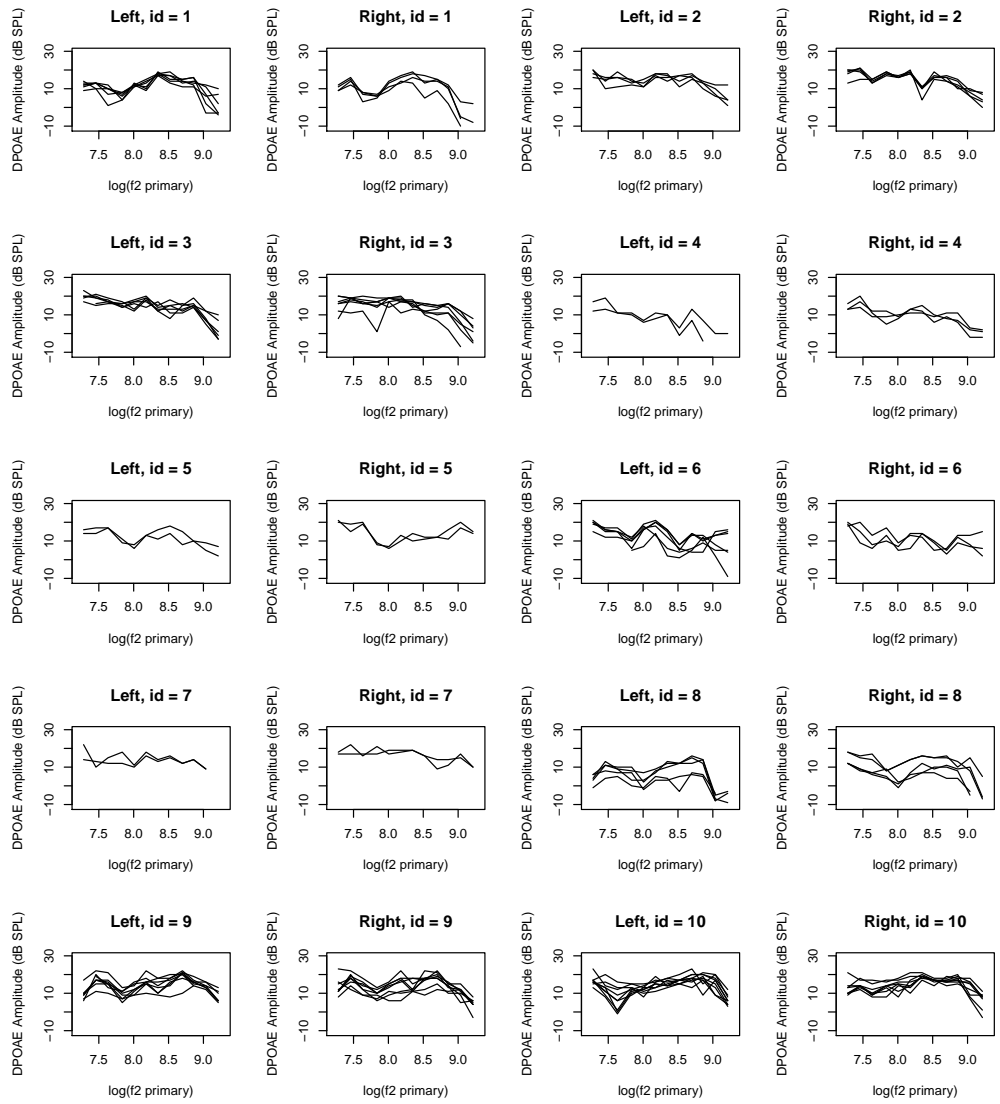


Figure B.2 Individual DP-grams (IDs 1 to 10).

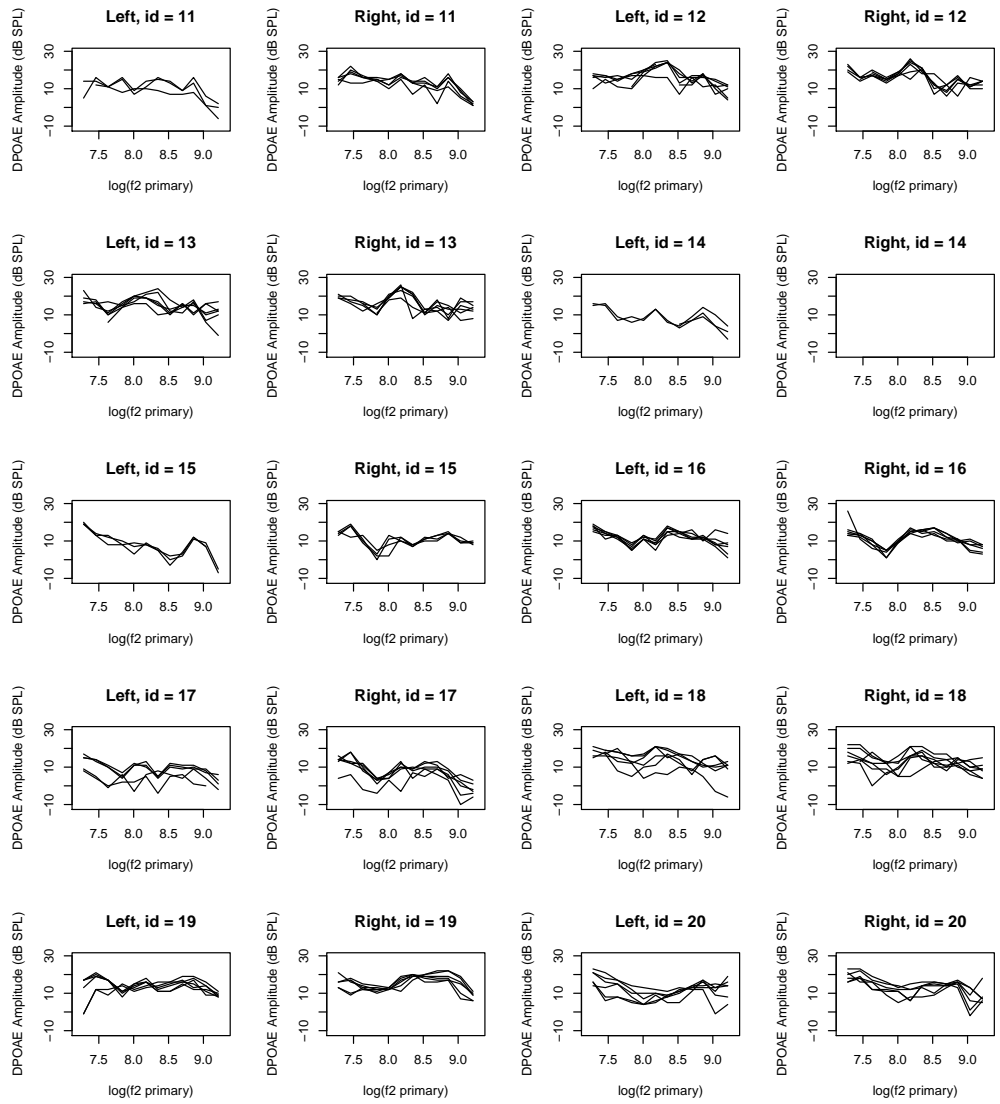


Figure B.3 Individual DP-grams (IDs 11 to 20).

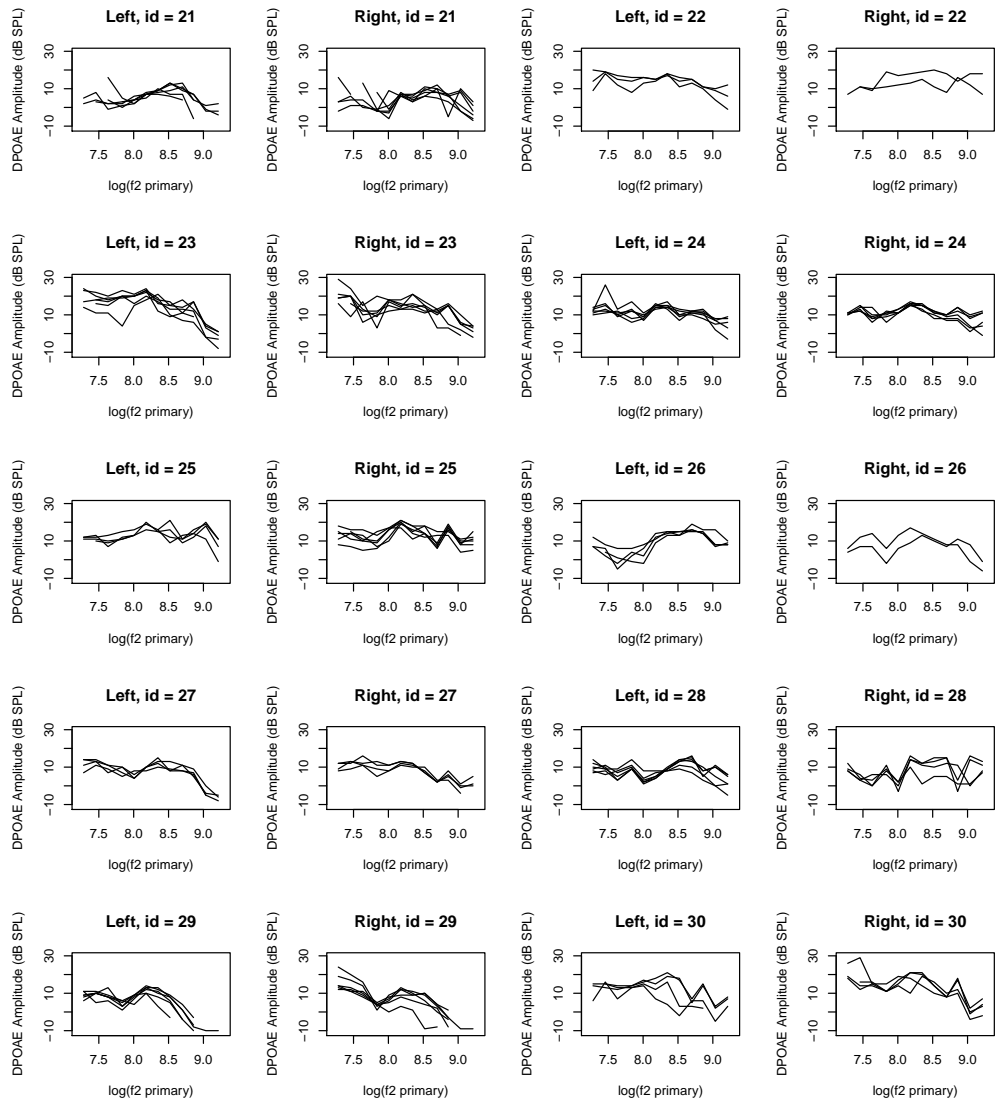


Figure B.4 Individual DP-grams (IDs 21 to 30).

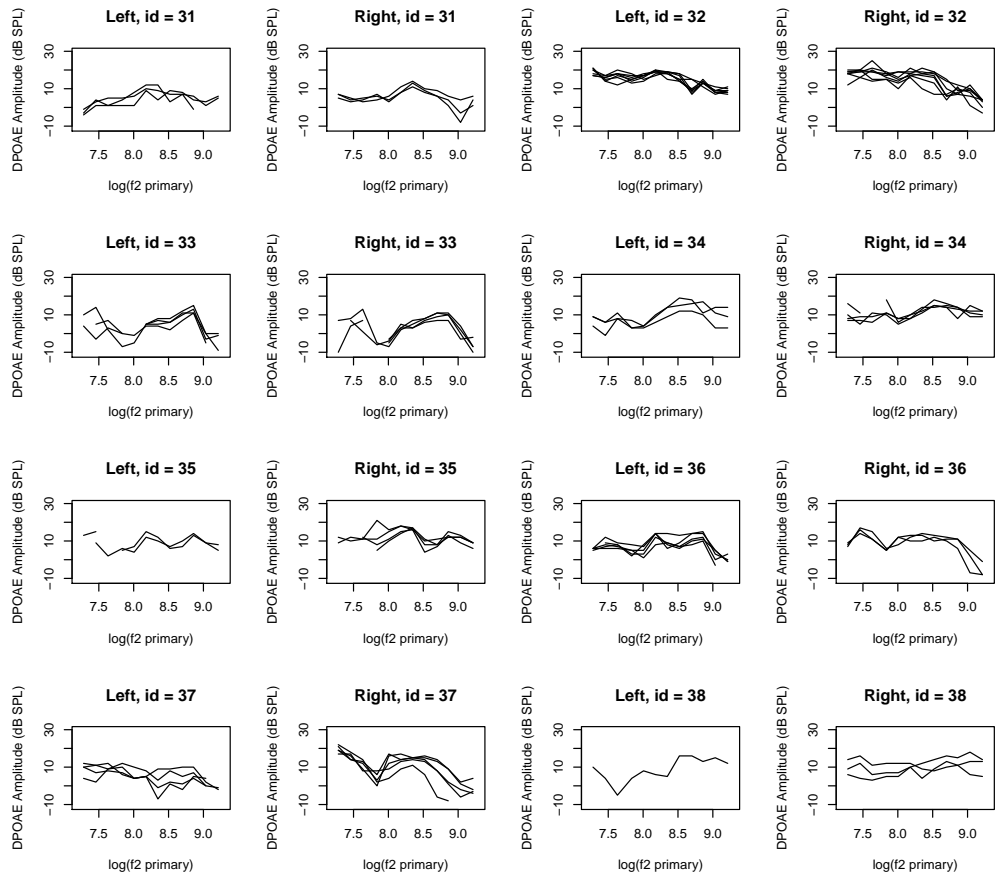


Figure B.5 Individual DP-grams (IDs 31 to 38).

