A Bayesian Joint Model Framework for Repeated Matrix-Variate Regression with Measurement Error Correction

by

Yue Wang

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Statistical Machine Learning

Department of Mathematical and Statistical Science

University of Alberta

Abstract

In this thesis, with the purpose of correcting for potential measurement errors in repeatedly-observed matrix-valued surrogates, and examining the underlying association between latent matrix covariates and a binary response, we propose a Bayesian joint model framework. This joint model method imposes a low-rank structure on the covariance matrix of additive measurement errors, and relates the binary response with low-dimensional features extracted from latent matrix covariates. Although in our framework, the latent matrix covariates are not directly observed and used as predictors in the proposed model, a unique formulation of associations between latent covariates and response is derived. Simulation studies demonstrate that our proposed method outperforms other naive methods (i.e., a naive joint model and a naive two-stage model) with respect to the estimates of underlying association. The advantage of the proposed method is more notable in the circumstances where a small sample size but high dimensional matrix covariates are presented. Finally, we apply this proposed framework to a case study that explores the association between a favorable response to antidepressant treatment and resting stage electroencephalography (EEG) data measured under different conditions. Results suggested that our method should be able to handle the attenuation bias induced from measurement errors and to reveal the most underlying association, compared to other competing methods.

Preface

This thesis is the original work of Yue Wang, and I am the only author of this thesis. Part of the results will be published in the future.

Acknowledgements

There are many people that I want to thank. First of all, I would like to sincerely acknowledge my supervisor, Dr. Bei Jiang, for her caring guidance and support during my master study. Throughout the two and a half years, her advice and encouragement helped me resolve many challenges in both personal life and academic study. It is my great honor to share her encyclopedic knowledge and have her deep belief in my abilities. Thank you for inspiring my dedication to Statistics.

I also want to thank everyone in the Prof Jiang & Kong's study group who have helped me a lot in academic study. I am very happy to be part of this group to discuss the up-to-date papers altogether. Special thanks to Dr. Kong for his support ofthe Mitacs program. This experience is really valuable to me. I would also thank Peng Yu for helping me in coding and debugging.

In the end, I really appreciate the support of my family and my boyfriend. Your company and encouragement helped me persist in research, especially during hard days.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The widely used therapies for effective disorders such as depression are antidepressant medications (Holsboer, 2008). However, Measuring the effectiveness of these medications in in-vivo studies is time-consuming and may present unexpected risks to the patients involved (Jiang et al., 2020). Therefore, there emerged a noticeable demand for predicting the response to antidepressants so as to choose the right medicines for patients prior to obtaining treatment (Holsboer, 2008). In recent decades, neuroimaging techniques have demonstrated their ability to provide antidepressant treatment biomarkers (Steiger & Kimura, 2010). Among those techniques, Electroencephalography (EEG) is a prominent method that measures and digitalizes the electrical brain activities under the specific mental state (Jiang et al., 2020). Various subsequent researches are conducted to relate the EEG records to antidepressant treatment outcomes (Alhaj, Wisniewski, & McAllister-Williams, 2011; Olbrich & Arns, 2013). Among these methods, the common practice to make the prediction is to rely on EEG data recorded under a single mental condition, for example, eyes-open state or eyes closed state (Iosifescu et al., 2009; Lee et al., 2011).

Although EEG gradually becomes one of the most used predictors of the antidepressant response and gives promising results (Wade & Iosifescu, 2016),

recent studies have shown that EEG data could potentially involve measurement errors from various aspects. For example, one type of measurement errors are the performance errors among patients arising from irregular impulses in attention tend to change the fluctuation patterns in the theta and alpha frequency bands in EEG records (van Driel, Ridderinkhof, & Cohen, 2012). Another commonly-acknowledged source of measurement errors occurs in the process of estimating the current sources of EEG recorded from the electrodes mounted on the scalp (Khosla, Don, & Kwong, 1999; Liu, Dale, & Belliveau, 2002). Due to the impedance irregularities and electrode-spacing errors, localization of the estimated current sources could be inaccurate. Even with the help of source density methods which are supposed to alleviate these errors through matrix transformation, such errors can hardly be eliminated (Tenke & Kayser, 2012).

Such measurement errors are in need of resolution as studies have shown that it could lead to significant bias in parameter estimation and inference results. For example, in an investigation of additive measurement error in matrix-valued explanatory variables (refer to matrix-valued covariates hereafter), Fang and Yi (2020) theoretically refined the bias in estimates induced by additive measurement errors in a logistic regression model. Their results showed that, as the severity of measurement errors increase, the negative impact brought by the bias to naive estimates increases accordingly. With the limited sample size and high dimensionality of matrix covariates, the estimation and inference results would be even worse. To address the impact of measurement errors, the authors suggested two methods of estimating the bias with additive measurement errors. In their experiments, their methods can successfully get rid of the bias induced from measurement errors and make correct inferences on the parameters. It is noted that this method is typically proposed for the circumstance when the sample size is larger than the dimen-

sionality of matrix-valued surrogates (i.e., repeatedly measured matrix covariates who are contaminated by measurement errors). When the sample size is smaller than matrix surrogates' dimension, the dimension reduction techniques (two-directional two-dimensional principal component analysis (Zhang & Zhou, 2005) was employed in their paper) need to be implemented first before correcting for measurement errors in matrix surrogates. This two-step implementation indeed weakens the interpretability of the association between original matrix-valued covariates and responses.

To the best of our knowledge, their work is the only attempt to incorporate the measurement errors in the matrix-variate regression problem. Thus, adjusting for measurement errors in matrix-valued covariates under other regression setting is of great research interest, leading to a research gap in matrix-variate regression models with measurement error correction and dimension reduction together. In this paper, to fill such a research gap, we propose a Bayesian joint model framework (hereafter, we refer to our framework as JMME) to simultaneously estimate the true latent matrix covariates from repeatedly observed matrix surrogates by incorporating measurement errors, extract low dimensional features from latent covariates, and relate them to binary responses through a probit regression model. Our study should relax the measurement error in matrix-valued covariates, and the proposed model further allows a bi-linear dimension reduction for the latent matrix-valued variables. Moreover, the Bayesian credible intervals allow the quantification of estimates' uncertainties for inference purposes (Jiang et al., 2020). Practically, we take advantage of the repeatedly measured EEG under two conditions: eyes-open state and eyes-closed state, and examine the association between the latent EEG baseline data (after measurement error correction) and the treatment outcomes from the selective serotonin reuptake inhibitors.

This work is highly related to the work from Jiang et al. (2020). However,

they directly related the noisy matrix covariates to the responses without considering the potential measurement errors. Therefore, our work can be considered as an extension on measurement errors correction through repeated measures in the matrix-variate regression context.

The contributions of this study are multi-fold:

1. A new framework, JMME, is proposed to adjust for measurement errors in the repeatedly observed matrix-valued covariates through a low-rank structure on the covariance matrix of measurement errors.

2. The proposed JMME framework is the first Bayesian matrix-variate regression model that handles the repeated measures where the sample size can be smaller than the dimensionality of matrix covariates given the existence of measurement error. Through simulation and case study, JMME is examined to be able to incorporate the measurement errors and reduce the dimensionality simultaneously.

3. The theoretical framework of JMME leads to an identifiable regression of binary outcomes on the latent matrix covariates after measurement error correction. Therefore, the inferences for associations between binary outcomes and the latent matrix covariates can be naturally obtained by Bayesian credible intervals from Markov chain Monte Carlo (MCMC) chains.

The rest of the thesis is organized as follows:

In Chapter 2, we discuss the literature related to this study. Specifically, we provide a review on recently researched dimension reduction techniques for matrix-valued covariates, the classic measurement error models handling errors in covariates, the influence of measurement errors

on principal component analysis, and the most recent study on matrix-variate logistic regression with measurement error correction.

In Chapter 3, we introduce the proposed JMME relating a binary response to the repeatedly measured erroneous matrix covariates through low-dimensional features. The corresponding hierarchical structure of JMME is presented. Its identifiability is also justified in this chapter. Then, the choices of prior distributions for model parameters are provided and the posterior distributions for the Gibbs sampling procedure are derived accordingly. At the end of this Chapter, we illustrate how to select the desired dimensions for the extracted features.

Chapter 4 demonstrates a simulation study to examine the performance of JMME. The simulation setups are described in detail. To examine the advantage of our joint model, we also implement two naive methods (i.e., without considering measurement errors) to compare the estimation correctness for model parameters. The simulation results are presented and analyzed in detail.

In Chapter 5, we apply the proposed JMME to our motivating dataset (EEG baseline data) to explore the association between latent EEG measures (corrected for measurement errors) and treatment responses with SSRI antidepressants. For comparison purposes, the same two naive methods are implemented to the same motivating dataset and the inference results are analyzed.

In Chapter 6, we conclude the findings with a discussion for the proposed joint model. The potential limitation and future works are also included.

Chapter 2

Literature Review

2.1 Dimension Reduction and Regression model for matrix-valued covariates

Matrix-valued covariates have attracted outstanding attention during recent decades due to the popularity of matrix structured data (i.e., image data, temporal and spatial EEG data) (Ding & Cook, 2016). When used as predictors, such matrix structured data, also known as the matrix-variates, exhibit two important properties: 1. They are usually of high dimensionality; 2. Their elements are dependent on the rows and columns (Ding & Cook, 2016). Such properties present challenges when regressing with the matrix-variates, especially with a small sample size, since simply vectorizing them could produce the predictors with even higher dimensionality and also destroy the natural matrix structure (Ding & Cook, 2016). To overcome the high dimensionality in the matrix-valued covariates, dimension reduction techniques were intensively researched so that low dimensional features can be extracted as predictors for consequent regression models. For example, as an extension of classical principal component analysis (PCA), Yang et al. (2004) proposed a two-dimensional principal component analysis (2D-PCA) that projects matrix-valued covariates onto their row space. Accordingly, Zhang and Zhou (2005) proposed an alternative 2D-PCA to project a matrix onto column space. They also introduced

a $(2D)^2$-PCA method that simultaneously projects the matrix-valued covariates on rows and columns spaces by eigenvalue-eigenvector decomposition. The multilinear principal component analysis (MPCA) then extends the dimension reduction to tensor variables with arbitrary order (Lu, Plataniotis, & Venetsanopoulos, 2008). Afterward, there are plenty of variant methods based on MPCA that can solve different problems. For example, Lai et al. (2014) proposed a multilinear sparse PCA that rewrites the MPCA formulation into sparse regression for face recognition problem; A non-Negative multilinear principal component is proposed for music genre classification, which reduces the dimension for third-order tensors while preserves the non-negativity in auditory data (Panagakis, Kotropoulos, & Arce, 2009).

In another track with the regression model context, a number of models have been investigated to handle matrix-variates. Hung and Wang (2013) proposed a logistic regression model that imposes a rank-1 structure to the coefficients of matrix-variates. To handle the small sample size situation, the likelihood function is penalized with l2 norms to obtain the maximum likelihood estimates for model parameters. This model is then extended to the generalized linear model framework (i.e., allows either continuous or discrete responses) by H. Zhou, Li, and Zhu (2013). They also generalized their GLM to tensor variables with arbitrary order by applying low-rank Candecomp/Parafac (CP) decomposition on the coefficients of tensor predictors. Penalties are applied to maximize the likelihood to obtain the estimation of coefficients. Later on, H. Zhou and Li (2014) investigated the sparsity regularization, i.e., regularization with the nuclear norm, in the GLM with matrix-variates. On the other hand, a Bayesian method is proposed by Jiang et al. (2017) that assumes a hierarchical structure for the low-rank CP decomposition. Another Bayesian joining model is investigated later, which utilizes a probabilistic MPCA model and regresses the response with extracted features simultaneously (Jiang et al., 2020). How-

ever, as mentioned earlier, all current dimension reduction techniques and the subsequent models do not take into account the fact that the observed matrix covariates may be measured with potential errors. They either directly regress the response with observed matrix covariates, or straightforwardly applied the dimension reduction techniques to the observations without adjusting for measurement errors.

2.2 Classical measurement error model

Measurement error is a common issue in many research areas, for example, social science (Heckman, Stixrud, & Urzua, 2006) and medication (Atkinson & Nevill, 1998), where the variables of interest cannot be directly or precisely observed. These errors can be easily ignored by the analysis. However, ignoring these errors in variables cannot always give accurate estimation and thus inference results for model parameters. The attenuation bias typically occurs when modeling directly on the erroneous predictors, leading to the parameter estimates shrunk toward zero (Schofield, 2015).

During the past decades, various models are investigated to handle different kinds of measurement errors. Among them, the models correcting for normally distributed errors that are independent of latent variables (refer to classical measurement errors) are well researched (Chen, Hong, & Nekipelov, 2007). For example, Carroll and Stefanski (1990) proposed a quasi-likelihood estimation for model parameters in linear regression with such classical measurement errors. A semiparametric estimation is investigated later in logistic model by Carroll and Wand (1991). Fuller (2009) described a moment-based method to correct for classical measurement errors in linear models. Richardson and Gilks (1993) proposed a Bayesian approach for logistic regression with such measurement errors, based on a conditional independence assumption: the response is independent of observed surrogates given the unobserved predictors.

Their model consists of three submodels: (1) an outcome model for unobserved predictors and responses; (2) a measurement model for classical measurement errors; (3) a prior model for parameters. Later on, a Mixed Effects Structural Equations (MESE) model is introduced to handle measurement errors in survey data (Schofield, 2015). As data grows, measurement error models are relaxed to high dimensional datasets. For example, Datta, Zou, et al. (2017) investigated a cocolasso model that applies a lasso penalty on least square estimators with additive measurement errors scenario. Although these models can successfully handle the measurement errors in covariates, little literature pays attention to the measurement errors in matrix-valued covariates. As the matrix-variate models are gradually becoming popular, the measurement errors in the matrix-valued covariates await further investigations.

2.3 The impact of measurement errors in dimension reduction

There is several literature that takes measurement errors into account while applying dimension reduction techniques. In the vector covariates settings, PCA is a popular technique to reduce dimensionality. It searches for linear combinations of the vector variables which can capture the largest variation. The coefficients of the linear combinations are called the principal components or the loadings. Hellton and Thoresen (2014) had explored the biases in the estimates of loadings when applying PCA on the error-contaminated data. Through Taylar's expansion on eigenvectors for the data matrix, it is proved that ignoring measurement errors during PCA could contribute to the large variability in the principal components (PC). Moreover, depending on the structure of measurement errors (i.e., correlated or heterogeneous), the bias induced from the errors may have different directions (Hellton & Thoresen, 2014). Such impacts would make the interpretation based on the loadings

difficult. Consequently, a maximum likelihood principal component analysis is investigated to extend the PCA with additive correlated measurement errors (Wentzell et al., 1997). Later on, a Bayesian method using probabilistic PCA is proposed by Sanguinetti et al. (2005) to incorporate the measurement errors.

In the matrix covariates setting, the MPCA with matrix-valued covariates can be considered as a parsimonious version of PCA with a rank-1 structure in PCA loadings (Hung et al., 2012). Therefore, similar to the results proved by Hellton and Thoresen (2014), when the matrix covariates are measured with errors, the estimated loadings for matrix covariates should be influenced by the measurement errors, leading to difficulties in interpretation based on the estimated loadings. Hence, it is worth trying to take measurement errors into account during the MPCA procedure for matrix-valued variables.

2.4 Matrix-variate regression with measurement error correction

Despite the vast amount of work on measurement error correction, to the best of our knowledge, very few have been investigating the measurement errors in matrix-variate regression. A notable prior work by Fang and Yi (2020) introduced an matrix-variate logistic regression with additive matrix-variate measurement errors in the matrix-valued observation. By maximizing the likelihood of data, they explicitly obtained the model parameters' estimates with and without measurement errors. Therefore, the bias induced by matrix-variate measurement errors can be quantified and hence estimated. As described in 1, their findings illustrated the negative impact of measurement errors on the logistic regression model. However, one shortcoming of this method is that it needs to first apply a $(2D)^2$ PCA method to reduce the dimension when the sample size is smaller than the dimension of matrix-valued covariates. Such a two-fold procedure weakens the interpretability of the coefficients

in the original matrix-variate scale. To address this concern, we proposed a Bayesian joint model that is able to not only adjust for measurement errors in the high-dimensional matrix covariates, but also reduce the dimension of true unobserved matrix covariates simultaneously. A probit regression model was built for the binary response and the extracted low dimensional features so that a closed form of posterior distributions can be obtained during the Gibbs sampling procedure.

Chapter 3

Matrix-variate Regression Joint Model Framework: JMME

In this chapter, we introduce the framework for our proposed JMME for matrix-valued covariates regression with measurement error correction. The identifiability of our framework is justified and the hierarchical structure of JMME is then established. In addition, this chapter describes the prior distributions that we chose for model parameters; their corresponding posterior distributions are clarified accordingly. In the end, we illustrate the procedure of selecting the optimal dimension for the extracted low-dimensional features.

3.1 Framework Formulation

In this section, we describe the proposed joint model framework that consists of three sub-models. For each subject $i, i \in \{1, \ldots, n\}$, let O_i represents the binary response for subject i and $\boldsymbol{X}_i^* \in \mathcal{R}^{p \times q}$ denotes the latent matrix-valued covariates which contains the information to predict O_i. We further assume that the exact measurement of $\boldsymbol{X}_i^*$ cannot be obtained directly, but the proxy measurements with errors can be observed repeatedly. For each subject i, we assume that, in total, the latent matrix-valued covariates are repeatedly observed for M times. These observations are called surrogate matrix covariates, and are denoted as $\boldsymbol{X}_{im} \in \mathcal{R}^{p \times q}$, for $m \in \{1, \ldots, M\}$. The vector of scalar

covariates, which is assumed to be observed precisely, is denoted as $\boldsymbol{Z}_i \in \mathcal{R}^{p_z}$ with p_z represents the number of scalar covariates.

3.1.1 Sub-model for Measurement Error between Repeated Contaminated Measures and Latent Matrix Covariates

A new measurement error sub-model is proposed to quantify the measurement error between the repeated surrogate covariates $\boldsymbol{X}_{im}$ and the latent covariates $\boldsymbol{X}_i^*$. Motivated by the classical measurement errors, we assumed an additive measurement error that is independent of O_i and $\boldsymbol{z}_i$. The sub-model is formulated as follows:

$$\mathcal{M}_{ME} : \boldsymbol{X}_{im} = \boldsymbol{X}_i^* + \boldsymbol{\epsilon}_{im}^*, \text{where } \text{vec}(\boldsymbol{\epsilon}_{im}^*) \sim \mathcal{N}\left(0, \text{diag}(\text{vec}(\widetilde{\boldsymbol{\Sigma}}_i))\right). \quad (3.1)$$

Note:

1. $\text{vec}(\cdot)$ denotes the vectorization operator.

2. $\boldsymbol{\epsilon}_{im}^*$ represents the random matrix-valued measurement error that is independent of $\boldsymbol{X}_i^*$.

3. $\widetilde{\boldsymbol{\Sigma}}_i$ is the covariance matrix of measurement errors. To preserve the natural matrix structure, it is defined as $\widetilde{\boldsymbol{\Sigma}}_i = \begin{pmatrix} \sigma_{i11}^2 & \dots & \sigma_{i1q}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{ip1}^2 & \dots & \sigma_{ipq}^2 \end{pmatrix}$, where σ_{ijk}^2 represents the variance for the measurement error at row j and column k for subject i.

4. The covariance matrix structure for $\boldsymbol{\epsilon}_{im}^*$ implies the independent measurement errors among elements in different rows and columns of matrix-valued covariates.

In order to reduce the number of parameters and maintain the matrix form of $\widetilde{\boldsymbol{\Sigma}}_i$, we further introduce a low rank structure for $\widetilde{\boldsymbol{\Sigma}}_i$ through rank $= r$:

$$\mathcal{M}_{Low} : \log(\widetilde{\boldsymbol{\Sigma}}_i) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^\top + \widetilde{\boldsymbol{\epsilon}}_i, \text{where } \text{vec}(\widetilde{\boldsymbol{\epsilon}}_i) \sim \mathcal{N}\left(0, \delta^2 \boldsymbol{I}\right). \quad (3.2)$$

Here:

1. log is the element-wise log calculation
2. $\widetilde{\boldsymbol{\epsilon}}_i \in \mathcal{R}^{p\times q}$ is an error matrix resulting from the low rank approximation
3. With $r \leq \min(p, q)$, the low rank structure of $\log(\widetilde{\boldsymbol{\Sigma}}_i)$ is obtained from $\widetilde{\boldsymbol{U}} \in \mathcal{R}^{p\times r}$ and $\widetilde{\boldsymbol{V}} \in \mathcal{R}^{q\times r}$

3.1.2 Sub-model for Latent Matrix Covariates through MPCA Formulation

This proposed sub-model aims to find a low dimensional representation of latent matrix covariates $\boldsymbol{X}_i^*$ which is able to approximate $\boldsymbol{X}_i^*$ while preserving innate matrix structure in $\boldsymbol{X}_i^*$ at the same time. Resorting to the probabilistic formulation of multi-linear principle component analysis (MPCA) (Jiang et al., 2020), we defined our MPCA sub-model as the following:

$$\mathcal{M}_{MPCA} : \boldsymbol{X}_i^* = \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i, \text{where } \text{vec}(\boldsymbol{\epsilon}_i) \sim \mathcal{N}\left(\boldsymbol{0}, \phi^{-1} I_{pq\times pq}\right) \tag{3.3}$$

Note:

1. $\boldsymbol{\epsilon}_i$ here represents the error matrix due to the dimension reduction procedure. It is independent of $\boldsymbol{u}_i$.
2. With $p_0 < p, q_0 < q$, $\text{vec}(\boldsymbol{u}_i) \in \mathcal{R}^{p_0\times q_0}$ is the desired low dimensional representation of $\boldsymbol{X}_i^*$.
3. We assume that $\text{vec}(\boldsymbol{u}_i) \sim \mathcal{N}\left(\text{vec}(\boldsymbol{\eta}), \boldsymbol{\xi} \otimes \boldsymbol{\lambda}\right)$, where $\boldsymbol{\xi}$ and $\boldsymbol{\lambda}$ are diagonal matrices representing the covariance matrix of $\boldsymbol{u}_i$, and $\text{vec}(\boldsymbol{\eta})$ stands for the mean matrix.
4. $\boldsymbol{A} \in \mathcal{R}^{p\times p_0}$ and $\boldsymbol{B} \in \mathcal{R}^{q\times q_0}$ are the MPCA projection matrices that mapping $\boldsymbol{X}_i^*$ onto $\boldsymbol{u}_i$. Followed by Jiang et al. (2020), we let $\boldsymbol{A}^\top\boldsymbol{A} = I_{p_0\times p_0}$, $\boldsymbol{B}^\top\boldsymbol{B} = I_{q_0\times q_0}$.

Sub-model 3.3 can be further decomposed to regulate the population mean and to standardize the matrix covariate $\boldsymbol{u}_i$. The decomposed sub-model is of the form

$$\begin{aligned}
\text{vec}(\boldsymbol{X}_i^*) &= (\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{u}_i) + \text{vec}(\boldsymbol{\epsilon}_i) \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})\left[\text{vec}(\boldsymbol{u}_i) - \text{vec}(\boldsymbol{\eta}) + \text{vec}(\boldsymbol{\eta})\right] + \text{vec}(\boldsymbol{\epsilon}_i) \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{\eta}) + (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2}(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{-1/2}(\text{vec}(\boldsymbol{u}_i) - \text{vec}(\boldsymbol{\eta})) + \text{vec}(\boldsymbol{\epsilon}_i) \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{\eta}) + (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2}\text{vec}(\tilde{\boldsymbol{u}}_i) + \text{vec}(\boldsymbol{\epsilon}_i) \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})\left[\text{vec}(\boldsymbol{\eta}) + (\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})\text{vec}(\tilde{\boldsymbol{u}}_i)\right] + \text{vec}(\boldsymbol{\epsilon}_i).
\end{aligned} \tag{3.4}$$

Now, $\tilde{\boldsymbol{u}}_i$ is the standardized matrix covariate so that $\text{vec}(\tilde{\boldsymbol{u}}_i) \sim \mathcal{N}(\boldsymbol{0}, I)$.

This sub-model 3.4 can be equivalently written, in the matrix form, as

$$\begin{aligned}
\mathcal{M}_{MPCA_decomp} : \boldsymbol{X}_i^* &= \boldsymbol{A}\boldsymbol{\eta}\boldsymbol{B}^\top + \boldsymbol{A}\boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{A}\left[\boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\right]\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i
\end{aligned} \tag{3.5}$$

with $\boldsymbol{u}_i = \boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}$.

Remark 1. Our JMME is highly related to the joint model framework proposed by Jiang et al. (2020). In their framework, a similar probabilistic formuation of MPCA is established as follows:

$$\boldsymbol{X}_i = \boldsymbol{\mu} + \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i, \tag{3.6}$$

with $\boldsymbol{\mu}$ stands for population mean and $\text{vec}(\boldsymbol{u}_i) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$. In our formulation, we relax this assumption and consider a more flexible covariance structure for $\boldsymbol{u}_i$: $\text{vec}(\boldsymbol{u}_i) \sim \mathcal{N}(\text{vec}(\boldsymbol{\eta}), \boldsymbol{\xi} \otimes \boldsymbol{\lambda})$. Motivated from the propeties of MPCA developed from Hung et al. (2012), we relax the covariance matrix of $\boldsymbol{u}_i$ as $(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})$ so that the $\boldsymbol{\lambda}$ corresponds to the variance among rows and the $\boldsymbol{\xi}$ corresponds to the variance among columns for $\boldsymbol{u}_i$. Our relaxed assumption also absorbed the population mean $\boldsymbol{\mu}$ into the low dimension features, further allowing a regularization on population mean.