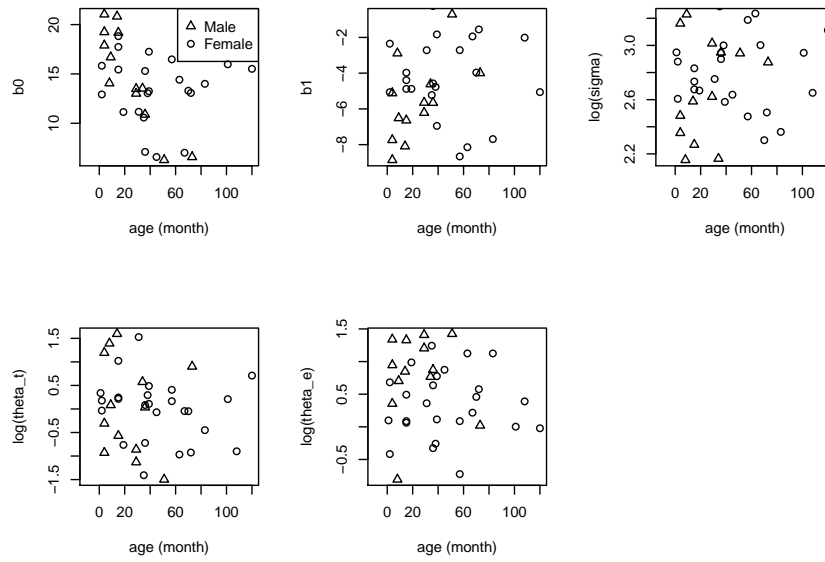


Figure B.6 Posterior means of  $b_0$ ,  $b_1$ ,  $\log(\sigma^2)$ ,  $\log(\theta_t)$ ,  $\log(\theta_e)$  versus age.

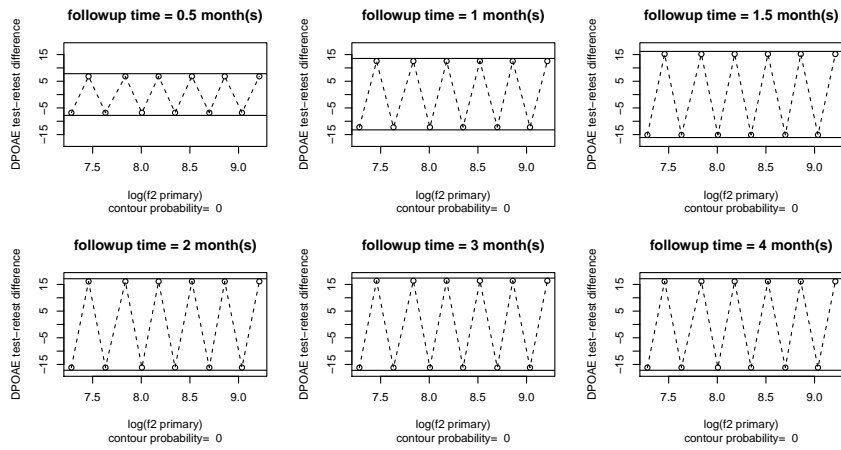


Figure B.7 Extreme test-retest DP-gram differences with contour probabilities; Case 1: oscillatory.

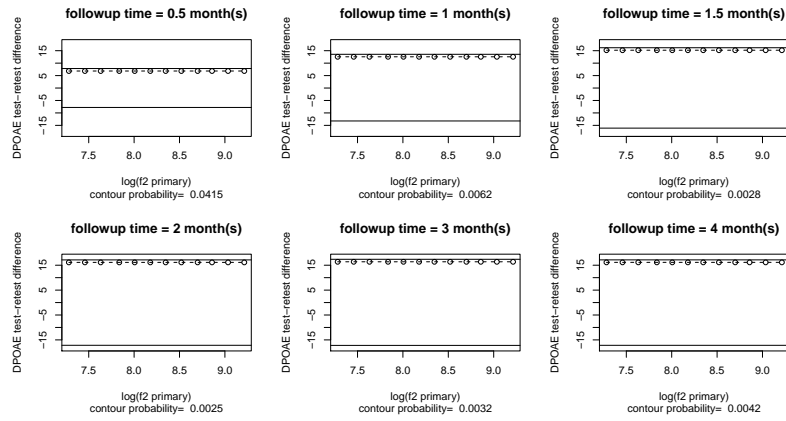


Figure B.8 Extreme test-retest DP-gram differences with contour probabilities; Case 2: overall vertical shift.

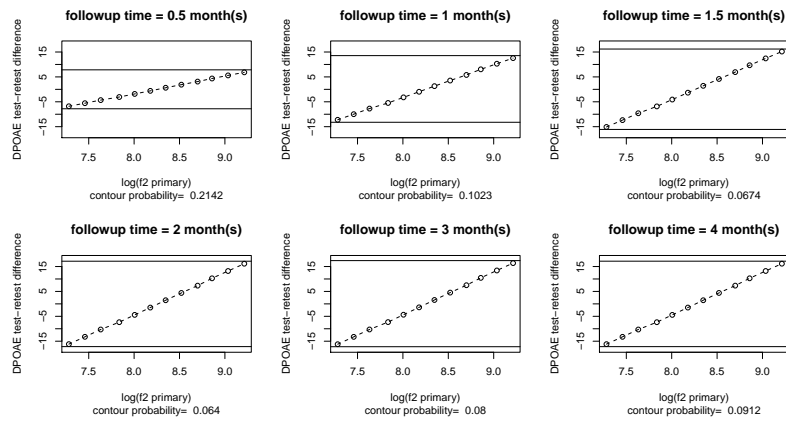


Figure B.9 Extreme test-retest DP-gram differences with contour probabilities; Case 3: monotone.



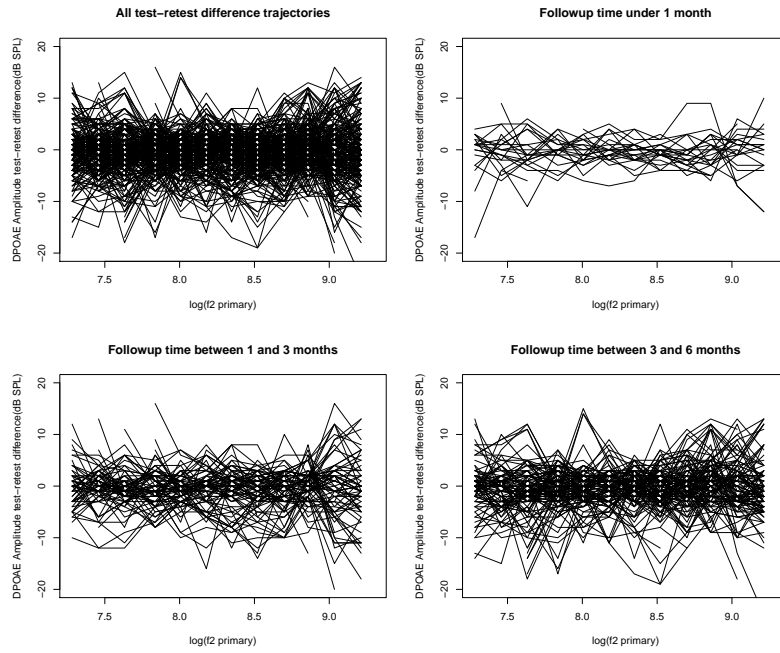


Figure B.10 Test-retest differences in DP-grams from baseline.

## B.2 HIERARCHICAL MODEL

### The Model

$$y_{ijkl} = \mu(f_j) + b_{i0} + b_{i1}f_j + e_{ijkl}$$

where

$$i = 1, 2, \dots, n \text{ (Subject index)}$$

$$j = 1, 2, \dots, F \text{ (F2 frequency levels)}$$

$$k = 1, 2, \dots, T_i \text{ (Time index for Subject } i)$$

$$l = 1, 2; \quad l = 1 \text{ if left ear and } l = 2 \text{ if right ear.}$$

Define the total number of observations for subject  $i$  and frequency  $f_j$  as

$$n_{ij} = \sum_{k=1}^{T_i} L_{ijk}$$

Where

$$L_{ijk} = 1 \text{ if only left or right ear was tested}$$

$$L_{ijk} = 2 \text{ if both ears were tested}$$

### The Response Variable

$$\mathbf{y}_{ijk} = \begin{cases} (y_{ijk1}, y_{ijkL_{ijk}})', & \text{if } L_{ijk} = 2 \\ (y_{ijk1}), & \text{if } L_{ijk} = 1 \end{cases}$$

$$\mathbf{y}_{ij} = (\mathbf{y}'_{ij1}, \mathbf{y}'_{ij2}, \dots, \mathbf{y}'_{ijT_i})'$$

### The Covariates

$$\mathbf{f} = (f_1, f_2, \dots, f_F)'$$

## The Overall Mean

$$\mu(f_j) = \sum_{s=1}^S \gamma_s \phi_s(f_j)$$

with constraints  $\gamma_1 = \gamma_S = 0$ .

## The Error Term

$$\mathbf{e}_{ijk} = \begin{cases} (e_{ijk1}, e_{ijkL_{ijk}})', & \text{if } L_{ijk} = 2 \\ (e_{ijk1}), & \text{if } L_{ijk} = 1 \end{cases}$$

$$\mathbf{e}_{ij} = (\mathbf{e}'_{ij1}, \mathbf{e}'_{ij2}, \dots, \mathbf{e}'_{ijT_i})'$$

and

$$\mathbf{e}_{ij} \sim N_{n_{ij}}(\mathbf{0}, \boldsymbol{\Sigma}_{ij})$$

where  $\boldsymbol{\Sigma}_{ij}$  is the covariance matrix of  $\mathbf{e}_{ij}$  and

$$\text{cov}(e_{ijk_1 l_1}, e_{ijk_2 l_2}) = \sigma_{ij}^2 e^{-\theta_{it}|t_{ik_1} - t_{ik_2}|^2 - \theta_{ie}|l_1 - l_2|^2}$$

## A Concise Version of the Model

Define

$$\mathbf{X}_{ijk} = \begin{cases} (\phi_2(f_j), \phi_3(f_j), \dots, \phi_{S-1}(f_j)), & \text{if } L_{ijk} = 1 \\ \begin{bmatrix} \phi_2(f_j), \phi_3(f_j), \dots, \phi_{S-1}(f_j) \\ \phi_2(f_j), \phi_3(f_j), \dots, \phi_{S-1}(f_j) \end{bmatrix}, & \text{if } L_{ijk} = 2 \end{cases}$$

Let

$$\mathbf{X}_{ij} = \begin{bmatrix} \mathbf{X}_{ij1} \\ \mathbf{X}_{ij2} \\ \vdots \\ \mathbf{X}_{ijT_i} \end{bmatrix}$$

Also define

$$\mathbf{Z}_{ijk} = \begin{cases} (1, f_j), & \text{if } L_{ijk} = 1 \\ \begin{bmatrix} 1 & f_j \\ 1 & f_j \end{bmatrix}, & \text{if } L_{ijk} = 2 \end{cases}$$
$$\mathbf{Z}_{ij} = \begin{bmatrix} \mathbf{Z}_{ij1} \\ \mathbf{Z}_{ij2} \\ \vdots \\ \mathbf{Z}_{ijT_i} \end{bmatrix}$$

Then, the model can be expressed in a more concise form:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\gamma} + \mathbf{Z}_{ij}\mathbf{b}_i + \mathbf{e}_{ij}$$

for  $i = 1, 2, \dots, n; j = 1, 2, \dots, F$ .