3.1.3 Sub-model for Binary Outcome

This sub-model defines a probit regression model that relates the binary outcome $O_i \in \{0, 1\}$ to the precisely measured scalar covariates $\boldsymbol{z_i}$ and the standardized low-dimensional features $\tilde{\boldsymbol{u}}_i$ extracted from latent matrix covariates. Motivated by the work from Jiang et al. (2020), the reason that we use the standardized $\tilde{\boldsymbol{u}}_i \in \mathcal{R}^{p_0 \times q_0}$ as the predictors instead of the latent true covariates $\boldsymbol{X}_i^* \in \mathcal{R}^{p \times q}$ is because such sub-model significantly reduces the number of model parameters from pq to $p_0 q_0$, while providing the unique transformation between the coefficients for $\tilde{\boldsymbol{u}}_i$ and the coefficients for $\boldsymbol{X}_i^*$. This unique transformation will be derived in details in the following section. Specifically, the probit regression model is of the form:

$$\mathcal{M}_{out} : \Phi^{-1}(p[O_i = 1]) = \psi + \boldsymbol{\gamma}^\top \boldsymbol{z}_i + \boldsymbol{\theta}^\top \text{vec}(\tilde{\boldsymbol{u}}_i), \tag{3.7}$$

Note:

1. $\Phi(\cdot)$ is the cumulative distribution function for a standard normal distribution;
2. $\psi \in \mathcal{R}$ is a constant for intercept in the regression sub-model;
3. $\boldsymbol{\gamma} \in \mathcal{R}^{p_z}$ is a vector of coefficients for scalar covariates z_i;
4. $\boldsymbol{\theta} \in \mathcal{R}^{p_0 q_0}$ is a vector of coefficients for $\text{vec}(\tilde{\boldsymbol{u}}_i)$.

3.2 Identifiability for JMME

It is noticed that, similar as in the model proposed by Jiang et al. (2020), the sub-models $\mathcal{M}_{MPCA\_decomp}$ (3.5) and $\mathcal{M}_{out}$ (3.7) are rotation invariance. That is, the identifiability can only be achieved up to orthogonal rotations. For example, if $\boldsymbol{G} \in \mathcal{R}^{p_0 \times p_0}$ and $\boldsymbol{H} \in \mathcal{R}^{q_0 \times q_0}$ are two arbitrary orthonormal matrices such that $\boldsymbol{G}\boldsymbol{G}^T = \boldsymbol{I}_{p_0}$ and $\boldsymbol{H}\boldsymbol{H}^T = \boldsymbol{I}_{q_0}$, then, with

(a) $\boldsymbol{A}^* = \boldsymbol{A}\boldsymbol{G}$,

(b) $\boldsymbol{B}^* = \boldsymbol{B}\boldsymbol{H}$,

(c) $\boldsymbol{\lambda}^{*1/2} = \boldsymbol{G}^T\boldsymbol{\lambda}^{1/2}\boldsymbol{G}$,

(d) $\boldsymbol{\xi}^{*1/2} = \boldsymbol{H}^T\boldsymbol{\xi}^{1/2}\boldsymbol{H}$,

(e) $\boldsymbol{\eta} = \boldsymbol{G}^T\boldsymbol{\eta}\boldsymbol{H}$,

(f) $\tilde{\boldsymbol{u}}_i^* = \boldsymbol{G}^T\tilde{\boldsymbol{u}}_i\boldsymbol{H}$,

the identical likelihood for $\mathcal{M}_{MPCA\_decomp}$ can be achieved as:

$$\begin{aligned}
&\boldsymbol{A}^*\left[\boldsymbol{\eta}^* + \boldsymbol{\lambda}^{*1/2}\tilde{\boldsymbol{u}}_i^*\boldsymbol{\xi}^{*1/2}\right]\boldsymbol{B}^{*\top} \\
&= \boldsymbol{A}^*\boldsymbol{\eta}^*\boldsymbol{B}^{*\top} + \boldsymbol{A}^*\boldsymbol{\lambda}^{*1/2}\tilde{\boldsymbol{u}}_i^*\boldsymbol{\xi}^{*1/2}\boldsymbol{B}^{*\top} + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{A}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{\eta}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{B}^T + \boldsymbol{A}\boldsymbol{G}\boldsymbol{G}^T\boldsymbol{\lambda}^{1/2}\boldsymbol{G}\boldsymbol{G}^T\tilde{\boldsymbol{u}}_i\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{\xi}^{1/2}\boldsymbol{H}\boldsymbol{H}^T\boldsymbol{B}^T + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{A}\boldsymbol{\eta}\boldsymbol{B}^\top + \boldsymbol{A}\boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i \\
&= \boldsymbol{A}\left[\boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\right]\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i
\end{aligned}$$

Furthermore, $\boldsymbol{\theta}^* = (\boldsymbol{H}^T \otimes \boldsymbol{G}^T)\boldsymbol{\theta}$ leads to

$$\begin{aligned}
\boldsymbol{\theta}^{*\top}\mathrm{vec}(\tilde{\boldsymbol{u}}_i^*) &= \boldsymbol{\theta}^\top(\boldsymbol{H} \otimes \boldsymbol{G})(\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\mathrm{vec}(\tilde{\boldsymbol{u}}_i) \\
&= \boldsymbol{\theta}^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i)
\end{aligned}$$

so that the same likelihood for $\mathcal{M}_{out}$ is achieved.

Since our goal is to examine the association between binary response o_i and latent matrix covariates $\boldsymbol{x}_i^*$ in $p \times q$ space, such rotation invariance can be neglected only when the estimation of association is identified (Jiang et al., 2020). To recover the regression model for o_i given latent matrix-valued covariate $\boldsymbol{x}_i^*$ and scalar covariates $\boldsymbol{z}_i$, we followed the similar procedure from Jiang et al. (2020) by introducing a intermediate variable w_i such that $o_i = \mathcal{I}(w_i > 0)$ and $w_i \sim \mathrm{N}(\psi + \boldsymbol{\gamma}^\top\boldsymbol{z}_i + \boldsymbol{\theta}^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i), 1)$. We further let $\boldsymbol{\nu}$ includes all the parameters in sub-models 3.1, 3.5 and 3.7 so that $\boldsymbol{\nu} = (\widetilde{\boldsymbol{\Sigma}}_i, \widetilde{\boldsymbol{U}}, \widetilde{\boldsymbol{V}}, \delta, \boldsymbol{A}, \boldsymbol{B}, \phi, \boldsymbol{\eta}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \psi, \boldsymbol{\gamma}, \boldsymbol{\theta})$.

Then, with the conditional distribution $p(\text{vec}(\boldsymbol{x}_i^*)|\text{vec}(\tilde{\boldsymbol{u}}_i), \boldsymbol{\nu})$ implied from $\mathcal{M}_{MPCA\_decomp}$ (sub-model 3.4), a joint normal distribution for $p(w_i, \boldsymbol{x}_i^*, \tilde{\boldsymbol{u}}_i|\boldsymbol{z}_i, \boldsymbol{\nu})$ can be inferred. As a result, this joint normal distribution further leads to a conditional normal distribution for $p(w_i|\boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu})$.

To determine this conditional distribution, we first derive the conditional distribution $p(\text{vec}(\tilde{\boldsymbol{u}}_i)|\text{vec}(\boldsymbol{x}_i^*), \boldsymbol{\nu})$:

$$
\begin{aligned}
&p(\text{vec}(\tilde{\boldsymbol{u}}_i)|\text{vec}(\boldsymbol{x}_i^*), \boldsymbol{\nu}) \propto p(\text{vec}(\tilde{\boldsymbol{u}}_i)|\boldsymbol{\nu}) p(\text{vec}(\boldsymbol{x}_i^*)|\text{vec}(\tilde{\boldsymbol{u}}_i), \boldsymbol{\nu}) \\
&\propto \exp\left\{ -\frac{1}{2}\text{vec}(\tilde{\boldsymbol{u}}_i)^\top \text{vec}(\tilde{\boldsymbol{u}}_i) - \frac{\phi}{2} \| \text{vec}(\boldsymbol{x}_i^*) - (\boldsymbol{B} \otimes \boldsymbol{A}) \left[\text{vec}(\boldsymbol{\eta}) + (\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2} \text{vec}(\tilde{\boldsymbol{u}}_i) \right] \|_F^2 \right\} \\
&\propto \exp\left\{ -\frac{1}{2} \left[\text{vec}(\tilde{\boldsymbol{u}}_i)^\top \boldsymbol{C}^{-1} \text{vec}(\tilde{\boldsymbol{u}}_i) - 2\text{vec}(\tilde{\boldsymbol{u}}_i)^\top \boldsymbol{C}^{-1} \boldsymbol{m} \right] \right\} \\
&\propto \exp\left\{ -\frac{1}{2} \left[\text{vec}(\tilde{\boldsymbol{u}}_i) - \boldsymbol{m} \right]^\top \boldsymbol{C}^{-1} \left[\text{vec}(\tilde{\boldsymbol{u}}_i) - \boldsymbol{m} \right] \right\} \text{ by completing the square,}
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{C}^{-1} &= \boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda}), \\
\boldsymbol{m} &= \phi \boldsymbol{C} (\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2} \left[(\boldsymbol{B} \otimes \boldsymbol{A})^\top \text{vec}(\boldsymbol{x}_i^*) - \text{vec}(\boldsymbol{\eta}) \right].
\end{aligned}
$$

This is a kernel for normal density function with mean $\boldsymbol{m}$ and variance $\boldsymbol{C}$. This derivation results a conditional normal distribution such that $\text{vec}(\tilde{\boldsymbol{u}}_i)|\text{vec}(\boldsymbol{x}_i^*) \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{C})$.

Next, we apply the law of total expectation and variance to obtain mean and variance for the conditional normal distribution $p(w_i|\boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu})$. It follows that,

$$
\begin{aligned}
\text{E}[w_i|\boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu}] &= \text{E}\left[\text{E}(w_i|\tilde{\boldsymbol{u}}_i, \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu}) | \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu} \right] \\
&= \psi + \boldsymbol{\gamma}^\top \boldsymbol{z}_i + \boldsymbol{\theta}^\top \text{E}[\text{vec}(\tilde{\boldsymbol{u}}_i)|\boldsymbol{x}_i^*, \boldsymbol{\nu}] \\
&= \psi + \boldsymbol{\gamma}^\top \boldsymbol{z}_i + \boldsymbol{\theta}^\top \boldsymbol{m},
\end{aligned}
$$

and

$$
\begin{aligned}
\text{Var}[w_i|\boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu}] &= \text{Var}\left[\text{E}(w_i|\tilde{\boldsymbol{u}}_i, \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu}) | \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu} \right] \\
&\quad + \text{E}\left[\text{Var}(w_i|\tilde{\boldsymbol{u}}_i, \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu}) | \boldsymbol{x}_i^*, \boldsymbol{z}_i, \boldsymbol{\nu} \right] \\
&= \boldsymbol{\theta}^\top \text{Var}[\text{vec}(\tilde{\boldsymbol{u}}_i|\boldsymbol{x}_i^*, \boldsymbol{\nu}] \boldsymbol{\theta} + 1 \\
&= \boldsymbol{\theta}^\top \boldsymbol{C} \boldsymbol{\theta} + 1.
\end{aligned}
$$

Therefore, a regression model for w_i given $\boldsymbol{x}_i^*$, $\boldsymbol{z}_i$ and $\boldsymbol{\nu}$ is revealed as follows,

$$w_i = \psi - \phi\boldsymbol{\theta}^\top \boldsymbol{C}(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})^{1/2}\mathrm{vec}(\boldsymbol{\eta}) + \boldsymbol{\gamma}^\top\boldsymbol{z}_i + \phi\boldsymbol{\theta}^\top\boldsymbol{C}(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})^{1/2}(\boldsymbol{B}\otimes\boldsymbol{A})^\top\mathrm{vec}(\boldsymbol{x}_i^*) + e_i,$$

where $e_i \sim \mathcal{N}(0, 1+\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta})$. Because that $\Pr(o_i = 1) = \Pr(w_i > 0) = \Pr((1+\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta})^{-\frac{1}{2}}w_i > 0)$, the following regression model for o_i given latent matrix-valued covariate $\boldsymbol{x}_i^*$, scalar covariates $\boldsymbol{z}_i$, and model parameters $\boldsymbol{\nu}$ is obtained as

$$\Phi^{-1}[\Pr\{o_i = 1\}] = \varphi + \boldsymbol{\alpha}^\top\boldsymbol{z}_i + \boldsymbol{\beta}^\top\mathrm{vec}(\boldsymbol{x}_i^*), \tag{3.8}$$

where

$$\varphi = (1+\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta})^{-\frac{1}{2}}\left[\psi - \phi\boldsymbol{\theta}^\top\boldsymbol{C}(\boldsymbol{\xi}^{1/2}\otimes\boldsymbol{\lambda}^{1/2})\mathrm{vec}(\boldsymbol{\eta})\right],$$

$$\boldsymbol{\alpha} = (1+\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta})^{-\frac{1}{2}}\boldsymbol{\gamma},$$

$$\boldsymbol{\beta} = \phi(1+\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta})^{-\frac{1}{2}}(\boldsymbol{B}\otimes\boldsymbol{A})(\boldsymbol{\xi}^{1/2}\otimes\boldsymbol{\lambda}^{1/2})\boldsymbol{C}\boldsymbol{\theta}.$$

With this regression model between o_i and $\boldsymbol{x}_i^*$, it is followed that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ stay the same with orthogonal rotations of $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{\lambda}$, $\boldsymbol{\xi}$ and $\boldsymbol{\theta}$. Specifically, with $\boldsymbol{A}^*$, $\boldsymbol{B}^*$, $\boldsymbol{\theta}^*$, $\boldsymbol{\lambda}^{*1/2}$ and $\boldsymbol{\xi}^{*1/2}$ aforementioned, we only need to show that $\boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta}$ and $(\boldsymbol{B}\otimes\boldsymbol{A})(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})^{1/2}\boldsymbol{C}\boldsymbol{\theta}$ remain unchanged:

$$\begin{aligned}
\boldsymbol{\theta}^{*\top}\boldsymbol{C}^*\boldsymbol{\theta}^* &= \boldsymbol{\theta}^\top(\boldsymbol{H}\otimes\boldsymbol{G})\left[\boldsymbol{I} + \phi(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})(\boldsymbol{H}\otimes\boldsymbol{G})\right]^{-1}(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^\top(\boldsymbol{H}\otimes\boldsymbol{G})\left[(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)\left[\boldsymbol{I} + \phi(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})\right](\boldsymbol{H}\otimes\boldsymbol{G})\right]^{-1}(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^\top(\boldsymbol{H}\otimes\boldsymbol{G})(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)\left[\boldsymbol{I} + \phi(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})\right]^{-1}(\boldsymbol{H}\otimes\boldsymbol{G})(\boldsymbol{H}^\top\otimes\boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^\top\left[\boldsymbol{I} + \phi(\boldsymbol{\xi}\otimes\boldsymbol{\lambda})\right]^{-1}\boldsymbol{\theta} \\
&= \boldsymbol{\theta}^\top\boldsymbol{C}\boldsymbol{\theta}
\end{aligned} \tag{3.9}$$

and

$$
\begin{aligned}
&(\boldsymbol{B}^* \otimes \boldsymbol{A}^*)(\boldsymbol{\xi}^{*1/2} \otimes \boldsymbol{\lambda}^{*1/2})\boldsymbol{C}^*\boldsymbol{\theta}^* \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})(\boldsymbol{H} \otimes \boldsymbol{G}) \left[\boldsymbol{I} + \phi(\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})(\boldsymbol{H} \otimes \boldsymbol{G})\right]^{-1} (\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})(\boldsymbol{H} \otimes \boldsymbol{G}) \left[(\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\left[\boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})\right](\boldsymbol{H} \otimes \boldsymbol{G})\right]^{-1} (\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})(\boldsymbol{H} \otimes \boldsymbol{G})(\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\left[\boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})\right]^{-1}(\boldsymbol{H} \otimes \boldsymbol{G})(\boldsymbol{H}^\top \otimes \boldsymbol{G}^\top)\boldsymbol{\theta} \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})\left[\boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})\right]^{-1}\boldsymbol{\theta} \\
&= (\boldsymbol{B} \otimes \boldsymbol{A})(\boldsymbol{\xi}^{1/2} \otimes \boldsymbol{\lambda}^{1/2})\boldsymbol{C}\boldsymbol{\theta}
\end{aligned}
\tag{3.10}
$$

Therefore, the rotation invariance in the model parameters is negligible and our proposed JMME implies an identifiable regression model of o_i on the latent matrix covariates $\boldsymbol{x}_i^*$, leading to a unique correspondence between the coefficients for low-dimension features $\widetilde{\boldsymbol{u}}_i$ and the coefficients for $\boldsymbol{x}_i^*$.

Remark 2. As mentioned in chapter 1, our JMME is an extension of the work from Jiang et al. (2020) for adjusting meausrement errors in matrix variate regression. Therefore, for comparison purpose, we implemented their joint model (named as naiveJM) to illustrate the improvement of our framework. The naiveJM is directly carried out with the observed matrix covariates $\boldsymbol{x}_i$, assuming no measurement errors involved. In naiveJM, the low dimension features $\hat{\boldsymbol{u}}_i^*$ is extracted to approximate $\boldsymbol{x}_i$ by

$$\hat{\boldsymbol{u}}_i^* = \boldsymbol{E}^T\boldsymbol{x}_i\boldsymbol{F} + \boldsymbol{\epsilon}_i^*, \text{where } \mathrm{vec}(\boldsymbol{\epsilon}_i^*) \sim \mathcal{N}\left(\boldsymbol{0}, \phi^{*-1}I_{pq\times pq}\right)$$

with projection matrices $\boldsymbol{E}$, $\boldsymbol{F}$ such that $\boldsymbol{E}^\top\boldsymbol{E} = I_{p_0\times p_0}$, $\boldsymbol{F}^\top\boldsymbol{F} = I_{q_0\times q_0}$, and $p_0 < p$, $q_0 < q$. Meanwhile, the probit regression in naiveJM for $\hat{\boldsymbol{u}}_i^*$ is of the form

$$\Phi^{-1}(p[O_i = 1]) = \psi^* + \boldsymbol{\gamma}^{\top *}\boldsymbol{z}_i + \boldsymbol{\theta}^{*\top}\mathrm{vec}(\hat{\boldsymbol{u}}_i^*)$$

As a result, the naiveJM leads to the coefficients $\boldsymbol{\beta}^*$ for matrix-valued covari-

ates $\boldsymbol{x}_i$ as

$$\boldsymbol{\beta}^* = \phi^*(1 + \boldsymbol{\theta}^{*\top}\boldsymbol{C}^*\boldsymbol{\theta}^*)^{-\frac{1}{2}}(\boldsymbol{F} \otimes \boldsymbol{E})\boldsymbol{C}^*\boldsymbol{\theta}^*$$

with $\boldsymbol{C}^{*-1} = (\phi^* + 1)\boldsymbol{I}$. Due to the omission of measurement errors, the naiveJM tend to be subject to attenuation bias such that the estimation of coefficients are shrunk towards zero. Consequently, naiveJM becomes less efficient than our JMME in the context of measurement errors. This phenomenon is demonstrated in both chapter 4 and chapter 5.

Remark 3. Despite of naiveJM, we also considered the a two-stage model without measurement error correction (named as naiveTSM) as another baseline method to compare with our JMME (Jiang et al., 2020). We implemented this method as follows:

Stage1 - For the observed matrix covariates $\boldsymbol{x}_i$, MPCA (Lu, Plataniotis, & Venetsanopoulos, 2008) is applied on $\boldsymbol{x}_i$ to obtain an estimation of low dimension features, $\hat{\boldsymbol{u}}_i^* \in \mathcal{R}^{p_0 \times q_0}$. Similar as the MPCA formulation in naiveJM, $\hat{\boldsymbol{u}}_i^*$ approximates the observed $\boldsymbol{x}_i$ by

$$\hat{\boldsymbol{u}}_i^* \approx \boldsymbol{E}^T\boldsymbol{x}_i\boldsymbol{F}$$

with $p_0 < p$, $q_0 < q$, $\boldsymbol{E}^\top\boldsymbol{E} = I_{p_0 \times p_0}$, and $\boldsymbol{F}^\top\boldsymbol{F} = I_{q_0 \times q_0}$;

Stage2 - We further fitted a probit regression model for binary outcome o_i taking the form

$$\Phi^{-1}(p[O_i = 1]) = \psi^* + \boldsymbol{\gamma}^{\top *}\boldsymbol{z}_i + \boldsymbol{\theta}^{*\top}\text{vec}(\hat{\boldsymbol{u}}_i^*)$$

Thus, we can recover the coefficient matrix for observed $\boldsymbol{x}_i$ by

$$\boldsymbol{\beta}^* = \boldsymbol{E}\boldsymbol{\theta}^*\boldsymbol{F}^T.$$

Similar as naiveJM, the naiveTSM ignores measurement errors in observations, leading to attenuation bias in coefficients estiamtes. As a consequence, naiveTSM should be less efficient than our JMME either, in the context of

measurement errors. Such phenomenon is illustrated in both chapter 4 and chapter 5.

3.3 Hierarchical structure for JMME

This section illustrates the hierarchical structure for the proposed framework combining $\mathcal{M}_{ME}$ (sub-model 3.1), $\mathcal{M}_{MPCA\_decomp}$ (sub-model 3.5) and $\mathcal{M}_{out}$ (sub-model 3.7). With model parameters $\boldsymbol{\nu}$ defined in section 3.2, the full likelihood based on the complete data ($\boldsymbol{o}, \tilde{\boldsymbol{u}}, \boldsymbol{x}, \boldsymbol{x^*}$, and covariates $\boldsymbol{z}$) is given as the follows,

$$
\begin{aligned}
&f(\boldsymbol{o}, \tilde{\boldsymbol{u}}, \boldsymbol{x}, \boldsymbol{x}^*, |\boldsymbol{z}, \boldsymbol{\nu}) \\
&\quad = \prod_{i=1}^{n}\prod_{m=1}^{M}\left[p(\boldsymbol{o}_i|\tilde{\boldsymbol{u}}_i, \boldsymbol{\nu}, \boldsymbol{z}) \times p(\boldsymbol{x}_{im}|\boldsymbol{x}_i^*, \boldsymbol{\nu}) \times p(\boldsymbol{x}_i^*|\tilde{\boldsymbol{u}}_i, \boldsymbol{\nu}) \times p(\tilde{\boldsymbol{u}}_i)\right] \\
&\quad = \prod_{i=1}^{n}\prod_{m=1}^{M}\Bigg[p_i^{\mathcal{I}(o_i=1)}(1-p_i)^{\mathcal{I}(o_i=0)} \\
&\quad \times \frac{1}{(2\pi)^{pq/2}\left|\boldsymbol{\Sigma_i}\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)^\top \boldsymbol{\Sigma}_i^{-1}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)\right\} \\
&\quad \times \left(\frac{\phi}{2\pi}\right)^{1/2}\exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\left[\boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\right]\boldsymbol{B}^\top\|_F^2\right\} \\
&\quad \times \frac{1}{(2\pi)^{p_0q_0/2}}\exp\left\{-\frac{1}{2}\mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i)\right\}\Bigg] \\
&\quad = \prod_{i=1}^{n}\prod_{m=1}^{M}\Bigg[p_i^{\mathcal{I}(o_i=1)}(1-p_i)^{\mathcal{I}(o_i=0)} \\
&\quad \times \frac{1}{(2\pi)^{pq/2}\left|\boldsymbol{\Sigma_i}\right|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)^\top \boldsymbol{\Sigma}_i^{-1}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)\right\} \\
&\quad \times \left(\frac{\phi}{2\pi}\right)^{1/2}\exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2\right\} \\
&\quad \times \frac{1}{(2\pi)^{p_0q_0/2}}\exp\left\{-\frac{1}{2}\mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i)\right\}\Bigg]
\end{aligned}
\tag{3.11}
$$

where $\|\cdot\|_F^2$ represents the Frobenius norm, and $p_i = \Phi(\psi + \boldsymbol{\gamma}^\top\boldsymbol{z}_i + \boldsymbol{\theta}^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i))$. Figure 3.1 is a graphical illustration of our JMME that depicts

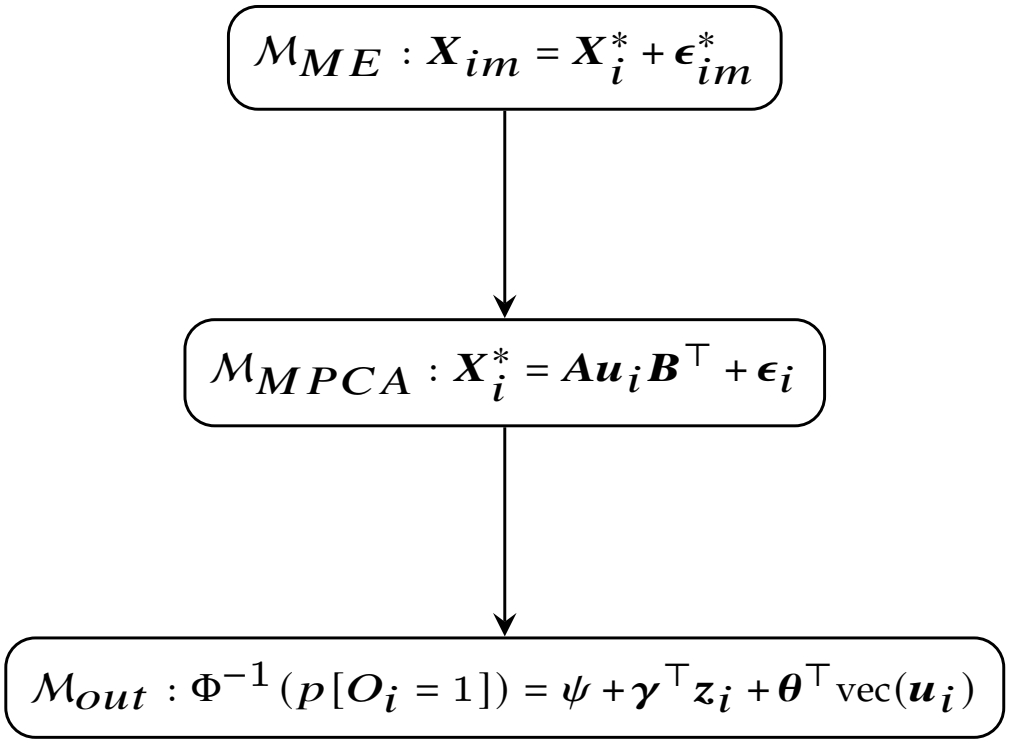

Figure 3.1: A graphical depiction of the hierarchical structure for proposed JMME

the structures among sub-models.