## The Priors on Linear Mean and Covariance Parameters

For subject  $i$ ,

$$\mathbf{r}_i = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{v}_i \end{bmatrix}, \quad \text{where } \mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \quad \text{and } \mathbf{v}_i = \begin{bmatrix} \log(\sigma_i^2) \\ \log(\theta_{it}) \\ \log(\theta_{ie}) \end{bmatrix}$$

$\mathbf{r}_1, \dots, \mathbf{r}_n \sim N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$  where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}$$

Further assume

$$\boldsymbol{\mu}_r \sim N_5(\mathbf{m}_0, \mathbf{M}_0)$$

$$\boldsymbol{\Sigma}_r^{-1} \sim \text{Wish}_5(\mathbf{Q}, q), q = 5$$

## The Prior of Penalized B-spline Parameters

$$\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{S-1})' \sim N_{S-2}(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$$

$$\boldsymbol{\mu}_\gamma = \mathbf{0}_{S-2} \text{ and } \boldsymbol{\Sigma}_\gamma^{-1} = \lambda \mathbf{D}' \mathbf{D}, \text{ where}$$

$$\mathbf{D} = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & 1 & -2 \end{bmatrix}_{(S-2) \times (S-2)}$$

The penalizing parameter  $\lambda$  follows Gamma distribution:

$$\lambda \sim \text{Gamma}(\alpha_1, \alpha_2), \quad \alpha_1 = 1 \text{ and } \alpha_2 = 0.005 \text{ or } 0.0005.$$

## Full Conditionals

The joint posterior distribution is

$$\begin{aligned} & p(\mathbf{r}, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}, \boldsymbol{\gamma}, \lambda | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto p(\mathbf{y} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\gamma}, \mathbf{r}, \lambda) p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\mu}_r | \mathbf{m}, \mathbf{M}) p(\boldsymbol{\Sigma}_r^{-1} | \mathbf{Q}, q) p(\boldsymbol{\gamma} | \boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma) p(\lambda | \alpha, \beta). \end{aligned}$$

$\boldsymbol{\mu}_r$  | else

$$\begin{aligned}
p(\boldsymbol{\mu}_r | \text{else}) &\propto p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\mu}_r | \mathbf{m}, \mathbf{M}) \\
&\propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\mathbf{r}_i - \boldsymbol{\mu}_r)\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_r - \mathbf{m})' \mathbf{M}^{-1} (\boldsymbol{\mu}_r - \mathbf{m})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_r - \mathbf{K})' \mathbf{v}^{-1} (\boldsymbol{\mu}_r - \mathbf{K})\right\}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{v} &= [\mathbf{M}^{-1} + n\boldsymbol{\Sigma}_r^{-1}]^{-1} \\
\mathbf{K} &= \mathbf{M}^{-1}\mathbf{m} + \boldsymbol{\Sigma}_r^{-1} \left[ \sum_{i=1}^n \mathbf{r}_i \right]
\end{aligned}$$

So,

$$\boldsymbol{\mu}_r | \text{else} \sim N_5(\mathbf{V}\mathbf{K}, \mathbf{v})$$

$\boldsymbol{\Sigma}_r^{-1}$  | else

$$\begin{aligned}
p(\boldsymbol{\Sigma}_r^{-1} | \text{else}) &\propto p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\Sigma}_r^{-1} | \mathbf{Q}, q) \\
&\propto \prod_{i=1}^n |\boldsymbol{\Sigma}_r^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\mathbf{r}_i - \boldsymbol{\mu}_r)\right\} |\boldsymbol{\Sigma}_r^{-1}|^{\frac{q-5-1}{2}} \\
&\quad \times \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{Q}^{-1}\boldsymbol{\Sigma}_r^{-1})\right\} \\
&\propto |\boldsymbol{\Sigma}_r^{-1}|^{\frac{(n+q)-5-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\left(\sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' + \mathbf{Q}^{-1}\right)\boldsymbol{\Sigma}_r^{-1}\right]\right\}
\end{aligned}$$

Thus,

$$\boldsymbol{\Sigma}_r^{-1} | \text{else} \sim \text{Wish}_5\left(\left[\sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' + \mathbf{Q}^{-1}\right]^{-1}, n + q\right)$$

Note that, if we let

$$\mathbf{e} = (\mathbf{r}_1 - \boldsymbol{\mu}_r, \dots, \mathbf{r}_n - \boldsymbol{\mu}_r)_{5 \times n}$$

Then,

$$\sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' = \mathbf{E}\mathbf{E}'$$

$\mathbf{b}_i$  | else

First, we need to derive the form of  $\Sigma_r^{-1}$  in terms of  $\Sigma_b$ ,  $\Sigma_v$ , and  $\Sigma_{bv}$ . Recall

$$\Sigma_r = \begin{bmatrix} \Sigma_b & \Sigma_{bv} \\ \Sigma'_{bv} & \Sigma_v \end{bmatrix}$$

then,

$$\begin{aligned} \Sigma_r^{-1} &= \begin{bmatrix} (\Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv})^{-1} & -(\Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv})^{-1}\Sigma_{bv}\Sigma_v^{-1} \\ -(\Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv})^{-1}\Sigma'_{bv}\Sigma_b^{-1} & (\Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (\Sigma_b^*)^{-1} & -(\Sigma_b^*)^{-1}\Sigma_{bv}\Sigma_v^{-1} \\ -(\Sigma_v^*)^{-1}\Sigma'_{bv}\Sigma_b^{-1} & (\Sigma_v^*)^{-1} \end{bmatrix} \end{aligned}$$

where

$$\begin{aligned} \Sigma_b^* &= \Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv} \\ \Sigma_v^* &= \Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv} \end{aligned}$$

Second, we will find the posterior distribution of  $\mathbf{b}_i$  given  $\mathbf{v}_i$ . Recall that

$$\mathbf{r}_i = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{v}_i \end{bmatrix} \quad \mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \quad \mathbf{v}_i = \begin{bmatrix} \log(\sigma_i^2) \\ \log(\theta_{it}) \\ \log(\theta_{ie}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}$$



So,

$$\begin{aligned}
p(\mathbf{b}_i|\mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) &\propto p(\mathbf{b}_i, \mathbf{v}_i|\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1}(\mathbf{r}_i - \boldsymbol{\mu}_r)\right\} \\
&= \exp\left\{-\frac{1}{2} \begin{bmatrix} \mathbf{b}_i - \boldsymbol{\beta} \\ \mathbf{v}_i - \boldsymbol{\tau} \end{bmatrix}' \boldsymbol{\Sigma}_r^{-1} \begin{bmatrix} \mathbf{b}_i - \boldsymbol{\beta} \\ \mathbf{v}_i - \boldsymbol{\tau} \end{bmatrix}\right\} \\
&= \exp\left\{-\frac{1}{2} \begin{bmatrix} \mathbf{b}_i^* \\ \mathbf{v}_i^* \end{bmatrix}' \begin{bmatrix} (\boldsymbol{\Sigma}_b^*)^{-1} & -(\boldsymbol{\Sigma}_b^*)^{-1} \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} \\ -(\boldsymbol{\Sigma}_v^*)^{-1} \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} & (\boldsymbol{\Sigma}_v^*)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{b}_i^* \\ \mathbf{v}_i^* \end{bmatrix}\right\} \\
&\propto (\mathbf{b}_i^* - \boldsymbol{\Sigma}_b^* \mathbf{W})' (\boldsymbol{\Sigma}_b^*)^{-1} (\mathbf{b}_i^* - \boldsymbol{\Sigma}_b^* \mathbf{W})
\end{aligned}$$

where

$$\mathbf{W} = (\boldsymbol{\Sigma}_b^*)^{-1} \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} \mathbf{v}^*$$

Therefore,

$$\mathbf{b}_i^* | \mathbf{v}_i^* \sim N_2(\boldsymbol{\Sigma}_b^* \mathbf{W}, \boldsymbol{\Sigma}_b^*)$$

and,

$$\mathbf{b}_i | \mathbf{v}_i \sim N_2(\boldsymbol{\mu}_{b_i|\mathbf{v}_i}, \boldsymbol{\Sigma}_{b_i|\mathbf{v}_i})$$

where

$$\boldsymbol{\mu}_{b_i|\mathbf{v}_i} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_b^* \mathbf{W} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} (\mathbf{v}_i - \boldsymbol{\tau})$$

$$\boldsymbol{\Sigma}_{b_i|\mathbf{v}_i} = \boldsymbol{\Sigma}_b^* = \boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Sigma}'_{bv}$$

Now we are ready to find the posterior of  $\mathbf{b}_i$ :

$$\begin{aligned}
p(\mathbf{b}_i|\text{else}) &\propto p(\mathbf{y}_i|\mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) p(\mathbf{b}_i|\mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\
&\propto \exp\left\{-\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \boldsymbol{\mu}_{b_i|\mathbf{v}_i})' \boldsymbol{\Sigma}_{b_i|\mathbf{v}_i}^{-1} (\mathbf{b}_i - \boldsymbol{\mu}_{b_i|\mathbf{v}_i})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \mathbf{GH})' \mathbf{G}^{-1} (\mathbf{b}_i - \mathbf{GH})\right\}
\end{aligned}$$

where

$$\begin{aligned} \mathbf{G} &= [\mathbf{Z}'_i \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i + \boldsymbol{\Sigma}_{b_i|v_i}^{-1}]^{-1} \\ \mathbf{H} &= \mathbf{Z}'_i \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\gamma}) + \boldsymbol{\Sigma}_{b_i|v_i}^{-1} \boldsymbol{\mu}_{b_i|v_i} \end{aligned}$$

Thus,

$$\mathbf{b}_i | \text{else} \sim N_2(\mathbf{GH}, \mathbf{G})$$

$\mathbf{v}_i$  | else

Following similar arguments of  $\mathbf{b}_i$ , we conclude that

$$\mathbf{v}_i | \mathbf{b}_i \sim N_3(\boldsymbol{\mu}_{v_i|b_i}, \boldsymbol{\Sigma}_{v_i|b_i})$$

where

$$\begin{aligned} \boldsymbol{\mu}_{v_i|b_i} &= \boldsymbol{\tau} + \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} (\mathbf{b}_i - \boldsymbol{\beta}) \\ \boldsymbol{\Sigma}_{v_i|b_i} &= \boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\Sigma}_{bv} \end{aligned}$$

The posterior distribution of  $\mathbf{v}_i$  can be derived as follows:

$$\begin{aligned} p(\mathbf{v}_i | \text{else}) &\propto p(\mathbf{y}_i | \mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) p(\mathbf{v}_i | \mathbf{b}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\ &\propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i|b_i})' \boldsymbol{\Sigma}_{v_i|b_i}^{-1} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i|b_i}) \right\} \end{aligned}$$

where  $\boldsymbol{\Sigma}_i$  is a function of  $\mathbf{v}_i$ . There is no closed form for this posterior distribution.

We adopted adaptive Metropolis algorithm to generate samples.

$\gamma$  | else

$$\begin{aligned}
p(\gamma|\text{else}) &\propto p(\mathbf{y}|\gamma)p(\gamma) \\
&\propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\gamma + \mathbf{Z}_i\mathbf{b}_i)]'\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - (\mathbf{X}_i\gamma + \mathbf{Z}_i\mathbf{b}_i)]\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)'\boldsymbol{\Sigma}_\gamma^{-1}(\gamma - \boldsymbol{\mu}_\gamma)\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\gamma - \mathbf{LJ})'\mathbf{l}^{-1}(\gamma - \mathbf{LJ})\right\}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{l} &= \left[\sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i + \boldsymbol{\Sigma}_\gamma^{-1}\right]^{-1} \\
\mathbf{J} &= \sum_{i=1}^n \mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{Z}_i\mathbf{b}_i) + \boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\mu}_\gamma
\end{aligned}$$

Thus,

$$\gamma|\text{else} \sim N_{S-2}(\mathbf{LJ}, \mathbf{l})$$

$\lambda$  | else

$$\begin{aligned}
p(\lambda|\text{else}) &\propto p(\gamma|\lambda)p(\lambda|\alpha_1, \alpha_2) \\
&\propto \frac{1}{\sqrt{|\boldsymbol{\Sigma}_\gamma|}} \exp\left\{-\frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)'\boldsymbol{\Sigma}_\gamma^{-1}(\gamma - \boldsymbol{\mu}_\gamma)\right\} \lambda^{\alpha_1-1} \exp\{-\alpha_2\lambda\} \\
&\propto \sqrt{|\lambda\mathbf{D}'\mathbf{D}|} \exp\left\{-\frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)'(\lambda\mathbf{D}'\mathbf{D})(\gamma - \boldsymbol{\mu}_\gamma)\right\} \lambda^{\alpha_1-1} \exp\{-\alpha_2\lambda\} \\
&\propto \lambda^{(\frac{S-2}{2}+\alpha_1)-1} \exp\left\{-\left[\frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)'(\mathbf{D}'\mathbf{D})(\gamma - \boldsymbol{\mu}_\gamma) + \alpha_2\right]\lambda\right\}
\end{aligned}$$

Therefore,

$$\lambda|\text{else} \sim \text{Gamma}\left(\frac{S-2}{2} + \alpha_1, \frac{1}{2}(\gamma - \boldsymbol{\mu}_\gamma)'(\mathbf{D}'\mathbf{D})(\gamma - \boldsymbol{\mu}_\gamma) + \alpha_2\right)$$

### B.3 AGE-GENDER MODEL

Instead of assuming

$$\mathbf{r}_1, \dots, \mathbf{r}_n \sim N_5(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \text{ where } \boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix},$$

we can let  $\boldsymbol{\mu}_r$  be dependent on age and gender of the patient. In particular, let

$$\mathcal{M}_r = \begin{bmatrix} \mathcal{B}' \\ \mathcal{T}' \end{bmatrix} \text{ and } \boldsymbol{\mu}_{ri} = \mathcal{M}_r A_i = \begin{bmatrix} \mathcal{B}' A_i \\ \mathcal{T}' A_i \end{bmatrix}$$

where

$$\mathcal{B}' = \begin{bmatrix} \beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{20} & \beta_{21} & \beta_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}'_1 \\ \boldsymbol{\beta}'_2 \end{bmatrix} \quad \mathcal{T}' = \begin{bmatrix} \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \\ \tau_{30} & \tau_{31} & \tau_{32} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\tau}'_1 \\ \boldsymbol{\tau}'_2 \\ \boldsymbol{\tau}'_3 \end{bmatrix} \quad A_i = \begin{bmatrix} 1 \\ a_i \\ g_i \end{bmatrix}$$

The joint posterior distribution of this age-gender-specific model becomes:

$$\begin{aligned} & p(\mathbf{r}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3, \boldsymbol{\Sigma}_r^{-1}, \boldsymbol{\gamma}, \lambda | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto p(\mathbf{y} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\gamma}, \mathbf{r}, \lambda) p(\mathbf{r} | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3, \boldsymbol{\Sigma}_r^{-1}) \prod_{k=1}^2 p(\boldsymbol{\beta}_k | \boldsymbol{\mu}_{\beta_k}, \boldsymbol{\Sigma}_{\beta_k}) \prod_{k=1}^3 p(\boldsymbol{\tau}_k | \boldsymbol{\mu}_{\tau_k}, \boldsymbol{\Sigma}_{\tau_k}) \\ & p(\boldsymbol{\Sigma}_r^{-1} | \mathbf{Q}, q) p(\boldsymbol{\gamma} | \boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}) p(\lambda | \alpha, \beta). \end{aligned}$$