3.4 Prior distributions

In this section, we demonstrate the prior distributions chose for the parameters in our proposed JMME.

3.4.1 Prior distributions for $\mathcal{M}_{ME}$

For the parameters in sub-model 3.1 and 3.2, we followed the commonly used conjugate priors for Bayesian methods (Jiang et al., 2020). The priors are assumed as follows:

1. for each element $\widetilde{u_j}$ in $\mathrm{vec}(\widetilde{\boldsymbol{U}})$, $\widetilde{u_j} \sim \mathcal{N}(\mu_u, \sigma_u^2)$, $j = 1, \ldots, pr$

2. for each element $\widetilde{v_k}$ in $\mathrm{vec}(\widetilde{\boldsymbol{V}})$,$\widetilde{v_k} \sim \mathcal{N}(\mu_v, \sigma_v^2), k = 1, \ldots, qr$

3. $\mu_u, \mu_v \sim \mathcal{N}(0, \sigma_0^2)$ with $\sigma_0^2 = 10$

4. σ_u^2, σ_v^2, $\delta^2 \sim \text{IG}(a_0, b_0)$ with $a_0 = 0.1, b_0 = 0.1$

3.4.2 Prior distributions for $\mathcal{M}_{MPCA}$

In the sub-model 3.5, the prior distributions for projection matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are special. As suggested by the literature (Hoff, 2007; Jiang et al., 2020), because $\boldsymbol{A} \in \mathcal{R}^{p \times p_0}$ is orthonormal, it belongs to a space of all $p \times p_0$ orthonormal matrices named Stiefel manifolds with notation $\nu_{p_0,p}$. Hence, following the literature (Hoff, 2007; Jiang et al., 2020), we adopted a uniform distribution on $\nu_{p_0,p}$ as the prior distribution for $\boldsymbol{A}$. Such a uniform distribution leads to a distinct conditional distribution of $p(\boldsymbol{A}_{[,j]}|\boldsymbol{A}_{[,-j]})$, where $\boldsymbol{A}_{[,j]} \in \mathcal{R}^{p \times 1}$ represents j^{th} column of matrix $\boldsymbol{A}$ and $\boldsymbol{A}_{[,-j]} \in \mathcal{R}^{p \times (p_0 - 1)}$ refers to matrix $\boldsymbol{A}$ without the j^{th} column. It is proved that $p(\boldsymbol{A}_{[,j]}|\boldsymbol{A}_{[,-j]}) \overset{d}{=} \boldsymbol{N}_{A[,-j]}\boldsymbol{a}_j$, where $\boldsymbol{N}_{A[,-j]} \in \mathcal{R}^{p \times (p-(p_0-1))}$ represents an orthonormal basis for the null space of $\boldsymbol{A}_{[,-j]}$ (i.e., for a space such that $\{\boldsymbol{t} \in \mathcal{R}^{p_0-1} | \boldsymbol{A}_{[,-j]}\boldsymbol{t} = \boldsymbol{0}\}, \boldsymbol{N}_{A[,-j]}$ represents a basis whose columns have norm 1), and $\boldsymbol{a}_j \in \mathcal{R}^{(p-(p_0-1)) \times 1}$ follows a uniform distribution on a $\mathcal{R}^{p-(p_0-1)}$ unit sphere. This conditional distribution further allows to accelerating the Gibbs sampling step from full posterior distribution on $\boldsymbol{A}$ (Hoff, 2007; Jiang et al., 2020). The derivation is presented in appendix A.

Similarly, because $\boldsymbol{B} \in \mathcal{R}^{q \times q_0}$ belongs to a Stiefel manifold $\nu_{p_0,p}$, the prior distribution for $\boldsymbol{B}$ is chosen as a uniform distribution on $\nu_{p_0,p}$. As a result, with notations $\boldsymbol{B}_{[,k]} \in \mathcal{R}^{q \times 1}$ representing k^{th} column of matrix $\boldsymbol{B}$ and $\boldsymbol{B}_{[,-k]} \in \mathcal{R}^{q \times (q_0 - 1)}$ referring to matrix $\boldsymbol{B}$ without the k^{th} column, $p(\boldsymbol{B}_{[,k]}|\boldsymbol{B}_{[,-k]}) \overset{d}{=}$

$\boldsymbol{N}_{B[,-k]}\boldsymbol{b}_k$. Here, $\boldsymbol{N}_{B[,-k]} \in \mathcal{R}^{q\times(q-(q_0-1))}$ stands for an orthonormal basis for the null sapce of $\boldsymbol{B}_{[,-K]}$, and $\boldsymbol{b}_k \in \mathcal{R}^{(q-(q_0-1))\times 1}$ follows a uniform distribution on a $\mathcal{R}^{q-(q_0-1)}$ unit sphere. The aforementioned acceleration of Gibbs sampling procedure from $\boldsymbol{B}$'s full posterior distribution is also achieved by this conditional distribution (Hoff, 2007; Jiang et al., 2020).

For other parameters in sub-model 3.5, we again followed the commonly used conjugate priors (Jiang et al., 2020):

1. $\text{vec}(\boldsymbol{\eta}) \sim \mathcal{N}\left(0, \sigma_0^2\boldsymbol{I}\right)$ with $\sigma_0^2 = 10$
2. for each diagonal element $\lambda_l^{1/2}$ in $\boldsymbol{\lambda}$, $\lambda_l^{1/2} \sim \mathcal{N}\left(\mu_\lambda, \sigma_\lambda^2\right)$, $l = 1, \ldots, p0$
3. for each diagonal element $\xi_l^{1/2}$ in $\boldsymbol{\xi}$, $\xi_h^{1/2} \sim \mathcal{N}\left(\mu_\xi, \sigma_\xi^2\right)$, $h = 1, \ldots, q0$
4. $\mu_\lambda, \mu_\xi \sim \mathcal{N}\left(0, \sigma_0^2\right)$ with $\sigma_0^2 = 10$
5. σ_λ^2, σ_ξ^2, $1/\phi \sim \text{IG}(a_0, b_0)$ with $a_0 = 0.1, b_0 = 0.1$

3.4.3 Prior distributions for $\mathcal{M}_{out}$

For the parameters in sub-model (3.7), similar as the work of Jiang et al. (2020), we defined the priors are follows:

1. $(\psi, \boldsymbol{\theta}^\top, \boldsymbol{\gamma}^\top)^\top \sim \mathcal{N}\left(0, \sigma_0^2\boldsymbol{I}\right)$ with $\sigma_0^2 = 10$

3.5 Gibbs Sampling Procedure

In this section, we described the Gibbs sampling procedure from model parameters' posterior distributions (also known as full conditional distributions) (Jiang et al., 2020). For the ease of derivation, we assume the rank r for the low rank structure in $\mathcal{M}_{low}$ (sub-model 3.2) as $r = 1$. The complete derivation is provided in appendix A.

1. For elements in $\widetilde{\boldsymbol{\Sigma}}_i = \begin{pmatrix} \sigma^2_{i11} & \dots & \sigma^2_{i1q} \\ \vdots & \ddots & \vdots \\ \sigma^2_{ip1} & \dots & \sigma^2_{ipq} \end{pmatrix}$, update $[\log(\sigma^2_{ijk})|\cdot] \sim$ the distribution with following kernel:

$$\exp\left\{-\frac{1}{2}\frac{(\log(\sigma^2_{ijk}) - \widetilde{u}_j\widetilde{v}_k)^2}{\delta^2}\right\} \times \prod_{m=1}^{M}\left[\left(\exp(\log(\sigma^2_{ijk}))\right)^{-1/2}\exp\left\{-\frac{1}{2}\frac{(\boldsymbol{x}_{imjk} - \boldsymbol{x}^*_{ijk})^2}{\exp(\log(\sigma^2_{ijk}))}\right\}\right],$$

by metropolis sampling, $i \in \{1, \dots, n\}, j \in \{1, \dots, p\}, k \in \{1, \dots, q\}$.

2. For elements in $\widetilde{\boldsymbol{U}} = \begin{pmatrix} \widetilde{u}_1 \\ \vdots \\ \widetilde{u}_p \end{pmatrix}$, update $[\widetilde{u}_j|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = \left(\frac{1}{\sigma^2_u} + \frac{n\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}}}{\delta^2}\right)^{-1},$$
$$M = E\left(\frac{\mu_u}{\sigma^2_u} + \sum_{i=1}^{n}\frac{\widetilde{\boldsymbol{V}}^\top\log(\boldsymbol{\Sigma}_{i[j,]})^\top}{\delta^2}\right).$$

3. Update $[\mu_u|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = (\frac{1}{\sigma^2_0} + \frac{p}{\sigma^2_u})^{-1},$$
$$M = E\left(\sum_{j=1}^{p}\frac{\widetilde{u}_j}{\sigma^2_u}\right).$$

4. Update $[(\sigma^2_u)^{-1}|\cdot] \sim \text{Gamma}\left\{a_0 + p/2, b_0 + \frac{1}{2}\sum_{j=1}^{p}(\widetilde{u}_j - \mu_u)^2\right\}$.

5. For elements in $\widetilde{\boldsymbol{V}} = \begin{pmatrix} \widetilde{v}_1 \\ \vdots \\ \widetilde{v}_q \end{pmatrix}$, update $[\widetilde{v}_k|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = \left(\frac{1}{\sigma^2_v} + \frac{n\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{U}}}{\delta^2}\right)^{-1},$$
$$M = E\left(\frac{\mu_v}{\sigma^2_v} + \sum_{i=1}^{n}\frac{\widetilde{\boldsymbol{U}}^\top\log(\boldsymbol{\Sigma}_{i[,k]})}{\delta^2}\right).$$

6. Update $[\mu_v|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = (\frac{1}{\sigma^2_0} + \frac{q}{\sigma^2_v})^{-1},$$
$$M = E\left(\sum_{k=1}^{q}\frac{\widetilde{v}_k}{\sigma^2_v}\right).$$

7. Update $[(\sigma_v^2)^{-1}|\cdot] \sim \text{Gamma}\left\{a_0 + q/2, b_0 + \frac{1}{2}\sum_{k=1}^{q}(\widetilde{v}_k - \mu_v)^2\right\}$.

8. Update $[(\delta^2)^{-1}|\cdot] \sim \text{Gamma}\left\{a_0 + npq/2, b_0 + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{q}(\log(\sigma_{ijk}^2) - \widetilde{u}_j\widetilde{v}_k)^2\right\}$.

9. Update $[\text{vec}(\boldsymbol{\eta})|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$, where

$$\boldsymbol{E} = (\frac{1}{\sigma_0^2} + n\phi)^{-1}\boldsymbol{I},$$
$$\boldsymbol{M} = \boldsymbol{E}\left(\sum_{i=1}^{n}\phi\left[(\boldsymbol{B}\otimes\boldsymbol{A})^\top\text{vec}(\boldsymbol{x}_i^*) - (\boldsymbol{\xi}\otimes\boldsymbol{\lambda})^{1/2}\text{vec}(\tilde{\boldsymbol{u}}_i)\right]\right).$$

10. For diagonal elements in $\boldsymbol{\lambda}^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_{p_0}^{1/2})$, update $[\lambda_l^{1/2}|\cdot] \sim \mathcal{N}(M, E)$, for $l = 1, \ldots, p_0$, where

$$E = \left(\frac{1}{\sigma_\lambda^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[l,]}\boldsymbol{\xi}\tilde{\boldsymbol{u}}_{i[l,]}^\top\right)^{-1},$$
$$M = E\left(\frac{\mu_\lambda}{\sigma_\lambda^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[l,]}\boldsymbol{\xi}^{1/2}\left[\boldsymbol{A}_{[,l]}^\top\boldsymbol{x}_i^*\boldsymbol{B} - \boldsymbol{\eta}_{[l,]}\right]^\top\right).$$

11. Update $[\mu_\lambda)|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = (\frac{1}{\sigma_0^2} + \frac{p_0}{\sigma_\lambda^2})^{-1},$$
$$M = E\frac{\sum_{l=1}^{p_0}\lambda_l^{1/2}}{\sigma_\lambda^2}.$$

12. Update $[(\sigma_\lambda^2)^{-1}|\cdot] \sim \text{Gamma}\left\{a_0 + \frac{p_0}{2}, b_0 + \frac{\sum_{l=1}^{p_0}\left(\lambda_l^{1/2} - \mu_\lambda\right)^2}{2}\right\}$.

13. For diagonal elements in $\boldsymbol{\xi}^{1/2} = \text{diag}(\xi^{1/2}, \ldots, \xi{q_0}^{1/2})$, update $[\xi_s^{1/2}|\cdot] \sim \mathcal{N}(M, E)$, for $s = 1, \ldots, q_0$, where

$$E = \left(\frac{1}{\sigma_\xi^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[,s]}^\top\boldsymbol{\lambda}\tilde{\boldsymbol{u}}_{i[,s]}\right)^{-1},$$
$$M = E\left(\frac{\mu_\xi}{\sigma_\xi^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[,s]}^\top\boldsymbol{\lambda}^{1/2}\left[\boldsymbol{A}^\top\boldsymbol{x}_i^*\boldsymbol{B}_{[,s]} - \boldsymbol{\eta}_{[,s]}\right]\right).$$

14. Update $[\mu_\xi)|\cdot] \sim \mathcal{N}(M, E)$, where

$$E = (\frac{1}{\sigma_0^2} + \frac{q_0}{\sigma_\xi^2})^{-1},$$
$$M = E\frac{\sum_{s=1}^{q_0} \xi_s^{1/2}}{\sigma_\xi^2}$$

15. Update $[(\sigma_\xi^2)^{-1}|\cdot] \sim \text{Gamma}\left\{a_0 + \frac{q_0}{2}, b_0 + \frac{\sum_{s=1}^{q_0}\left(\xi_s^{1/2} - \mu_\xi\right)^2}{2}\right\}$.