$\boldsymbol{\beta}_1$  | else

$$\begin{aligned} & p(\boldsymbol{\beta}_1 | \text{else}) \\ & \propto p(\mathbf{r} | \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \boldsymbol{\tau}_3, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\beta}_1 | \boldsymbol{\mu}_{\beta_1}, \boldsymbol{\Sigma}_{\beta_1}) \\ & \propto \prod_{i=1}^n p(\mathbf{r}_i | \boldsymbol{\mu}_{ri}, \boldsymbol{\Sigma}_r) p(\boldsymbol{\beta}_1 | \boldsymbol{\mu}_{\beta_1}, \boldsymbol{\Sigma}_{\beta_1}) \\ & \propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2} [\mathbf{r}_i - \mathcal{M}_r A_i]' \boldsymbol{\Sigma}_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \right\} \\ & \quad \times \exp \left\{ -\frac{1}{2} [\boldsymbol{\beta}_1 - \boldsymbol{\mu}_{\beta_1}]' \boldsymbol{\Sigma}_{\beta_1}^{-1} [\boldsymbol{\beta}_1 - \boldsymbol{\mu}_{\beta_1}] \right\} \end{aligned}$$

Note that

$$\begin{aligned}
& [\mathbf{r}_i - \mathcal{M}_r A_i]' \boldsymbol{\Sigma}_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \\
&= \begin{bmatrix} b_{i0} - \beta'_1 A_i \\ b_{i1} - \beta'_2 A_i \\ v_{i1} - \tau'_1 A_i \\ v_{i2} - \tau'_2 A_i \\ v_{i3} - \tau'_3 A_i \end{bmatrix}' \begin{bmatrix} \sigma_{11}^{-1} & \cdots & \sigma_{15}^{-1} \\ \vdots & \ddots & \vdots \\ \sigma_{51}^{-1} & \cdots & \sigma_{55}^{-1} \end{bmatrix} \begin{bmatrix} b_{i0} - \beta'_1 A_i \\ b_{i1} - \beta'_2 A_i \\ v_{i1} - \tau'_1 A_i \\ v_{i2} - \tau'_2 A_i \\ v_{i3} - \tau'_3 A_i \end{bmatrix} \\
&\propto \sigma_{11}^{-1} (b_{i0} - \beta'_1 A_i)^2 \\
&\quad + 2 \left\{ \sigma_{12}^{-1} (b_{i0} - \beta'_1 A_i) (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^3 \sigma_{1,k+2}^{-1} (b_{i0} - \beta'_1 A_i) (v_{ik} - \tau'_k A_i) \right\} \\
&\propto \beta'_1 \underbrace{[\sigma_{11}^{-1} A_i A_i']}_{V_i} \beta_1 - \underbrace{\left\{ [\sigma_{11}^{-1} b_{i0} + \sigma_{12}^{-1} (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^3 \sigma_{1,k+2}^{-1} (v_{ik} - \tau'_k A_i)] A_i' \right\}}_{W_i} \beta_1 \\
&\quad - \beta'_1 \underbrace{\left\{ A_i [\sigma_{11}^{-1} b_{i0} + \sigma_{12}^{-1} (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^3 \sigma_{1,k+2}^{-1} (v_{ik} - \tau'_k A_i)] \right\}}_{W'_i}.
\end{aligned}$$

Then,

$$\begin{aligned}
& \sum_{i=1}^n [\mathbf{r}_i - \mathcal{M}_r A_i]' \boldsymbol{\Sigma}_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \propto \beta'_1 \left[ \sum_{i=1}^n V_i \right] \beta_1 - \left[ \sum_{i=1}^n W_i \right] \beta_1 - \beta'_1 \left[ \sum_{i=1}^n W'_i \right] \\
p(\beta_1 | \text{else}) &\propto \exp \left\{ -\frac{1}{2} \left[ \beta_1 - \left( \sum_{i=1}^n V_i + \boldsymbol{\Sigma}_{\beta_1}^{-1} \right)^{-1} \underbrace{\left( \sum_{i=1}^n W_i + \boldsymbol{\mu}'_{\beta_1} \boldsymbol{\Sigma}_{\beta_1}^{-1} \right)}_W \right]' \underbrace{\left( \sum_{i=1}^n V_i + \boldsymbol{\Sigma}_{\beta_1}^{-1} \right)}_{V^{-1}} \right. \\
&\quad \left. \times \left[ \beta_1 - \left( \sum_{i=1}^n V_i + \boldsymbol{\Sigma}_{\beta_1}^{-1} \right)^{-1} \left( \sum_{i=1}^n W'_i + \boldsymbol{\mu}'_{\beta_1} \boldsymbol{\Sigma}_{\beta_1}^{-1} \right) \right] \right\}
\end{aligned}$$

Thus,

$$\beta_1 | \text{else} \sim N_2 \left( VW, V \right)$$

where

$$W = \sum_{i=1}^n W'_i + \boldsymbol{\mu}'_{\beta_1} \boldsymbol{\Sigma}_{\beta_1}^{-1} \quad \text{and} \quad V = \left( \sum_{i=1}^n V_i + \boldsymbol{\Sigma}_{\beta_1}^{-1} \right)^{-1}.$$

$\beta_2, \tau_1, \dots, \tau_3$  can be derived in a similar way.

$\Sigma_r^{-1} \mid \text{else}$

The derivation of the posterior distribution of  $\Sigma_r^{-1}$  is the same as that in Section 2 except that  $\boldsymbol{\mu}_{ri}$  replaces  $\boldsymbol{\mu}_r$  in the expression, i.e.

$$\Sigma_r^{-1} \mid \text{else} \sim \text{Wish}_5 \left( \left[ \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_{ri})(\mathbf{r}_i - \boldsymbol{\mu}_{ri})' + \mathbf{Q}^{-1} \right]^{-1}, n + q \right)$$

where, recall that

$$\boldsymbol{\mu}_{ri} = \begin{bmatrix} \mathcal{B}'A_i \\ \mathcal{T}'A_i \end{bmatrix} = \left[ \beta'_1 A_i \quad \beta'_2 A_i \quad \tau'_1 A_i \quad \tau'_2 A_i \quad \tau'_3 A_i \right]'$$

$\mathbf{b}_i \mid \text{else}$

The conditional distribution of  $\mathbf{b}_i$  given  $\mathbf{v}_i$  is slightly different.

$$\mathbf{b}_i \mid \mathbf{v}_i \sim N_2 \left( \boldsymbol{\mu}_{b_i | v_i}^*, \Sigma_{b_i | v_i} \right),$$

where

$$\boldsymbol{\mu}_{b_i | v_i}^* = \mathcal{B}'A_i + \Sigma_{bv} \Sigma_v^{-1} (\mathbf{v}_i - \mathcal{T}'A_i)$$

$$\Sigma_{b_i | v_i} = \Sigma_b - \Sigma_{bv} \Sigma_v^{-1} \Sigma_{bv}'$$

It follows that

$$\mathbf{b}_i \mid \text{else} \sim N_2(\mathbf{GH}^*, \mathbf{G})$$

where

$$\mathbf{G} = \left[ \mathbf{Z}'_i \Sigma_i^{-1} \mathbf{Z}_i + \Sigma_{b_i | v_i}^{-1} \right]^{-1}$$

$$\mathbf{H}^* = \mathbf{Z}'_i \Sigma_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\gamma}) + \Sigma_{b_i | v_i}^{-1} \boldsymbol{\mu}_{b_i | v_i}^*$$

$\mathbf{v}_i$  | else

Following similar arguments of  $\mathbf{b}_i$ , we conclude that

$$\mathbf{v}_i | \mathbf{b}_i \sim N_3(\boldsymbol{\mu}_{v_i | b_i}^*, \boldsymbol{\Sigma}_{v_i | b_i})$$

where

$$\begin{aligned}\boldsymbol{\mu}_{v_i | b_i}^* &= \mathcal{T}' A_i + \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} (\mathbf{b}_i - \mathcal{B}' A_i) \\ \boldsymbol{\Sigma}_{v_i | b_i} &= \boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\Sigma}_{bv}\end{aligned}$$

The posterior distribution of  $\mathbf{v}_i$  can be derived as follows:

$$\begin{aligned}p(\mathbf{v}_i | \text{else}) &\propto p(\mathbf{y}_i | \mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_{r_i}, \boldsymbol{\Sigma}_r) p(\mathbf{v}_i | \mathbf{b}_i, \boldsymbol{\mu}_{r_i}, \boldsymbol{\Sigma}_r) \\ &\propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i | b_i}^*)' \boldsymbol{\Sigma}_{v_i | b_i}^{-1} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i | b_i}^*) \right\}\end{aligned}$$

where  $\boldsymbol{\Sigma}_i$  is a function of  $\mathbf{v}_i$ . Since there is no closed form for this posterior distribution, we adopted adaptive Metropolis algorithm to generate posterior samples.

The posterior distributions of  $\boldsymbol{\gamma}$  and  $\boldsymbol{\lambda}$  are the same as those in Section 2.

## B.4 LINEAR MIXED MODEL WITH INDIVIDUAL VARIANCES

### The Model

$$y_{ijkl} = \mu(f_{ijk}) + b_{i0} + b_{i1}f_{ijk} + e_{ijkl}$$

where

$$\mu(f_{ijk}) = \sum_{s=2}^{S-1} \gamma_s \phi_s(f_{ijk})$$

$$e_{ijkl} \stackrel{iid}{\sim} N(0, \sigma_i^2)$$

$i = 1, 2, \dots, n$ . (Subject index)

$j = 1, 2, \dots, T_i$ . (Time index for Subject  $i$ )

$k = 1, 2, \dots, F_{ij}$ . (Frequency index for Subject  $i$  at Time  $j$ )

$l = 1, 2$ .  $l = 1$  if left ear, and  $l = 2$  if right ear.

Define

$$\mathbf{X}_{ijkl} = (\phi_2(f_{ijk}), \phi_3(f_{ijk}), \dots, \phi_{S-1}(f_{ijk}))$$

$$\mathbf{Z}_{ijkl} = (1, f_{ijk}), \quad \text{if } L_{ijk} = 1$$

$$\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{S-1})'$$

$$\mathbf{b}_i = (b_{i0}, b_{i1})'$$

Then,

$$y_{ijkl} = \mathbf{X}_{ijkl}\boldsymbol{\gamma} + \mathbf{Z}_{ijkl}\mathbf{b}_i + e_{ijkl}$$

Define the total number of observations for subject  $i$  as

$$n_i = \sum_{j=1}^{T_i} \sum_{k=1}^{F_{ij}} L_{ijk}$$



Where

$L_{ijk} = 1$  if only left or right ear was tested

$L_{ijk} = 2$  if both ears were tested

## The Priors

### The Prior of $\mathbf{b}_i$

$$\mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \sim N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma}_b)$$

Further assume

$$\boldsymbol{\beta} \sim N_2(\mathbf{m}, \mathbf{M})$$

$$\boldsymbol{\Sigma}_b^{-1} \sim \text{Wish}_2(\mathbf{Q}, q), q = 2$$

### The Prior of $\sigma_i^{-2}$

$$\sigma_i^{-2} \stackrel{iid}{\sim} \text{Gamma}(a_\sigma, b_\sigma)$$

### The Prior of Penalized B-spline Parameters

Same as Section 2.1.7.

## Full Conditionals

The joint posterior distribution is

$$\begin{aligned} & p(\mathbf{b}, \boldsymbol{\beta}_r, \boldsymbol{\Sigma}_b^{-1}, \sigma^2, \boldsymbol{\gamma}, \lambda | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto p(\mathbf{y} | \mathbf{X}, \mathbf{Z}, \boldsymbol{\gamma}, \mathbf{b}, \sigma^2, \lambda) p(\mathbf{b} | \boldsymbol{\beta}, \boldsymbol{\Sigma}_b^{-1}) p(\boldsymbol{\beta}) p(\boldsymbol{\Sigma}_b^{-1}) p(\boldsymbol{\gamma}) p(\sigma^2) p(\lambda). \end{aligned}$$

$\beta$  | else

$$\begin{aligned}
p(\beta|\text{else}) &\propto p(\mathbf{b}|\beta, \Sigma_b^{-1})p(\beta|\mathbf{m}, \mathbf{M}) \\
&\propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \beta)' \Sigma_b^{-1}(\mathbf{b}_i - \beta)\right\} \exp\left\{-\frac{1}{2}(\beta - \mathbf{m})' \mathbf{M}^{-1}(\beta - \mathbf{m})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\beta - \mathbf{K})' \mathbf{V}^{-1}(\beta - \mathbf{K})\right\}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{V} &= [\mathbf{M}^{-1} + n\Sigma_b^{-1}]^{-1} \\
\mathbf{K} &= \mathbf{M}^{-1}\mathbf{m} + \Sigma_b^{-1} \left[ \sum_{i=1}^n \mathbf{b}_i \right]
\end{aligned}$$

So,

$$\beta|\text{else} \sim N_2(\mathbf{V}\mathbf{K}, \mathbf{V})$$

$\Sigma_b^{-1}$  | else

$$\begin{aligned}
p(\Sigma_b^{-1}|\text{else}) &\propto p(\mathbf{b}|\beta, \Sigma_b^{-1})p(\Sigma_b^{-1}|\mathbf{Q}, q) \\
&\propto \prod_{i=1}^n |\Sigma_b^{-1}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \beta)' \Sigma_b^{-1}(\mathbf{b}_i - \beta)\right\} |\Sigma_b^{-1}|^{\frac{q-2-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{Q}^{-1}\Sigma_b^{-1})\right\} \\
&\propto |\Sigma_b^{-1}|^{\frac{(n+q)-2-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[\left(\sum_{i=1}^n (\mathbf{b}_i - \beta)(\mathbf{b}_i - \beta)' + \mathbf{Q}^{-1}\right)\Sigma_b^{-1}\right]\right\}
\end{aligned}$$

Thus,

$$\Sigma_b^{-1}|\text{else} \sim \text{Wish}_2\left(\left[\sum_{i=1}^n (\mathbf{b}_i - \beta)(\mathbf{b}_i - \beta)' + \mathbf{Q}^{-1}\right]^{-1}, n + q\right)$$

Note that, if we let

$$\mathbf{E} = (\mathbf{b}_1 - \beta, \dots, \mathbf{b}_n - \beta)_{2 \times n}$$

then

$$\sum_{i=1}^n (\mathbf{b}_i - \beta)(\mathbf{b}_i - \beta)' = \mathbf{E}\mathbf{E}'$$

$\mathbf{b}_i$  | else

Let

$$\mathbf{y}_i = (\mathbf{y}'_{i1}, \dots, \mathbf{y}'_{iT_i})'$$

where  $\mathbf{y}_{ij} = (\mathbf{y}'_{ij1}, \dots, \mathbf{y}'_{ijF_{ij}})'$  and  $\mathbf{y}_{ijk} = (y_{ijk1}, y_{ijk2})'$  if  $L_{ijk} = 2$  and  $\mathbf{y}_{ijk} = (y_{ijk1})$  otherwise. Define

$$\mathbf{X}_{ijk} = \mathbf{X}_{ijkl}, \text{ if } L_{ijk} = 1$$

$$\mathbf{X}_{ijk} = \begin{bmatrix} \phi_2(f_{ijk1}), \phi_3(f_{ijk1}), \dots, \phi_{S-1}(f_{ijk1}) \\ \phi_2(f_{ijk2}), \phi_3(f_{ijk2}), \dots, \phi_{S-1}(f_{ijk2}) \end{bmatrix}, \text{ if } L_{ijk} = 2$$

Let

$$\mathbf{X}_{ij} = \begin{bmatrix} \mathbf{X}_{ij1} \\ \mathbf{X}_{ij2} \\ \vdots \\ \mathbf{X}_{ijF_{ij}} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \mathbf{X}_{i2} \\ \vdots \\ \mathbf{X}_{iT_i} \end{bmatrix}$$

Also define

$$\mathbf{Z}_{ijk} = (1, f_{ijkl}), \text{ if } L_{ijk} = 1$$

$$\mathbf{Z}_{ijk} = \begin{bmatrix} 1 & f_{ijk1} \\ 1 & f_{ijk2} \end{bmatrix}, \text{ if } L_{ijk} = 2$$

$$\mathbf{Z}_{ij} = \begin{bmatrix} \mathbf{Z}_{ij1} \\ \mathbf{Z}_{ij2} \\ \vdots \\ \mathbf{Z}_{ijF_{ij}} \end{bmatrix}, \quad \mathbf{Z}_i = \begin{bmatrix} \mathbf{Z}_{i1} \\ \mathbf{Z}_{i2} \\ \vdots \\ \mathbf{Z}_{iT_i} \end{bmatrix}$$

Then,

$$\begin{aligned}
p(\mathbf{b}_i|\text{else}) &\propto p(\mathbf{y}_i|\mathbf{b}_i)p(\mathbf{b}_i|\boldsymbol{\beta}, \boldsymbol{\Sigma}_b) \\
&\propto \exp\left\{-\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]'\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]\right\} \\
&\quad \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \boldsymbol{\beta})'\boldsymbol{\Sigma}_b^{-1}(\mathbf{b}_i - \boldsymbol{\beta})\right\} \\
&\propto \exp\left\{-\frac{1}{2}(\mathbf{b}_i - \mathbf{GH})'\mathbf{G}^{-1}(\mathbf{b}_i - \mathbf{GH})\right\}
\end{aligned}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_i &= \sigma_i^2 \mathbf{I}_{n_i}, n_i = \sum_{j=1}^{T_i} \sum_{k=1}^{F_{ij}} L_{ijk} \\
\mathbf{G} &= [\mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i + \boldsymbol{\Sigma}_b^{-1}]^{-1} \\
\mathbf{H} &= \mathbf{Z}_i'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\gamma}) + \boldsymbol{\Sigma}_b^{-1}\boldsymbol{\beta}
\end{aligned}$$

Thus,

$$\mathbf{b}_i|\text{else} \sim N_2(\mathbf{GH}, \mathbf{G})$$

$\sigma_i^{-2}$  | else

$$\begin{aligned}
p(\sigma_i^{-2}|\text{else}) &\propto p(\mathbf{y}_i|\sigma_i^{-2})p(\sigma_i^{-2}) \\
&\propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]'\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]\right\} \\
&\quad \times (\sigma_i^{-2})^{a_\sigma-1} \exp(-b_\sigma\sigma_i^{-2}) \\
&\propto (\sigma_i^{-2})^{\frac{n_i}{2}+a_\sigma-1} \exp\left\{-\left[\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]'[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)] + b_\sigma\right]\sigma_i^{-2}\right\}
\end{aligned}$$

Therefore,

$$\sigma_i^{-2} \sim \text{Gamma}\left(a_\sigma + \frac{n_i}{2}, b_\sigma + \frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]'[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]\right)$$

The posterior distributions of  $\boldsymbol{\gamma}$  and  $\boldsymbol{\lambda}$  are the same as those in Section 2.

## B.5 LINEAR MIXED MODEL WITH OVERALL COMMON VARIANCE

This model is almost the same to the previous model except that the variance of the error term is the same for all measurements, i.e.

$$e_{ijkl} \stackrel{iid}{\sim} N(0, \sigma^2)$$

Consequently, the Prior of  $\sigma^2$  becomes

$$\sigma^{-2} \stackrel{iid}{\sim} \text{Gamma}(a_\sigma, b_\sigma)$$

The posterior distribution of  $\sigma^2$  can be derived as follows:

$$\begin{aligned} & p(\sigma^{-2} | \text{else}) \\ & \propto \prod_{i=1}^n p(\mathbf{y}_i | \sigma^{-2}) p(\sigma^{-2}) \\ & \propto \prod_{i=1}^n |\Sigma_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \Sigma_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right\} \\ & \quad \times (\sigma^{-2})^{a_\sigma - 1} \exp(-b_\sigma \sigma^{-2}) \\ & \propto (\sigma^{-2})^{\frac{\sum_{i=1}^n n_i}{2} + a_\sigma - 1} \exp \left\{ -\left[ \frac{1}{2} \sum_{i=1}^n [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] + b_\sigma \right] \sigma^{-2} \right\} \end{aligned}$$

Therefore,

$$\sigma^{-2} \sim \text{Gamma} \left( a_\sigma + \frac{1}{2} \sum_{i=1}^n n_i, b_\sigma + \frac{1}{2} \sum_{i=1}^n [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right)$$

## B.6 HIERARCHICAL MODEL WITH CORRELATION AMONG F2

In this model, we assume there is correlation among frequency levels. So,  $\theta_{if} > 0$  but it does not tend to infinity. We use the same index system as the previous two linear mixed models. i.e.

$$y_{ijkl} = \mu(f_{ijkl}) + b_{i0} + b_{i1} f_{ijkl} + e_{ijkl}$$

where

$i = 1, 2, \dots, n$ . (Subject index)

$j = 1, 2, \dots, T_i$ . (Time index)

$k = 1, 2, \dots, F_{ij}$ . (Frequency level index)

$l = 1, 2$ .  $l = 1$  if left ear, and  $l = 2$  if right ear.

The  $\mathbf{r}$  vector is 6-dimensional, with  $\log(\theta_{if})$  in it, i.e.

$$\mathbf{r}_i = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{v}_i \end{bmatrix} \quad \text{where } \mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \quad \text{and } \mathbf{v}_i = \begin{bmatrix} \log(\sigma_i^2) \\ \log(\theta_{it}) \\ \log(\theta_{if}) \\ \log(\theta_{ie}) \end{bmatrix}$$

$\mathbf{r}_1, \dots, \mathbf{r}_n \sim N_6(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$  where

$$\boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}, \quad \boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}$$

The posterior density  $p(\mathbf{r}, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}, \boldsymbol{\gamma}, \lambda | \mathbf{y})$  augmented with  $\mathbf{r} = [\mathbf{r}_1 \dots \mathbf{r}_n]$  is proportional to  $p(\mathbf{y} | \boldsymbol{\gamma}, \mathbf{r}, \lambda) p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\mu}_r | \mathbf{m}, \mathbf{M}) p(\boldsymbol{\Sigma}_r^{-1} | \mathbf{Q}, q) p(\boldsymbol{\gamma} | \boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma) p(\lambda | \alpha, \beta)$ .

$\boldsymbol{\mu}_r | \text{else}$

$$\begin{aligned} p(\boldsymbol{\mu}_r | \text{else}) &\propto p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\mu}_r | \mathbf{m}, \mathbf{M}) \\ &\propto \prod_{i=1}^n \exp\left\{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\mathbf{r}_i - \boldsymbol{\mu}_r)\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_r - \mathbf{m})' \mathbf{M}^{-1} (\boldsymbol{\mu}_r - \mathbf{m})\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_r - \mathbf{KV})' \mathbf{v}^{-1} (\boldsymbol{\mu}_r - \mathbf{KV})\right\} \end{aligned}$$

where

$$\begin{aligned}\mathbf{v} &= [\mathbf{M}^{-1} + n\boldsymbol{\Sigma}_r^{-1}]^{-1} \\ \mathbf{K} &= \mathbf{M}^{-1}\mathbf{m} + \boldsymbol{\Sigma}_r^{-1} \left[ \sum_{i=1}^n \mathbf{r}_i \right]\end{aligned}$$

So,

$$\boldsymbol{\mu}_r | \text{else} \sim N_6(\mathbf{V}\mathbf{K}, \mathbf{v})$$

$\boldsymbol{\Sigma}_r^{-1} | \text{else}$

$$\begin{aligned}p(\boldsymbol{\Sigma}_r^{-1} | \text{else}) &\propto p(\mathbf{r} | \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r^{-1}) p(\boldsymbol{\Sigma}_r^{-1} | \mathbf{Q}, q) \\ &\propto \prod_{i=1}^n |\boldsymbol{\Sigma}_r^{-1}|^{\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)' \boldsymbol{\Sigma}_r^{-1} (\mathbf{r}_i - \boldsymbol{\mu}_r) \right\} |\boldsymbol{\Sigma}_r^{-1}|^{\frac{q-6-1}{2}} \exp\left\{ -\frac{1}{2} \text{tr}(\mathbf{Q}^{-1} \boldsymbol{\Sigma}_r^{-1}) \right\} \\ &\propto |\boldsymbol{\Sigma}_r^{-1}|^{\frac{(n+q)-6-1}{2}} \exp\left\{ -\frac{1}{2} \text{tr} \left[ \left( \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' + \mathbf{Q}^{-1} \right) \boldsymbol{\Sigma}_r^{-1} \right] \right\}\end{aligned}$$

Thus,

$$\boldsymbol{\Sigma}_r^{-1} | \text{else} \sim \text{Wish}_6 \left( \left[ \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' + \mathbf{Q}^{-1} \right]^{-1}, n + q \right)$$

Note that, if we let

$$\mathbf{e} = (\mathbf{r}_1 - \boldsymbol{\mu}_r, \dots, \mathbf{r}_n - \boldsymbol{\mu}_r)_{6 \times n}.$$

Then,

$$\sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_r)(\mathbf{r}_i - \boldsymbol{\mu}_r)' = \mathbf{E}\mathbf{E}'$$

$\mathbf{b}_i | \text{else}$

First, we need to derive the form of  $\boldsymbol{\Sigma}_r^{-1}$  in terms of  $\boldsymbol{\Sigma}_b$ ,  $\boldsymbol{\Sigma}_v$ , and  $\boldsymbol{\Sigma}_{bv}$ . Recall

$$\boldsymbol{\Sigma}_r = \begin{bmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\Sigma}_{bv} \\ \boldsymbol{\Sigma}'_{bv} & \boldsymbol{\Sigma}_v \end{bmatrix}$$

then,

$$\begin{aligned}\Sigma_r^{-1} &= \begin{bmatrix} (\Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv})^{-1} & -(\Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv})^{-1}\Sigma_{bv}\Sigma_v^{-1} \\ -(\Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv})^{-1}\Sigma'_{bv}\Sigma_b^{-1} & (\Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv})^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (\Sigma_b^*)^{-1} & -(\Sigma_b^*)^{-1}\Sigma_{bv}\Sigma_v^{-1} \\ -(\Sigma_v^*)^{-1}\Sigma'_{bv}\Sigma_b^{-1} & (\Sigma_v^*)^{-1} \end{bmatrix}\end{aligned}$$

where

$$\Sigma_b^* = \Sigma_b - \Sigma_{bv}\Sigma_v^{-1}\Sigma'_{bv}$$

$$\Sigma_v^* = \Sigma_v - \Sigma'_{bv}\Sigma_b^{-1}\Sigma_{bv}$$

Second, we will find the posterior distribution of  $\mathbf{b}_i$  given  $\mathbf{v}_i$ . Recall that

$$\mathbf{r}_i = \begin{bmatrix} \mathbf{b}_i \\ \mathbf{v}_i \end{bmatrix} \quad \mathbf{b}_i = \begin{bmatrix} b_{i0} \\ b_{i1} \end{bmatrix} \quad \mathbf{v}_i = \begin{bmatrix} \log(\sigma_i^2) \\ \log(\theta_{it}) \\ \log(\theta_{if}) \\ \log(\theta_{ie}) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix}$$