16. Update $[\text{vec}(\tilde{\boldsymbol{u}}_i)|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$, for $i = 1, \ldots, n$, where

$$\boldsymbol{E} = (\boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda}) + \boldsymbol{\theta}\boldsymbol{\theta}^\top)^{-1},$$
$$\boldsymbol{M} = \boldsymbol{E}\left(\phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2}\left[\boldsymbol{B} \otimes \boldsymbol{A}\right)^\top \text{vec}(\boldsymbol{x}_i^*) - \text{vec}(\boldsymbol{\eta})\right] + \boldsymbol{\theta}\tilde{w}_i\right),$$

and $\tilde{w}_i = w_i - \psi - \boldsymbol{\gamma}^\top \boldsymbol{z}_i$.

17. Calculate the $\boldsymbol{u}_i = \boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}$, and update the following parameters with $\boldsymbol{u}_i$.

18. Update $[\phi|\cdot] \sim \text{Gamma}\{a_0 + npq/2, b_0 + \sum_{i=1}^n \|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2/2\}$.

19. Update $[\boldsymbol{A}_{[,j]}|\cdot] \overset{d}{=} \boldsymbol{N}_{A[,-j]}\boldsymbol{a}_j$, for $j \in \{1, \ldots, p_0\}$, where $\boldsymbol{a}_j \sim \text{vMF}(\boldsymbol{\eta}^A)$ with

$$\boldsymbol{\eta}^A = \phi \boldsymbol{N}_{A[,-j]}^\top \sum_{i=1}^n \sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]} \boldsymbol{x}_i^{*-j} \boldsymbol{B}_{[,k]},$$
$$\boldsymbol{x}_i^{*-j} = \boldsymbol{x}_i^* - \sum_{j' \neq j} \sum_{k=1}^{q_0} \boldsymbol{A}_{[,j']} \boldsymbol{u}_{i[j',k]} \boldsymbol{B}_{[,k]}^\top.$$

20. Update $[\boldsymbol{B}_{[,k]}|\cdot] \overset{d}{=} \boldsymbol{N}_{B[,-k]}\boldsymbol{b}_j$, for $k \in \{1, \ldots, q_0\}$, where $\boldsymbol{b}_j \sim \text{vMF}(\boldsymbol{\eta}^B)$ with

$$\boldsymbol{\eta}^B = \phi \boldsymbol{N}_{B[,-k]}^\top \sum_{i=1}^n \sum_{j=1}^{p_0} \boldsymbol{u}_{i[j,k]} \tilde{\boldsymbol{x}}_i^{*-k} \boldsymbol{A}_{[,j]},$$
$$\tilde{\boldsymbol{x}}_i^{*-k} = \boldsymbol{x}_i^\top - \sum_{k' \neq k} \sum_{j=1}^{p_0} \boldsymbol{B}_{[,k']} \boldsymbol{u}_{i[j,k']} \boldsymbol{A}_{[,j]}^\top.$$

21. Update $[(\text{vec}(\boldsymbol{x}_i^*)|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$,for $i \in \{1, \ldots, n\}$, where

$$\boldsymbol{E} = (\phi \boldsymbol{I} + M\boldsymbol{\Sigma}_i^{-1})^{-1},$$
$$\boldsymbol{M} = \boldsymbol{E}(\phi(\boldsymbol{B} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{u}_i) + \sum_{m=1}^{M} \boldsymbol{\Sigma}_i^{-1}\text{vec}(\boldsymbol{x}_{im})).$$

22. Update w_i, for $i \in \{1, \ldots, n\}$, we have

$$[w_i|\cdot] \sim \mathcal{N}(\psi + \boldsymbol{\gamma}^T \boldsymbol{z}_i + \boldsymbol{\theta}^\top \text{vec}(\tilde{\boldsymbol{u}}_i), 1)\mathcal{I}(w_i > 0) \text{ if } o_i = 1$$
$$[w_i|\cdot] \sim \mathcal{N}(\psi + \boldsymbol{\gamma}^T \boldsymbol{z}_i + \boldsymbol{\theta}^\top \text{vec}(\tilde{\boldsymbol{u}}_i), 1)\mathcal{I}(w_i < 0) \text{ if } o_i = 0$$

23. Update $[(\psi, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$ where

$$\boldsymbol{E} = (\frac{1}{\sigma_0^2}\boldsymbol{I} + \sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i \tilde{\boldsymbol{z}}_i^\top)^{-1},$$
$$\boldsymbol{M} = \boldsymbol{E}\sum_{i=1}^{n} w_i \tilde{\boldsymbol{z}}_i \text{ with } \tilde{\boldsymbol{z}}_i = (1, \boldsymbol{z}_i^\top, \text{vec}(\tilde{\boldsymbol{u}}_i)^\top)^\top.$$

3.6 Selection of Dimensionality

It should be noticed that, before applying our proposed JMME framework, the dimension for the extracted features $\boldsymbol{u}_i \in \mathcal{R}^{p_0 \times q_0}$ in sub-model 3.3 needs to be predetermined. To select the optimal (p_0, q_0), we considered the Akaike information criterion (AIC) for the conditional likelihood $f(O|\boldsymbol{X}^*, \nu)$ because we aim to examine the association between responses and latent matrix covariates $\boldsymbol{X}^*$.

To calculate the desired AIC, we first derive the likelihood $\mathcal{L}(\boldsymbol{o}|\boldsymbol{x}^*, \nu)$. Implied by model (3.8), the likelihood takes the form

$$\begin{aligned} \mathcal{L}(\boldsymbol{o}|\boldsymbol{x}^*, \nu) &= \prod_{i=1}^{n} p(\boldsymbol{o}_i|\boldsymbol{x}_i^*, \nu) \\ &= \prod_{i=1}^{n} \left[p_i^{\mathcal{I}(o_i=1)}(1-p_i)^{\mathcal{I}(o_i=0)} \right] \end{aligned} \tag{3.12}$$

where $p_i = \Phi(\varphi + \boldsymbol{\alpha}^\top \boldsymbol{z}_i + \boldsymbol{\beta}^\top \text{vec}(\boldsymbol{x}_i^*))$.

Table 3.1: A list of model parameters in likelihood $\mathcal{L}(\boldsymbol{o}|\boldsymbol{x}^*)$ and their corresponding dimensionality

Model Parameters in $\mathcal{L}(\boldsymbol{o}\|\boldsymbol{x}^*)$	Dimensionality
$\boldsymbol{A}$	$p \times p_0$
$\boldsymbol{B}$	$q \times q_0$
$\boldsymbol{\eta}$	$p_0 \times q_0$
$\boldsymbol{\lambda}^{1/2}$	p_0
$\boldsymbol{\xi}^{1/2}$	q_0
ϕ	1
ψ	1
$\boldsymbol{\gamma}$	p_z
$\boldsymbol{\theta}$	$p_0 \times q_0$

Therefore, the AIC for conditional likelihood $f(\boldsymbol{o}|\boldsymbol{x})$ is of the form:

$$AIC = -2\log(\hat{\mathcal{L}}(\boldsymbol{o}|\boldsymbol{x}^*)) + 2k, \tag{3.13}$$

where k represents the totoal number of parameters in the likelihood (model 3.12) and $\hat{\mathcal{L}}$ is the posterior mean of likelihood 3.12 from MCMC chain. In terms of total number of parameters, Table 3.1 tells all the parameters used in the $o|\boldsymbol{x}^*$ regression model along with their dimensions needed for likelihood calculation. Consequently,

$$k = p \times p_0 + q \times q_0 + p_0 \times q_0 + p_0 + q_0 + 1 + 1 + p_z + p_0 \times q_0.$$

Chapter 4

Simulation Study

To examine the correctness and effectiveness of our proposed framework, a simulation study was implemented in various trials with different matrix surrogates dimensionality setups. The study aims to evaluate the performance of our JMME framework from the perspective of correctly estimating coefficients for latent matrix covariates. As baselines, the naiveTSM and naiveJM methods introduced in Chapter 3 were also implemented so as to highlight the improvement of our proposed method.

4.1 Simulation Procedure

4.1.1 Data setup

To test the three methods, this study followed the procedure described in Jiang et al. (2020). Each simulation trial was coupled with both a fixed low-dimensional structure ($(p_0, q_0) = (2, 2)$) and a fixed number of repeatedly observed matrix-valued covariates ($M = 2$ in the measurement error sub-model (3.1)). For different choices of $n \in \{50, 100\}$ and $(p, q) \in \{(5, 5), (10, 10), (15, 15)\}$, totally 600 simulation trials were performed on each of the three methods. For each simulation trial, a simulated synthetic dataset with binary outcomes was generated through the following steps:

1) With fixed $(p_0,\, q_0) = (2,\, 2)$, the elements in $\text{vec}(\tilde{\boldsymbol{u}}_i)$ are generated inde-

pendently from a uniform distribution on (-1, 1);

2) Generate z_i from a uniform distribution on (-0.5, 0.5);

3) With $\mathbf{1}^\top = (1, \ldots, 1)^\top$, o_i is generated from a Bernoulli distribution with $p = 0.2z_i + \mathbf{1}^\top \text{vec}(\tilde{\boldsymbol{u}}_i)$;

4) The elements in $\boldsymbol{\eta}$ are generated independently from a uniform distribution on (-2, 2);

5) The diagonal elements of $\boldsymbol{\lambda}^{1/2}$ are generated independently from a uniform distribution on (-1, 1), and the off-diagonal elements are 0;

6) The diagonal elements of $\boldsymbol{\xi}^{1/2}$ are generated independently from a uniform distribution on (-1, 1), and the off-diagonal elements are 0;

7) The corresponding $\boldsymbol{u}_i$ can be calculated as $\boldsymbol{u}_i = \boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}$ so that $\text{vec}(\boldsymbol{u}_i) \sim \mathcal{N}(\text{vec}(\boldsymbol{\eta}), \boldsymbol{\xi} \otimes \boldsymbol{\lambda})$ holds;

8) For each $\{(o_i, z_i, \boldsymbol{u}_i) : i = 1, \ldots, n\}$ generated from 1) - 7), we further simulated $(\boldsymbol{x}_i^* : i = 1, \ldots, n)$ and $(\boldsymbol{x}_{im} : i = 1, \ldots, n, m = 1, \ldots, M)$ corresponding to different pairs of (p, q) , with fixed $M = 2$, as follows:

   a) Generate $\boldsymbol{A} \sim \text{uniform}(\nu_{p,p0})$, a uniform distribution on the Stiefel manifold with dimension (p, p_0), using rstiefel package in R software;

   b) Generate $\boldsymbol{B} \sim \text{uniform}(\nu_{q,q0})$, a uniform distribution on the Stiefel manifold with dimension (p, p_0), using rstiefel package in R software;

   c) Generate $\boldsymbol{x_i}^* = \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top + \boldsymbol{\epsilon}_i$, with $\text{vec}(\boldsymbol{\epsilon}_i) \sim \mathcal{N}(\mathbf{0}, 0.2^2\boldsymbol{I}_{pq\times pq})$;

d) The elements of $\widetilde{\boldsymbol{U}}$ are generated independently from a uniform distribution on (-1, 1);

e) The elements of $\widetilde{\boldsymbol{V}}$ are generated independently from a uniform distribution on (-1, 1);

f) Generate $\log(\widetilde{\boldsymbol{\Sigma}}_i) = \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^\top + \widetilde{\boldsymbol{\epsilon}}_i$ with $\text{vec}(\widetilde{\boldsymbol{\epsilon}}_i) \sim \mathcal{N}(0, 0.4\boldsymbol{I})$;

g) For $m \in \{1, 2\}$, generate $\boldsymbol{x}_{im} = \boldsymbol{x}_i^* + \boldsymbol{\epsilon}_{im}^*$ under the condition that $\text{vec}(\boldsymbol{\epsilon}_{im}^*) \sim \mathcal{N}\left(0, \text{diag}(\text{vec}(\widetilde{\boldsymbol{\Sigma}}_i))\right)$;

4.1.2 Evaluation Criteria

For each simulated dataset, we applied the Gibbs sampling procedure described in section 3.5. The posterior samples of model parameters were obtained from the MCMC chains by keeping every 5th draw from 10000 iterations after a burn-in period of 20,000 iterations (Jiang et al., 2020). In order to evaluate the estimation performance of the coefficient $\boldsymbol{\beta}$ for $\text{vec}(\boldsymbol{x})_i^*$, we calculated the root mean squared error (RMSE) for the estimates. More precisely, under each choice of (p, q), for the t^{th} simulated dataset, $t = 1, \ldots, 100$, we defined the RMSE as

$$RMSE^{(t)} = \sqrt{\frac{1}{pq}\|\widehat{\boldsymbol{\beta}}^{(t)} - \boldsymbol{\beta}^{(t)}\|_F^2}, \tag{4.1}$$

where $\boldsymbol{\beta}^{(t)}$ is the true coefficient matrix for the unobserved true matrix covariates in $p \times q$ dimension and $\widehat{\boldsymbol{\beta}}^{(t)}$ is its posterior mean estimate obtained from the MCMC chains. As proved at section 2.8, the formulation for estimating $\boldsymbol{\beta}$ from model parameters follows the equation 3.8.