So,

$$\begin{aligned}p(\mathbf{b}_i|\mathbf{v}_i, \boldsymbol{\mu}_r, \Sigma_r^{-1}) &\propto p(\mathbf{b}_i, \mathbf{v}_i|\boldsymbol{\mu}_r, \Sigma_r^{-1}) \\ &\propto \exp\left\{-\frac{1}{2}(\mathbf{r}_i - \boldsymbol{\mu}_r)'\Sigma_r^{-1}(\mathbf{r}_i - \boldsymbol{\mu}_r)\right\} \\ &= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{b}_i - \boldsymbol{\beta} \\ \mathbf{v}_i - \boldsymbol{\tau} \end{bmatrix}'\Sigma_r^{-1}\begin{bmatrix} \mathbf{b}_i - \boldsymbol{\beta} \\ \mathbf{v}_i - \boldsymbol{\tau} \end{bmatrix}\right\} \\ &= \exp\left\{-\frac{1}{2}\begin{bmatrix} \mathbf{b}_i^* \\ \mathbf{v}_i^* \end{bmatrix}'\begin{bmatrix} (\Sigma_b^*)^{-1} & -(\Sigma_b^*)^{-1}\Sigma_{bv}\Sigma_v^{-1} \\ -(\Sigma_v^*)^{-1}\Sigma'_{bv}\Sigma_b^{-1} & (\Sigma_v^*)^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{b}_i^* \\ \mathbf{v}_i^* \end{bmatrix}\right\} \\ &\propto (\mathbf{b}_i^* - \Sigma_r^*\mathbf{W})'(\Sigma_b^*)^{-1}(\mathbf{b}_i^* - \Sigma_r^*\mathbf{W})\end{aligned}$$



where

$$\mathbf{W} = (\boldsymbol{\Sigma}_b^*)^{-1} \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} \mathbf{v}^*$$

Therefore,

$$\mathbf{b}_i^* | \mathbf{v}_i^* \sim N_2(\boldsymbol{\Sigma}_b^* \mathbf{W}, \boldsymbol{\Sigma}_b^*)$$

and,

$$\mathbf{b}_i | \mathbf{v}_i \sim N_2(\boldsymbol{\mu}_{b_i | v_i}, \boldsymbol{\Sigma}_{b_i | v_i})$$

where

$$\begin{aligned} \boldsymbol{\mu}_{b_i | v_i} &= \boldsymbol{\beta} + \boldsymbol{\Sigma}_b^* \mathbf{W} = \boldsymbol{\beta} + \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} (\mathbf{v}_i - \boldsymbol{\tau}) \\ \boldsymbol{\Sigma}_{b_i | v_i} &= \boldsymbol{\Sigma}_b^* = \boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{bv} \boldsymbol{\Sigma}_v^{-1} \boldsymbol{\Sigma}'_{bv} \end{aligned}$$

Now we are ready to find the posterior of  $\mathbf{b}_i$ :

$$\begin{aligned} p(\mathbf{b}_i | \text{else}) &\propto p(\mathbf{y}_i | \mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) p(\mathbf{b}_i | \mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\ &\propto \exp\left\{ -\frac{1}{2} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right\} \\ &\quad \exp\left\{ -\frac{1}{2} (\mathbf{b}_i - \boldsymbol{\mu}_{b_i | v_i})' \boldsymbol{\Sigma}_{b_i | v_i}^{-1} (\mathbf{b}_i - \boldsymbol{\mu}_{b_i | v_i}) \right\} \\ &\propto \exp\left\{ -\frac{1}{2} (\mathbf{b}_i - \mathbf{GH})' \mathbf{G}^{-1} (\mathbf{b}_i - \mathbf{GH}) \right\} \end{aligned}$$

where

$$\begin{aligned} \mathbf{G} &= [\mathbf{Z}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{Z}_i + \boldsymbol{\Sigma}_{b_i | v_i}^{-1}]^{-1} \\ \mathbf{H} &= \mathbf{Z}_i' \boldsymbol{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\gamma}) + \boldsymbol{\Sigma}_{b_i | v_i}^{-1} \boldsymbol{\mu}_{b_i | v_i} \end{aligned}$$

Thus,

$$\mathbf{b}_i | \text{else} \sim N_2(\mathbf{GH}, \mathbf{G})$$

$\mathbf{v}_i | \text{else}$

Following similar arguments of  $\mathbf{b}_i$ , we conclude that

$$\mathbf{v}_i | \mathbf{b}_i \sim N_4(\boldsymbol{\mu}_{v_i | b_i}, \boldsymbol{\Sigma}_{v_i | b_i})$$

where

$$\boldsymbol{\mu}_{v_i | b_i} = \boldsymbol{\tau} + \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} (\mathbf{b}_i - \boldsymbol{\beta})$$

$$\boldsymbol{\Sigma}_{v_i | b_i} = \boldsymbol{\Sigma}_r - \boldsymbol{\Sigma}'_{bv} \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\Sigma}_b v$$

The posterior distribution of  $\mathbf{v}_i$  can be derived as follows:

$$\begin{aligned} p(\mathbf{v}_i | \text{else}) &\propto p(\mathbf{y}_i | \mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) p(\mathbf{v}_i | \mathbf{b}_i, \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\ &\propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)]' \boldsymbol{\Sigma}_i^{-1} [\mathbf{y}_i - (\mathbf{X}_i \boldsymbol{\gamma} + \mathbf{Z}_i \mathbf{b}_i)] \right\} \\ &\quad \times \exp\left\{ -\frac{1}{2} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i | b_i})' \boldsymbol{\Sigma}_{v_i | b_i}^{-1} (\mathbf{v}_i - \boldsymbol{\mu}_{v_i | b_i}) \right\} \end{aligned}$$

where  $\boldsymbol{\Sigma}_i$  is a function of  $\mathbf{v}_i$ .

There is no close form for this posterior distribution. We adopt adaptive M-H algorithm to generate samples.

The posterior distributions of  $\boldsymbol{\gamma}$  and  $\boldsymbol{\lambda}$  are the same as those in Section 2.

## B.7 AGE-GENDER MODEL WITH CORRELATION AMONG F2

Instead of assuming

$$\mathbf{r}_1, \dots, \mathbf{r}_n \sim N_6(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r), \text{ where } \boldsymbol{\mu}_r = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\tau} \end{bmatrix},$$

we can let  $\boldsymbol{\mu}_r$  be dependent on age and gender of the patient. In particular, let

$$\mathcal{M}_r = \begin{bmatrix} \mathcal{B}' \\ \mathcal{T}' \end{bmatrix} \quad \text{and} \quad \boldsymbol{\mu}_{r_i} = \mathcal{M}_r A_i = \begin{bmatrix} \mathcal{B}' A_i \\ \mathcal{T}' A_i \end{bmatrix}$$

where

$$\mathcal{B}' = \begin{bmatrix} \beta_{10} & \beta_{11} & \beta_{12} \\ \beta_{20} & \beta_{21} & \beta_{22} \end{bmatrix} = \begin{bmatrix} \beta'_1 \\ \beta'_2 \end{bmatrix} \quad \mathcal{T}' = \begin{bmatrix} \tau_{10} & \tau_{11} & \tau_{12} \\ \tau_{20} & \tau_{21} & \tau_{22} \\ \tau_{30} & \tau_{31} & \tau_{32} \\ \tau_{40} & \tau_{41} & \tau_{42} \end{bmatrix} = \begin{bmatrix} \tau'_1 \\ \tau'_2 \\ \tau'_3 \\ \tau'_4 \end{bmatrix} \quad A_i = \begin{bmatrix} 1 \\ a_i \\ g_i \end{bmatrix}$$

The joint posterior distribution of this age-gender specific model becomes:

$$\begin{aligned} & p(\mathbf{r}, \beta_1, \beta_2, \tau_1, \tau_2, \tau_3, \tau_4, \Sigma_r^{-1}, \gamma, \lambda | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}) \\ & \propto p(\mathbf{y} | \mathbf{X}, \mathbf{Z}, \gamma, \mathbf{r}, \lambda) p(\mathbf{r} | \beta_1, \beta_2, \tau_1, \tau_2, \tau_3, \tau_4, \Sigma_r^{-1}) \prod_{k=1}^2 p(\beta_k | \boldsymbol{\mu}_{\beta_k}, \Sigma_{\beta_k}) \prod_{k=1}^4 p(\tau_k | \boldsymbol{\mu}_{\tau_k}, \Sigma_{\tau_k}) \\ & p(\Sigma_r^{-1} | \mathbf{Q}, q) p(\gamma | \boldsymbol{\mu}_\gamma, \Sigma_\gamma) p(\lambda | \alpha, \beta). \end{aligned}$$

$\beta_1$  | else

$p(\beta_1 | \text{else})$

$$\begin{aligned}
& \propto p(\mathbf{r} | \beta_1, \beta_2, \tau_1, \tau_2, \tau_3, \tau_4, \Sigma_r^{-1}) p(\beta_1 | \mu_{\beta_1}, \Sigma_{\beta_1}) \\
& \propto \prod_{i=1}^n p(\mathbf{r}_i | \mu_{r_i}, \Sigma_r) p(\beta_1 | \mu_{\beta_1}, \Sigma_{\beta_1}) \\
& \propto \prod_{i=1}^n \exp \left\{ -\frac{1}{2} [\mathbf{r}_i - \mathcal{M}_r A_i]' \Sigma_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \right\} \\
& \quad \times \exp \left\{ -\frac{1}{2} [\beta_1 - \mu_{\beta_1}]' \Sigma_{\beta_1}^{-1} [\beta_1 - \mu_{\beta_1}] \right\} \\
& \quad [\mathbf{r}_i - \mathcal{M}_r A_i]' \Sigma_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \\
& = \begin{bmatrix} b_{i0} - \beta'_1 A_i \\ b_{i1} - \beta'_2 A_i \\ v_{i1} - \tau'_1 A_i \\ v_{i2} - \tau'_2 A_i \\ v_{i3} - \tau'_3 A_i \\ v_{i4} - \tau'_4 A_i \end{bmatrix}' \begin{bmatrix} \sigma_{11}^{-1} & \cdots & \sigma_{16}^{-1} \\ \vdots & \ddots & \vdots \\ \sigma_{61}^{-1} & \cdots & \sigma_{66}^{-1} \end{bmatrix} \begin{bmatrix} b_{i0} - \beta'_1 A_i \\ b_{i1} - \beta'_2 A_i \\ v_{i1} - \tau'_1 A_i \\ v_{i2} - \tau'_2 A_i \\ v_{i3} - \tau'_3 A_i \\ v_{i4} - \tau'_4 A_i \end{bmatrix} \\
& \propto \sigma_{11}^{-1} (b_{i0} - \beta'_1 A_i)^2 \\
& \quad + 2 \left\{ \sigma_{12}^{-1} (b_{i0} - \beta'_1 A_i) (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^4 \sigma_{1,k+2}^{-1} (b_{i0} - \beta'_1 A_i) (v_{ik} - \tau'_k A_i) \right\} \\
& \propto \beta'_1 \underbrace{[\sigma_{11}^{-1} A_i A_i']}_{V_i} \beta_1 - \underbrace{\left\{ [\sigma_{11}^{-1} b_{i0} + \sigma_{12}^{-1} (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^4 \sigma_{1,k+2}^{-1} (v_{ik} - \tau'_k A_i)] A_i' \right\}}_{W_i} \beta_1 \\
& \quad - \beta'_1 \underbrace{\left\{ A_i [\sigma_{11}^{-1} b_{i0} + \sigma_{12}^{-1} (b_{i1} - \beta'_2 A_i) + \sum_{k=1}^4 \sigma_{1,k+2}^{-1} (v_{ik} - \tau'_k A_i)] \right\}}_{W'_i}.
\end{aligned}$$

Then,

$$\begin{aligned}
& \sum_{i=1}^n [\mathbf{r}_i - \mathcal{M}_r A_i]' \Sigma_r^{-1} [\mathbf{r}_i - \mathcal{M}_r A_i] \propto \beta_1' \left[ \sum_{i=1}^n V_i \right] \beta_1 - \left[ \sum_{i=1}^n W_i \right] \beta_1 - \beta_1' \left[ \sum_{i=1}^n W_i' \right] \\
p(\beta_1 | \text{else}) & \propto \exp \left\{ -\frac{1}{2} \left[ \beta_1 - \underbrace{\left( \sum_{i=1}^n V_i + \Sigma_{\beta_1}^{-1} \right)^{-1} \left( \sum_{i=1}^n W_i' + \boldsymbol{\mu}'_{\beta_1} \Sigma_{\beta_1}^{-1} \right)}_W \right]' \underbrace{\left( \sum_{i=1}^n V_i + \Sigma_{\beta_1}^{-1} \right)}_{V^{-1}} \right. \\
& \quad \left. \left[ \beta_1 - \left( \sum_{i=1}^n V_i + \Sigma_{\beta_1}^{-1} \right)^{-1} \left( \sum_{i=1}^n W_i' + \boldsymbol{\mu}'_{\beta_1} \Sigma_{\beta_1}^{-1} \right) \right] \right\}
\end{aligned}$$

Thus,

$$\beta_1 | \text{else} \sim N_2 \left( VW, V \right)$$

where

$$W = \sum_{i=1}^n W_i' + \boldsymbol{\mu}'_{\beta_1} \Sigma_{\beta_1}^{-1} \quad \text{and} \quad V = \left( \sum_{i=1}^n V_i + \Sigma_{\beta_1}^{-1} \right)^{-1}$$

$\beta_2, \boldsymbol{\tau}_1, \dots, \boldsymbol{\tau}_4$  can be derived in a similar way.