In addition, we compared our JMME framework with the naiveJM and naiveTSM methods. All three methods were implemented with the same num-

ber of low-dimensional features ($(p_0, q_0) = (2, 2)$). It is noted that our simulation procedure generated 2 repeatedly observed error-prone covariates in the matrix form $\{\boldsymbol{x}_{im}, m = 1, 2\}$. Given the fact that the naiveJM and naiveTSM methods cannot handle repeated measurements (M=2), we set them up with the following three observation-setup options:

Option 1) Assuming that $\boldsymbol{x}_{i1}$ is the only measured matrix covariates, datasets of the form $\{o_i, z_i, \boldsymbol{x}_{i1}\}$ are used for the two baseline methods;

Option 2) Assuming that $\boldsymbol{x}_{i2}$ is the only measured matrix covariates, datasets of the form $\{o_i, z_i, \boldsymbol{x}_{i2}\}$ are used for the two baseline methods;

Option 3) Use the mean of the two observations, $\frac{1}{2}\sum_{m=1}^{2}\boldsymbol{x}_{im}$, as an estimation of accurately measured matrix covariates. Hence $\{o_i, z_i, \frac{1}{2}\sum_{m=1}^{2}\boldsymbol{x}_{im}\}$ are used for the two baseline methods.

For each option, we implemented the two baseline methods as follows. For thenaiveJM, we used the same prior distributions for all model parameters as JMME's. Moreover, to make the results comparable, the MCMC chains for their method were also obtained with the same number of iterations as ours, and the posterior samples of model parameters were accessed by keeping every 5^{th} draw from 10,000 iterations after 20,000 burn-in iterations. For each saved posterior sample, the coefficients for observed matrix covariates were estimated through the formulation described in remark 3.2. As a result, the posterior mean of coefficients was obtained by averaging out the posterior estimates of coefficients, and the RMSE was calculated accordingly.

For naiveTSM, we first applied MPCA on the matrix-valued covariates with the rTensor package in R software (Li, Bien, & Wells, 2018). Then, a probit regression model for the outcome o_i was implemented with scalar covariates and vectorized low-dimensional features extracted from MPCA. The coeffi-

cients for observed matrix covariates were calculated as described in remark 3.2 and the RMSE was obtained accordingly.

4.2 Simulation Results

Table 4.1, 4.2 and 4.3 display the average RMSE of coefficient estimates $\widehat{\boldsymbol{\beta}} \in \mathcal{R}^{pq\times 1}$ in binary response prediction over the 100 simulated datasets, with $n = 50$. Similarly, table 4.4, 4.5 and 4.6 display the cases when $n = 100$ for three methods. Overall, for all three methods, there exists a decreasing trend among the average RMSEs when the sample size increased from n=50 to n=100, leading to a more precise estimates of the coefficients $\boldsymbol{\beta}$ when more samples are used in the models. Among all the results, our proposed JMME achieved the smallest RMSE within all scenarios, suggesting its ability of providing more accurate estimates of the associations between binary outcome and latent matrix-valued covariates in $p \times q$ dimensional space.

According to Table 4.3 and Table 4.6, among all levels of (p, q), the naiveTSM method performed the worst in estimating the coefficients $\boldsymbol{\beta}$. This observation fits our expectation because, on the one hand, the measurement errors in the observed matrix covariates could deteriorate the estimates of coefficients for latent matrix covariates without explicit correction. On the other hand, the MPCA procedure could be subject to estimation errors when extracting lower-dimensional features $\hat{\boldsymbol{u}}_i^*$. Such estimation errors could further negatively impact the estimates of coefficients.

As for the naiveJM method, it slightly improved the estimation performance, compared to the naiveTSM model. This is because that the naiveJM explicitly modeled the estimation error during the MPCA procedure so that the estimation bias is avoided (Jiang et al., 2020). However, its drawback lies in the situation when measurement errors exist: it could still induce the bias in coefficient estimation due to the lack of measurement error correction. This

RMSE for Proposed JMME Framework with n = 50

	$(p,q)=(5,5)$	$(p,q)=(10,10)$	$(p,q)=(15,15)$
Use repeated measures	0.079	0.043	0.031

Table 4.1: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with n = 50 under different (p,q) scenarios and observation-setup options, using the proposed JMME framework.

RMSE for NaiveJM Method with $n = 50$

	$((p,q)=(5,5)$	$(p,q)=(10,10)$	$(p,q)=(15,15)$
Option 1	0.435	0.219	0.146
Option 2	0.427	0.218	0.146
Option 3	0.423	0.218	0.145

Table 4.2: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with $n = 50$, different (p,q) scenarios and observation-setup options, using naiveJM method.

explains the substantial reduction in RMSEs provided by the proposed JMME model on the basis of the naiveJM model.

When comparing among the observation-setup options, using the average of two observations as an estimate of true latent matrix covariates slightly outperform others in terms of the coefficients estimation results in naiveTSM. The finding is consistent for naiveJM: using the average of two observations leads to a minor reduction in RMSEs compared to naiveJM using single $\boldsymbol{x}_i m$. This is partially due to the fact that taking average may reduce the variability in the original matrix covariates, so that the assumed additive measurement errors could be diminished. However, such reduction is insufficient for measurement error correction, and the RMSEs are still larger than those obtained by our proposed JMME model, as shown in Table 4.1.

RMSE for NaiveTSM Method with $n = 50$

	$(p,q) = (5,5)$	$(p,q) = (10,10)$	$(p,q) = (15,15)$
Option 1	0.656	0.307	0.207
Option 2	0.624	0.306	0.211
Option 3	0.610	0.301	0.204

Table 4.3: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with $n = 50$, under different (p, q) scenarios and observation-setup options, using naiveTSM method.

RMSE for Proposed JMME Framework with n = 100

	$(p,q) = (5,5)$	$(p,q) = (10,10)$	$(p,q) = (15,15)$
Use repeated measures	0.071	0.039	0.028

Table 4.4: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with n = 100 under different (p, q) scenarios and observation-setup options, using the proposed JMME framework.

RMSE for NaiveJM Method with $n = 100$

	$((p,q) = (5,5)$	$(p,q) = (10,10)$	$(p,q) = (15,15)$
Option 1	0.395	0.201	0.134
Option 2	0.394	0.201	0.134
Option 3	0.383	0.199	0.133

Table 4.5: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with $n = 100$, under different (p, q) scenarios and observation-setup options, using naiveJM method.

RMSE for NaiveTSM Method with n=100

	$(p,q) = (5,5)$	$(p,q) = (10,10)$	$(p,q) = (15,15)$
Option 1	0.397	0.201	0.134
Option 2	0.397	0.201	0.134
Option 3	0.386	0.200	0.134

Table 4.6: The average of the root mean squared errors (RMSEs) of the estimated coefficients $\hat{\boldsymbol{\beta}}$ across 100 simulated datasets with $n = 100$, under different (p, q) scenarios and observation-setup options, using naiveTSM method.

Chapter 5

Case Study: Antidepressant Response Prediction using EEG Data

In this chapter, we applied the proposed JMME framework on the error-prone EEG data to predict the antidepressant outcomes and explore the association between latent EEG data and treatment outcomes. To illustrate that our model is able to successfully adjust for the potential measurement errors in the EEG data and restore the associations from attenuation bias, we compared our model with the joint modeling without measurement error correction (naiveJM) described in remark 3.2 and the naive two-stage modeling method (naiveTSM) described in remark 3.2.

5.1 Data Description

The case study used two types of data sources collected from a group of 80 patients with major depressive disorder: error-prone CSD-EEG data collected through physical equipment, and the patients' responses to an antidepressant treatment in the class of selective serotonin reuptake inhibitors. The goal of the case study is to study the association between latent EEG (i.e., the unobserved EEG corrected for measurement errors) and patients' treatment responses. The CSD-EEG was obtained through a procedure of collecting continuous

scalp EEG data measured from physical electrodes and transform them into matrices using the current source density (CSD) analysis. The rows of the resulting matrices correspond to the the positions of electrodes placed over brain areas and the columns are frequencies ranging from 0.25Hz to 74.75 Hz, with 0.25 Hz resolution (Jiang et al., 2020). Suggested by prior theoretical and empirical studies (Tenke et al., 2017), we made the hypothesis that the CSD-EEG data measures of power spectra in the theta-frequency band (4-7 Hz) at the posterior brain regions, under eyes open and eyes closed conditions, are the most related to the patients' response to an antidepressant treatment. Based on this assumption, we acquired for each patient two 14×13 matrices of CSD-EEG measures: The first one corresponds to the EEG recorded under eyes-open condition (denoted as $\boldsymbol{x}_{i1}$ for patient i) and the second matrix corresponds to eyes-closed condition (denoted as $\boldsymbol{x}_{i2}$ for patient i). The columns of each matrix correspond to 14 electrodes (P9, P10, P7, P8, P5, P6, PO7, PO8, PO3, PO4, O1, O2, POZ, OZ) at the posterior brain region and the rows correspond to theta-frequency band (4-7Hz). As suggested in the literature, the CSD-EEG matrices are log-transformed for normalization purposes (Jiang et al., 2020). As described in Chapter 1, since the CSD-EEG data cannot accurately access the patients' brain activities due to the potential measurement errors, we further hypothesized additive measurement errors (as proposed in $\mathcal{M}_{ME}$ in our joint model) among CSD-EEG baseline measures, and let $\boldsymbol{x}_i^*$ represents the latent EEG covariates, estimated from the two sets of CSD-EEG under eyes-open and eyes-closed conditions.

In addition, we also consider the effect of gender and depression chronicity when predicting treatment outcomes. The gender factor takes value 1 for a patient being female and 0 for a patient being male; the chronicity takes value 1 if a patient is being depressed for at least 24 months during the past 4 to 5 years and 0 otherwise. These two scalar covariates are denoted as $\boldsymbol{z}_i$ for

subject i.

To formulate the treatment outcomes, the 17-item Hamilton Depression Rating Scale (HAMD-17) at baseline for each patient was recorded to measure the severity of depression. Ranging from 0 to 52, the higher the HAMD-17 score is, the more severe the depression is for a patient. These scores were recorded at weeks 1, 2, 3, 4, 6, and 8 during the treatment. After 8 weeks, the treatment outcome, denoted as o_i, is a binary variable such that $o_i = 1$ (the patient favorably responded to the SSRI) if the HAMD-17 score is reduced by 50% or more (Israel, 2006) and $O_i = 0$ otherwise (Jiang et al., 2020). Among 80 subjects, 46 were with $O_i = 1$ and the remaining 34 were with $O_i = 0$.

In summary, the complete data for our motivating study consists of two sets of contaminated CSD-EEG baseline signals $\{(\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2})\}_i^n$ in 14×13 matrix form, two accurately observed scalar covariates $\{\boldsymbol{x}_i\}_i^n$ and the binary treatment outcome $\{o_i\}_i^n$.

5.2 Implementation

We applied our proposed JMME model with the prior distributions specified in chapter 3.4. To obtain the posterior samples for model parameters, we ran five MCMC chains, each consists of 20000 discarded burn-in samples and consecutive 20000 iterations on a high performance computing cluster, and keep every 5^{th} sample as a posterior sample. In total, the Gibbs sampling procedure generated 2000 posterior sample points for association analysis. Note that the dimensionality of extracted features in lower $p_0 \times q_0$ space needs to be decided for applying our JMME framework. As described in chapter 3, we calculated the averaged AIC for all combinations of (p_0, q_0) ranging from 1 to 5 and the smallest AIC leads to the desired dimensionality for JMME. Table 5.1 shows the resulting AIC for all fitting models with smallest AIC bolded. The AIC values first decreased when p_0 and q_0 changed from 1 to 2, and then increased

Table 5.1: AIC under different combination of (p0, q0) for proposed joint model

p0/q0	1	2	3	4	5
1	180.621	212.574	244.547	276.904	308.904
2	214.86	**171.685**	258.245	294.706	337.367
3	248.4	211.948	295.719	349.542	392.306
4	282.495	251.372	341.395	391.907	442.027
5	316.311	292.28	378.798	436.638	488.283

when p_0 and q_0 increased from 2 to 5. Overall, there exists a U-shape trend in the AIC for all fitting JMME models, and the smallest AIC was obtained at $(p_0, q_0) = (2, 2)$. Therefore, the estimation of coefficients $\boldsymbol{\beta}$ for latent EEG in antidepressant response prediction is calculated using $(p_0, q_0) = (2, 2)$ in the equation derived in chapter 3.