$\Sigma_r^{-1} | \text{else}$

The derivation of the posterior distribution of  $\Sigma_r^{-1}$  is the same as that in Section 2 except that  $\boldsymbol{\mu}_{ri}$  replaces  $\boldsymbol{\mu}_r$  in the expression, i.e.

$$\Sigma_r^{-1} | \text{else} \sim \text{Wish}_6 \left( \left[ \sum_{i=1}^n (\mathbf{r}_i - \boldsymbol{\mu}_{ri})(\mathbf{r}_i - \boldsymbol{\mu}_{ri})' + \mathbf{Q}^{-1} \right]^{-1}, n + q \right)$$

where, recall that

$$\boldsymbol{\mu}_{ri} = \begin{bmatrix} \mathcal{B}' A_i \\ \mathcal{T}' A_i \end{bmatrix} = \left[ \beta_1' A_i \quad \beta_2' A_i \quad \boldsymbol{\tau}_1' A_i \quad \boldsymbol{\tau}_2' A_i \quad \boldsymbol{\tau}_3' A_i \quad \boldsymbol{\tau}_4' A_i \right]'$$

$\mathbf{b}_i | \text{else}$

The conditional distribution of  $\mathbf{b}_i$  given  $\mathbf{v}_i$  is slightly different.

$$\mathbf{b}_i | \mathbf{v}_i \sim N_2 \left( \boldsymbol{\mu}_{b_i | v_i}^*, \Sigma_{b_i | v_i} \right),$$

where

$$\begin{aligned}\boldsymbol{\mu}_{b_i|v_i}^* &= \mathcal{B}'A_i + \boldsymbol{\Sigma}_{bv}\boldsymbol{\Sigma}_v^{-1}(\mathbf{v}_i - \mathcal{T}'A_i) \\ \boldsymbol{\Sigma}_{b_i|v_i} &= \boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_{bv}\boldsymbol{\Sigma}_v^{-1}\boldsymbol{\Sigma}'_{bv}\end{aligned}$$

It follows that

$$\mathbf{b}_i|\text{else} \sim N_2(\mathbf{GH}^*, \mathbf{G})$$

where

$$\begin{aligned}\mathbf{G} &= [\mathbf{Z}'_i\boldsymbol{\Sigma}_i^{-1}\mathbf{Z}_i + \boldsymbol{\Sigma}_{b_i|v_i}^{-1}]^{-1} \\ \mathbf{H}^* &= \mathbf{Z}'_i\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\gamma}) + \boldsymbol{\Sigma}_{b_i|v_i}^{-1}\boldsymbol{\mu}_{b_i|v_i}^*\end{aligned}$$

$\mathbf{v}_i$  | else

Following similar arguments of  $\mathbf{b}_i$ , we conclude that

$$\mathbf{v}_i|\mathbf{b}_i \sim N_4(\boldsymbol{\mu}_{v_i|b_i}^*, \boldsymbol{\Sigma}_{v_i|b_i})$$

where

$$\begin{aligned}\boldsymbol{\mu}_{v_i|b_i}^* &= \mathcal{T}'A_i + \boldsymbol{\Sigma}'_{bv}\boldsymbol{\Sigma}_b^{-1}(\mathbf{b}_i - \mathcal{B}'A_i) \\ \boldsymbol{\Sigma}_{v_i|b_i} &= \boldsymbol{\Sigma}_v - \boldsymbol{\Sigma}'_{bv}\boldsymbol{\Sigma}_b^{-1}\boldsymbol{\Sigma}_{bv}\end{aligned}$$

The posterior distribution of  $\mathbf{v}_i$  can be derived as follows:

$$\begin{aligned}p(\mathbf{v}_i|\text{else}) &\propto p(\mathbf{y}_i|\mathbf{b}_i, \mathbf{v}_i, \boldsymbol{\mu}_{ri}, \boldsymbol{\Sigma}_r)p(\mathbf{v}_i|\mathbf{b}_i, \boldsymbol{\mu}_{ri}, \boldsymbol{\Sigma}_r) \\ &\propto |\boldsymbol{\Sigma}_i|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]'\boldsymbol{\Sigma}_i^{-1}[\mathbf{y}_i - (\mathbf{X}_i\boldsymbol{\gamma} + \mathbf{Z}_i\mathbf{b}_i)]\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{v}_i - \boldsymbol{\mu}_{v_i|b_i}^*)'\boldsymbol{\Sigma}_{v_i|b_i}^{-1}(\mathbf{v}_i - \boldsymbol{\mu}_{v_i|b_i}^*)\right\}\end{aligned}$$

where  $\boldsymbol{\Sigma}_i$  is a function of  $\mathbf{v}_i$ .

Since there is no closed form for this posterior distribution, we adopted adaptive Metropolis algorithm to generate samples. The posterior distributions of  $\boldsymbol{\gamma}$  and  $\boldsymbol{\lambda}$  are the same as those in Section 2.

## B.8 SAMPLE FORTRAN 90 CODE

```
!Growth curve model, hearing data (both ears)
!Hierarchical model
program gcb2
  use numerical_libraries
  use module_1
  use module_2
  implicit none

  integer,parameter::rd=20000 ! number of iterates
  integer,parameter::thin=1
  integer,parameter::rd0=rd*0.1 ! number of iterates before burn in
  integer,parameter::rd1=rd*0.2 ! number of iterates before adaption
  integer,parameter::size=4148 ! number of observations in data set
  integer,parameter::n=38 ! number of subjects
  integer,parameter::nn=456 ! number of subject*f2 combinations
  integer,parameter::S=20 ! number of knots in b-spline
  double precision,parameter::c1=dbl(1),c2=dbl(0.0005)
  ! lambda-Gamma(c1,c2), Lambda is the penalty coefficient
  double precision,parameter::eps=0.0000001,eps2=0.001
  double precision,dimension(S-2,S-2)::D2
  double precision,dimension(S-2,S-2)::DD

  double precision,parameter::p=0.5 !proportion of tail chains
  !*****

  integer,parameter::ss=5
  double precision,dimension(ss,1)::m0 !mean of mu_r
  double precision,dimension(ss,ss)::MM0 !covariance matrix of mu_r
  double precision,dimension(ss,ss)::Q !sig_r~Iwish(Q^(-1), q)
  integer::qq=ss
  double precision,dimension(S-2,1)::mu_g !mean of gamma
  double precision,dimension(S-2,S-2)::sig_g !covariance matrix of gamma
  double precision,dimension(S-2,S-2)::prec_g !precision matrix of gamma

  double precision,dimension(ss,1)::r0 !mean of r
  double precision,dimension(ss,ss)::sig_r0 !covariance of r
  double precision,dimension(ss+1,1)::r00 !mean of r
  double precision,dimension(ss+1,ss+1)::sig_r00 !covariance of r
  !*****
  double precision,dimension(ss*rd,1)::mu_r !mean vector of r
  double precision,dimension(ss*rd,ss)::prec_r !precision of r
  double precision,dimension(ss,ss)::sig_r !covariance of r
  double precision,dimension(2,1)::mu_b !mean of b, b=(b0,b1)
  double precision,dimension(3,1)::mu_v !mean of v
```

```

!*****
double precision,dimension(rd*2,n)::b
double precision,dimension(rd*2*n,1)::b_chain
double precision,dimension(rd*3,n)::v
double precision,dimension(rd*3*n,1)::v_chain
double precision,dimension(rd*ss,n)::r
double precision,dimension(rd*ss*n,1)::r_chain
double precision,dimension((S-2)*rd,1)::gam
double precision,dimension(rd)::lamb

!*****
integer,parameter::m=31 !number of distinct frequencies for prediction
integer,parameter::m1=6
double precision,dimension(m,S-1)::XXp
double precision,dimension(m,S-2)::Xp !Design matrix of spline part
double precision,dimension(m,2)::Zp !Design matrix of linear part
double precision,dimension(m)::fp !Distinct values of f2
double precision,dimension(rd*p/thin*m,1)::Yp
double precision,dimension(rd*p/thin*m,m1)::delta_p
double precision,dimension(m,1)::mu_yp,mu_dp
double precision,dimension(m,m)::sig_yp,sig_dp
double precision,dimension(m1)::delta_t !followup time

!***** counting integers *****
integer::it,i,jj,j1,j2,k1,k2,l,k,ii

!*****
double precision,dimension(ss,ss)::VV,Minv
double precision,dimension(ss,n)::E
double precision,dimension(ss,1)::D,D0

double precision,dimension(3)::vi,vip,vii

double precision,dimension(:,),allocatable::prec_i
double precision,dimension(:,),allocatable::sig_i
double precision,dimension(:,),allocatable::sig_ii
double precision,dimension(S-2,1)::J
double precision,dimension(2,1)::bb
double precision,dimension(S-2,S-2)::W

double precision,dimension(2,2)::sig_b
double precision,dimension(3,3)::sig_v
double precision,dimension(2,3)::sig_bv
double precision,dimension(3,2)::sig_vb
double precision,dimension(2,1)::m_bv
double precision,dimension(2,2)::s_bv

```



```

double precision,dimension(3,1)::m_vb
double precision,dimension(3,3)::s_vb

double precision,dimension(2,2)::G
double precision,dimension(2,1)::H
double precision,dimension(ss,1)::sum_r
double precision::u,a
double precision,dimension(:,:),allocatable::mu_y,mu_y0

double precision,dimension(n)::ct
double precision,dimension(3,3*n)::vmat
double precision,dimension(3,n)::vsum
double precision,dimension(3,3)::sv

double precision,dimension(S-2,1)::gam_it
double precision::GDG
double precision,dimension(S-2,1)::DG
integer,dimension(nn)::obs_ij
integer,dimension(nn+1)::cobs_ij
integer,dimension(n)::obs
integer,dimension(n+1)::cobs
integer,dimension(n,13)::cobs_mat
double precision,dimension(S-2,S-2)::XPX
double precision,dimension(S-2,1)::XPY
!*****
!***** data *****
double precision,dimension(size,11)::both
double precision,dimension(size,S-2)::X
double precision,dimension(size,1)::Y
double precision,dimension(size,2)::Z
double precision,dimension(:,:),allocatable::Xi,Yi,Zi
double precision,dimension(size)::dt,df,de
double precision,dimension(:),allocatable::dti,eari
!*****
double precision,dimension(size,1)::mu_spl
double precision::mu,tau

double precision,dimension(size)::sump
double precision::pinv
!*****
open(unit=1,file="C:\Growth Curve Hearing\data\both_new.txt")
  do i=1,size
    read(1,*)both(i,:)
  end do
close(unit=1)

open(unit=2,file="C:\Growth Curve Hearing\data\X_both.txt")

```

```

do i=1,size
    read(2,*)X(i,:)
end do
close(unit=2)

open(unit=4,file="C:\Growth Curve Hearing\data\D2.txt")
do i=1,S-2
    read(4,*)D2(i,:)
end do
close(unit=4)

open(unit=5,file="C:\Growth Curve Hearing\data\obs_both.txt")
do i=1,n
    read(5,*)obs(i)
end do
close(unit=5)

open(unit=6,file="C:\Growth Curve Hearing\data\cobs_both.txt")
do i=1,n+1
    read(6,*)cobs(i)
end do
close(unit=6)

open(unit=7,file="C:\Growth Curve Hearing\data\cobs_ij.txt")
do i=1,nn+1
    read(7,*)cobs_ij(i)
end do
close(unit=7)

open(unit=8,file="C:\Growth Curve Hearing\data\cobs_mat.txt")
do i=1,n
    read(8,*)cobs_mat(i,:)
end do
close(unit=8)

open(unit=9,file="C:\Growth Curve Hearing\data\XXp_both.txt")
do i=1,m
    read(9,*)XXp(i,:)
end do
close(unit=9)

open(unit=10,file="C:\Growth Curve Hearing\data\r0_both.txt")
do i=1,ss+1
    read(10,*)r00(i,:)
end do
close(unit=10)

```

```

open(unit=11,file="C:\Growth Curve Hearing\data\sig_r0_both.txt")
  do i=1,ss+1
    read(11,*)sig_r00(i,:)
  end do
close(unit=11)
!*****
fp=XXp(:,1);Xp=XXp(:,2:(S-1));
Zp(:,1)=VecConstant(m,db1e(1));Zp(:,2)=XXp(:,1)

r0(1:4,1)=r00(1:4,1)
r0(5,1)=r00(6,1)
sig_r0(1:4,1:4)=sig_r00(1:4,1:4)
sig_r0(5,1:4)=sig_r00(6,1:4)
sig_r0(1:4,5)=sig_r00(1:4,6)
sig_r0(5,5)=sig_r00(6,6)

m0=r0
MM0=sig_r0/n

Q=inv(ss,sig_r0)/qq
mu_g=MatConstant(S-2,1,db1e(0))

do i=1,S-2
  gam(i,1)=mu_g(i,1)
enddo

Minv=inv(ss,MM0)

do i=1,ss
  do jj=1,n
    r(i,jj)=r0(i,1)
  enddo
enddo

b(1:2,:)=r(1:2,:)
v(1:2,:)=r(3:4,:)
v(3,:)=r(6,:)

do i=1,ss
  mu_r(i,1)=r0(i,1)
enddo

prec_r(1:ss,:)=Q*qq
ct=VecConstant(n,db1e(0))
vsum=MatConstant(3,n,db1e(0))
vmat=MatConstant(3,3*n,db1e(0))

```

```

DD=matmul(transpose(D2),D2)

de=both(:,6);dt=both(:,7)

Y=reshape(both(:,9),(/size,1/))
Z(:,1)=dble(1);Z(:,2)=both(:,11)
lamb(1)=rgamma(c1,c2)

ii=0 !counting index for the predicted responses
D0=MatMul(Minv,m0)
sump=VecConstant(size,dble(0))

mu_dp=reshape(VecConstant(m,dble(0)),(/m,1/))
delta_t=(/0.5,1.0,1.5,2.0,3.0,4.0/)
!*****
!***** Start sampling *****
do it=1,rd-1
  print*,it

  sum_r=MatConstant(ss,1,dble(0))
  do i=1,n
    sum_r=sum_r+reshape(r((it-1)*ss+1:it*ss,i),(/ss,1/))
  enddo

  VV=inv(ss,Minv+n*prec_r((it-1)*ss+1:it*ss,:))
  D=D0+MatMul(prec_r((it-1)*ss+1:it*ss,:),sum_r)

  mu_r(it*ss+1:it*ss+ss,1)=rmvn(ss,MatMul(VV,D),VV)
  do i=1,n
    do jj=1,ss
      E(jj,i)=r((it-1)*ss+jj,i)-mu_r(it*ss+jj,1)
    enddo
  enddo

  prec_r(it*ss+1:it*ss+ss,:)
=wishart(ss,qq+n,inv(ss,inv(ss,Q)+MatMul(E,transpose(E))))
  sig_r=inv(ss,prec_r(it*ss+1:it*ss+ss,:))
  mu_b=reshape(mu_r(it*ss+1:it*ss+2,1),(/2,1/))
  mu_v=reshape(mu_r(it*ss+3:it*ss+ss,1),(/3,1/))

  sig_b=sig_r(1:2,1:2)
  sig_v=sig_r(3:ss,3:ss)
  sig_bv=sig_r(1:2,3:ss)
  sig_vb=sig_r(3:ss,1:2)

  XPX=MatConstant(S-2,S-2,dble(0));XPY=MatConstant(S-2,1,dble(0));

```

```

do i=1,n
  vi=v((it-1)*3+1:it*3,i)
  allocate(sig_i(obs(i),obs(i)))
  allocate(sig_ii(obs(i),obs(i)))
  allocate(prec_i(obs(i),obs(i)))
  allocate(Xi(obs(i),S-2))
  allocate(Yi(obs(i),1))
  allocate(Zi(obs(i),2))
  allocate(mu_y(obs(i),1))
  allocate(dti(obs(i)))
  allocate(eari(obs(i)))

  Xi=X(cobs(i)+1:cobs(i+1),:)
  Yi=Y(cobs(i)+1:cobs(i+1),:)
  Zi=Z(cobs(i)+1:cobs(i+1),:)
  dti=dt(cobs(i)+1:cobs(i+1))
  eari=de(cobs(i)+1:cobs(i+1))

  sig_i=dble(0)
  do k=1,12
    do j1=cobs_mat(i,k)+1,cobs_mat(i,k+1)
      do j2=cobs_mat(i,k)+1,cobs_mat(i,k+1)
        sig_i(j1,j2)=exp(-exp(vi(2))*(dti(j1)-dti(j2))**2
          -exp(vi(3))*(eari(j1)-eari(j2))**2)
      enddo
    enddo
  enddo

  sig_i=sig_i*exp(vi(1))
  prec_i=inv(obs(i),sig_i)

  m_bv=mu_b
    +matmul(sig_bv,matmul(inv(3,sig_v),reshape(vi,(/3,1/))-mu_v))
  s_bv=sig_b-matmul(sig_bv,matmul(inv(3,sig_v),sig_vb))

  G=inv(2,matmul(transpose(Zi),matmul(prec_i,Zi))+inv(2,s_bv))
  H=matmul(transpose(Zi),matmul(prec_i,Yi
    -matmul(Xi,reshape(gam((it-1)*(S-2)+1:it*(S-2),1),
(/S-2,1/)))))+matmul(inv(2,s_bv),m_bv)

  b(it*2+1:it*2+2,i)=rmvn(2,matmul(G,H),G)

  m_vb=mu_v+matmul(sig_vb,matmul(inv(2,sig_b),
    reshape(b(it*2+1:(it+1)*2,i),(/2,1/))-mu_b))
  s_vb=sig_v-matmul(sig_vb,matmul(inv(2,sig_b),sig_bv))

  sv=s_vb

```

```

if (it>rd0) then
  vsum(:,i)=vsum(:,i)+vi
  vmat(:,(i-1)*3+1:i*3)=vmat(:,(i-1)*3+1:i*3)+xxt(3,vi)
endif

if (it>rd1) then
  sv=vmat(:,(i-1)*3+1:i*3)/dble(it-rd0-1)-xxt(3,vsum(:,i))
  /(dble(it-rd0)*dble(it-rd0-1))
  do l=1,3
    sv(l,l)=sv(l,l)+eps
  enddo
  sv=dble(2.4**2)*sv/dble(3)
endif

sig_ii=dble(0)
vip=rmvn(3,vi,sv)

do k=1,12
  do j1=cobs_mat(i,k)+1,cobs_mat(i,k+1)
    do j2=cobs_mat(i,k)+1,cobs_mat(i,k+1)
      sig_ii(j1,j2)=exp(-exp(vip(2))*(dti(j1)-dti(j2))**2
                        -exp(vip(3))*(eari(j1)-eari(j2))**2)
    enddo
  enddo
enddo

sig_ii=sig_ii*exp(vip(1))

mu_y=MatMul(Xi,reshape(gam((it-1)*(S-2)+1:it*(S-2),1),(/S-2,1/)))
+MatMul(Zi,reshape(b(it*2+1:it*2+2,i),(/2,1/)))

a=pmvn(obs(i),Yi,mu_y,sig_ii)*pmvn(3,vip,m_vb,s_vb)/
(pmvn(obs(i),Yi,mu_y,sig_i)*pmvn(3,vi,m_vb,s_vb))

call random_number(u)
if(u<a) then
  v(it*3+1:it*3+3,i)=vip
  ct(i)=ct(i)+1
else
  v(it*3+1:it*3+3,i)=vi ! v_i is updated
end if

vi=v((it)*3+1:(it+1)*3,i)
sig_i=dble(0)
do k=1,12
  do j1=cobs_mat(i,k)+1,cobs_mat(i,k+1)

```

```

do j2=cobs_mat(i,k)+1,cobs_mat(i,k+1)
  sig_i(j1,j2)=exp(-exp(vi(2))*(dti(j1)-dti(j2))**2
                    -exp(vi(3))*(eari(j1)-eari(j2))**2)
enddo
enddo
enddo

sig_i=sig_i*exp(vi(1))
prec_i=inv(obs(i),sig_i)

XPX=XPX+reshape(matmul(transpose(Xi),matmul(prec_i,Xi)),(/S-2,S-2/))
XPY=XPY+reshape(matmul(transpose(Xi),matmul(prec_i,(Yi-matmul(Zi,
reshape(b(it*2+1:(it+1)*2,i),(/2,1/)))))),(/S-2,1/))

deallocate(sig_i)
deallocate(sig_ii)
deallocate(prec_i)
deallocate(Xi)
deallocate(Yi)
deallocate(Zi)
deallocate(mu_y)
deallocate(dti)
deallocate(eari)
end do

prec_g=lamb(it)*DD
J=reshape(XPY+matmulself(S-2,S-2,1,prec_g,mu_g),(/S-2,1/))
W=inv(S-2,XPX+prec_g)
gam(it*(S-2)+1:(it+1)*(S-2),1)=rmvn(S-2,matmul(W,J),W)
gam_it=reshape(gam(it*(S-2)+1:(it+1)*(S-2),1),(/S-2,1/))

DG=matmul(D2,gam_it-mu_g)
GDG=VecMulSum(S-2,DG,DG)
lamb(it+1)=rgamma(c1+(S-dble(2))/dble(2),c2+0.5*GDG)

r(it*ss+1:it*ss+2,:)=b(it*2+1:it*2+2,:)
r(it*ss+3:it*ss+ss,:)=v(it*3+1:it*3+3,:)

!Prediction of new observations

if (it>rd*p-1 .and. it/thin==int(it/thin)) then
  ii=ii+1
  mu_yp=matmul(Xp,reshape(gam(it*(S-2)+1:(it+1)*(S-2),1),(/S-2,1/)))
  +matmul(Zp,reshape(mu_r(it*ss+1:it*ss+2,1),(/2,1/)))

  sig_yp=diag_matrix(m)
  sig_yp=sig_yp*exp(mu_r(it*ss+3,1))

```

```

Yp(((ii-1)*m+1):(ii*m),1)=rmvn(m,mu_yp,sig_yp)

do k=1,m1
    sig_dp=2*(1-exp(-exp(mu_r(it*ss+4,1))*delta_t(k)**2))*sig_yp
    delta_p(((ii-1)*m+1):(ii*m),k)=rmvn(m,mu_dp,sig_dp)
enddo

!Calculate CPO
mu_spl=matmul(X, gam_it)

do i=1,n
    do k=cobs(i)+1,cobs(i+1)
        mu=b(it*2+1,i)+b(it*2+2,i)*both(k,11)+mu_spl(k,1)!
        tau=1/exp(v(it*3+1,i))
        pinv=1/pn(Y(k,1),mu,tau)
        sump(k)=sump(k)+pinv
    enddo
enddo

endif
enddo

b_chain=reshape(b,(/2*n*rd,1/))
v_chain=reshape(v,(/3*n*rd,1/))
r_chain=reshape(r,(/ss*n*rd,1/))

!*****
open(unit=11,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\beta.txt")
do i=1,2*n*rd
    write(11,*)b_chain(i,:)
end do
close(unit=11)

open(unit=12,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\vchain.txt")
do i=1,3*n*rd
    write(12,*)v_chain(i,:)
end do
close(unit=12)

open(unit=13,file="C:\Growth Curve Hearing\MCMC\NoGA_simple
\mu_r_chain.txt")
do i=1,ss*rd
    write(13,*)mu_r(i,:)
end do
close(unit=13)

```



```

open(unit=14,file="C:\Growth Curve Hearing\MCMC\NoGA_simple
\gamma_chain.txt")
  do i=1,(S-2)*rd
    write(14,*)gam(i,:)
  end do
close(unit=14)

open(unit=15,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\lamb.txt")
  do i=1,rd
    write(15,*)lamb(i)
  end do
close(unit=15)

open(unit=16,file="C:\Growth Curve Hearing\MCMC\NoGA_simple
\rchain.txt")
  do i=1,ss*n*rd
    write(16,*)r_chain(i,:)
  end do
close(unit=16)

open(unit=17,file="C:\Growth Curve Hearing\MCMC\NoGA_simple
\prec_r_1.txt")
  do i=1,ss*rd
    write(17,*)prec_r(i,1:3)
  end do
close(unit=17)

open(unit=18,file="C:\Growth Curve Hearing\MCMC\NoGA_simple
\prec_r_2.txt")
  do i=1,ss*rd
    write(18,*)prec_r(i,4:ss)
  end do
close(unit=18)

open(unit=19,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\Yp.txt")
  do i=1,rd*p/thin*m
    write(19,*)Yp(i,:)
  end do
close(unit=19)

open(unit=20,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\dp.txt")
  do i=1,rd*p/thin*m
    do jj=1,m1
      write(20,*)delta_p(i,jj)
    enddo
  end do
close(unit=20)

```

```
open(unit=100,file="C:\Growth Curve Hearing\MCMC\NoGA_simple\sump.txt")
  do i=1,size
    write(100,*)sump(i)
  end do
close(unit=100)
end program
```



# APPENDIX C

## PERMISSION TO REPRINT OF CHAPTER 2

3/31/2016

RightsLink Printable License

### JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Mar 31, 2016

---

This Agreement between Junshu Bao ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	3839630049157
License date	Mar 31, 2016
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	The Canadian Journal of Statistics
Licensed Content Title	Bayesian nonparametric multivariate ordinal regression
Licensed Content Author	Junshu Bao, Timothy E. Hanson
Licensed Content Date	May 28, 2015
Pages	21
Type of use	Dissertation/Thesis
Requestor type	Author of this Wiley article
Format	Print and electronic
Portion	Full article
Will you be translating?	No
Title of your thesis / dissertation	Development and Application of Bayesian Semiparametric Models for Dependent Data
Expected completion date	Apr 2016
Expected size (number of pages)	183
Requestor Location	Junshu Bao 113 Bartram St.  LEXINGTON, SC 29072 United States Attn: Junshu Bao
Billing Type	Invoice
Billing Address	Junshu Bao 113 Bartram St.  LEXINGTON, SC 29072 United States Attn: Junshu Bao
Total	0.00 USD
Terms and Conditions	

### TERMS AND CONDITIONS

<https://s100.copyright.com/AppDispatchServlet>

1/5