To illustrate the improvement of JMME method and highlight the significance of measurement errors in CSD-EEG baseline data, the naiveJM and naiveTSM were implemented to compare with our JMME method. According to Table 5.1 , JMME is best performed when $(p_0, q_0) = (2, 2)$. Therefore, we applied the naiveJM and the naiveTSM with $(p_0, q_0) = (2, 2)$. Since the two baseline methods cannot handle the repeated CSD-based EEG measures, similar to the simulation study design in Chapter 4, three observation-setup options were provided to adopted the two CSD-EEG datasets:

Option 1) Use only the CSD-EEG data measured under eyes-open state $\boldsymbol{x}_{i1}$ as the matrix-valued covariates, assuming no measurement errors involved;

Option 2) Use only the CSD-EEG data measured under eyes-closed state $\boldsymbol{x}_{i2}$ as the matrix-valued covariates, assuming no measurement errors involved;

Option 3) Use the mean of the CSD-EEG data measured under eyes-open state and eyes-closed state, i.e., $\frac{1}{2}\sum_{m=1}^{2}\boldsymbol{x}_{im}$, as an estimation of the latent CSD-EEG for the two methods.

For $(p_0, q_0) = (2, 2)$ with each observation-setup option, the two baseline methods are implemented as follows. For the naiveJM method, the prior distributions for all model parameters are chosen to be the same as for our approach (as described in chapter 3.4). Moreover, to make the results equitable, the MCMC chains for their method were also obtained with the same number of iterations as ours. That is, the posterior samples of model parameters were accessed by keeping every 5^{th} draw from 20,000 iterations after 20,000 burn-in iterations. The posterior mean of coefficients is then obtained from the formulation described in remark 3.2.

For the naiveTSM method, we first applied the MPCA procedure on CSD-EEG data to extract low-dimensional features of EEG with rTensor package (Li, Bien, & Wells, 2018). Then, a probit regression model was implemented to regress the response o_i with scalar covariates and vectorized low-dimensional features. The coefficients for observed EEG were calculated as described in remark 3.2.

5.3 Results

In this section, we demonstrated the coefficients for CSD-EEG data estimated from three methods under the optimal choice of $(p_0, q_0) = (2, 2)$. Figures 5.1 to 5.5 show the estimated coefficients and the significance results for CSD-EEG data. In these figures, the filling color of each cell indicates the value of estimated coefficients, and the discoveries of significant effects are marked as yellow asterisks. Due to the advantage of the Bayesian framework, the significant effects at 5% significance level are determined if 0 is not contained in the 95% credible intervals of the coefficients, and is denoted as yellow asterisks in the figures.

5.3.1 JMME

For the JMME method, our results show that neither chronicity ($\hat{\gamma}_1 = -0.49$ with 95% CI: (-3.53, 2.91)) nor gender ($\hat{\gamma}_2 = 0.54$ with 95% CI: (-3.09, 3.87)) contribute to the prediction on SSRI responder. For the latent 14×13 CSD-EEG covariates, among all 14 electrodes considered in our study, P8, P7, P5, O2, and O1 electrodes do not have important effects over the theta frequency band on predicting the treatment outcome. The remaining 9 electrodes jointly have significant effects on the 4-5Hz and 6-7Hz frequency range. In total, our JMME method successfully identified 90 combinations of electrodes and frequencies that jointly played significant roles in predicting antidepressant response. Our findings corroborate previous literature that CSD-EEG data recorded at the posterior brain regions and theta frequency band contain useful information in antidepressant response prediction (Tenke et al., 2017).

5.3.2 naiveJM

As for the naiveJM method, figures 5.2 to 5.4 demonstrate the results under the three observation-setup options With $(p_0, q_0) = (2, 2)$. It is shown that the naiveJM gave similar estimation and inference results of coefficients $\boldsymbol{\beta}$ between Option 1 and Option 2. Figure 5.2 and Figure 5.3 suggest that no electrodes or frequencies are considered as significant for predicting the treatment response. This result is improved using the mean of the EEG data under two conditions (Option 3) where 3 electrodes (P7, P5, O2) at 5.75Hz were detected as significant, as shown in Figure 5.4.

When comparing the values of coefficient estimates, Figure 5.2 and Figure 5.3 shows that the estimates for coefficients when using EEG with Option 1 and 2 are very close to zero, which is an indication of attenuation bias in the estimation. In contrast, the range of estimates in Option 3 (Figure 5.4) for coefficients are slightly increased, though they are still approaching zero. As

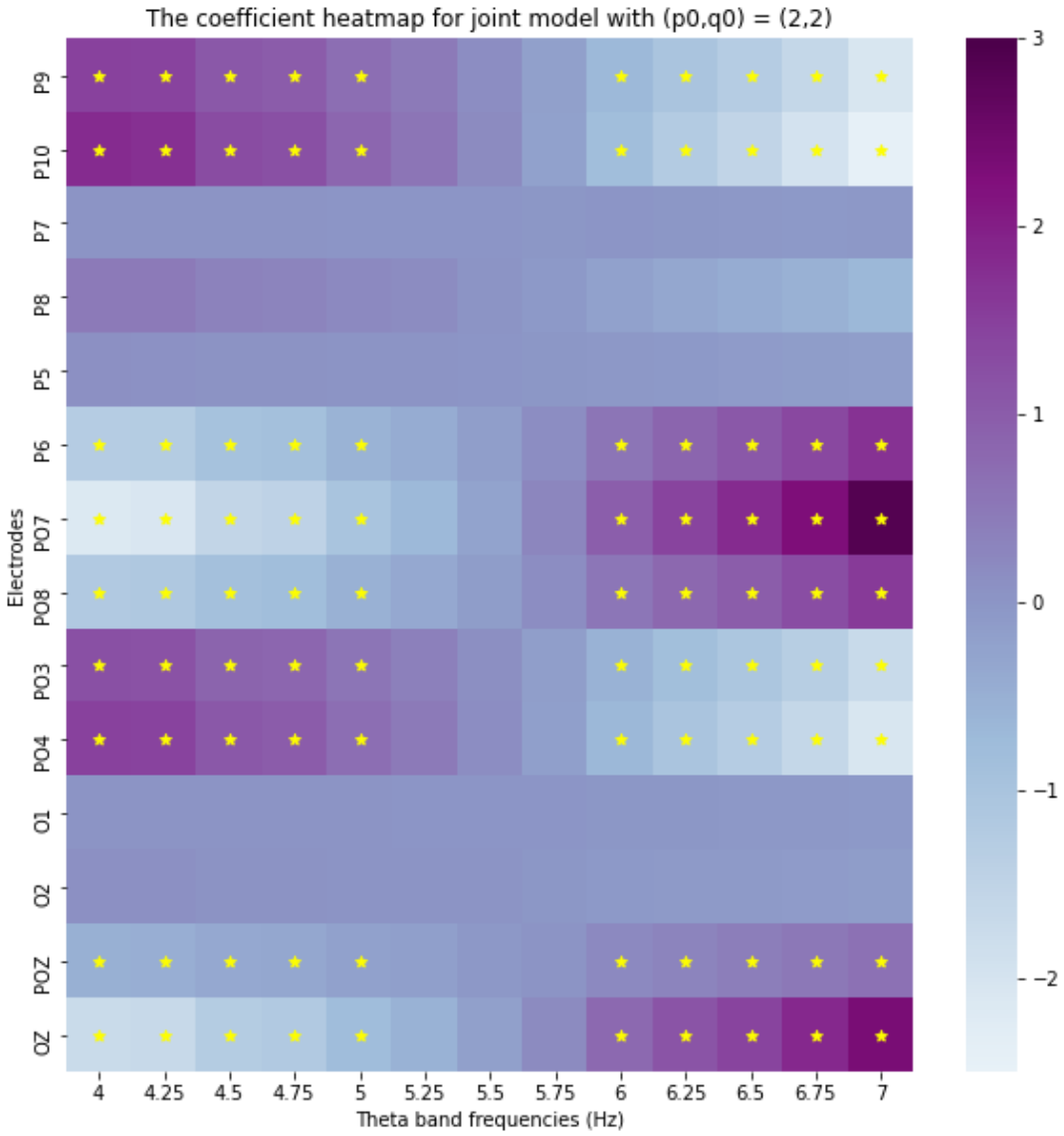

Figure 5.1: Estimation and significance results for the JMME framework, when $(p_0, q_0) = (2, 2)$

pointed out in Chapter 4, such phenomenon is reasonable due to the reduction in variability in EEG by taking the average, and hence the measurement errors are potentially reduced, which further leads to improved estimation of associations. However, compared to the results from our proposed JMME method, simply taking the average is insufficient to adjust for measurement errors, and the effects of other electrodes and frequencies are still underestimated. Moreover, the inconsistency of identified significant effects between the naiveJM method and our proposed JMME method also follows the literature that, when the measurement errors exist in predictors, ignoring them in modeling may lead to inconsistent inference results.

5.3.3 naiveTSM

As shown in Figures 5.5 to 5.7, the results of naiveTSM are very similar to naiveJM's. The estimates of coefficients for EEG are close to zero when regressing with EEG with Option 1 and 2, leading to no important effects determined for predicting patients' responses. The negative impact brought by the attenuation bias was slightly alleviated when the naiveTSM method was applied with Option 3. As a result, the estimates following this option have a slightly larger range for estimated coefficients, and 3 significant effects are determined with electrodes P7, P5, and O2 with 5.75Hz. However, compared with the results from our JMME method, naiveTSM estimated the coefficients still close to zero and discovered fewer significant effects, confirming the finding that modeling without measurement error correction could lead to severe attenuation bias.

5.3.4 Overall

By comparing the results from the three methods, the influence of measurement errors in EEG data for antidepressant response prediction is compelling.

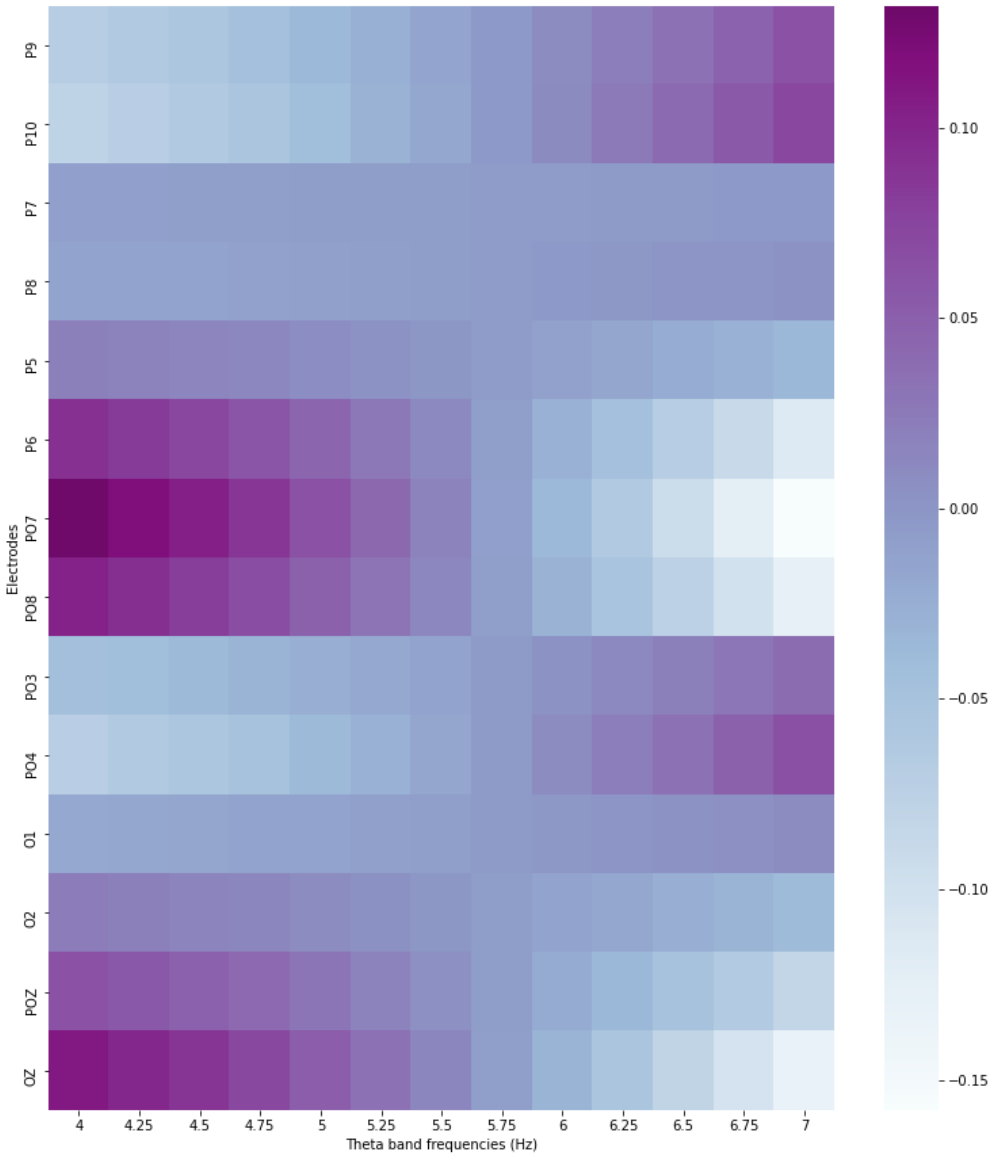

Figure 5.2: Estimation and significance results for naiveJM with option 1, when $(p_0, q_0) = (2, 2)$

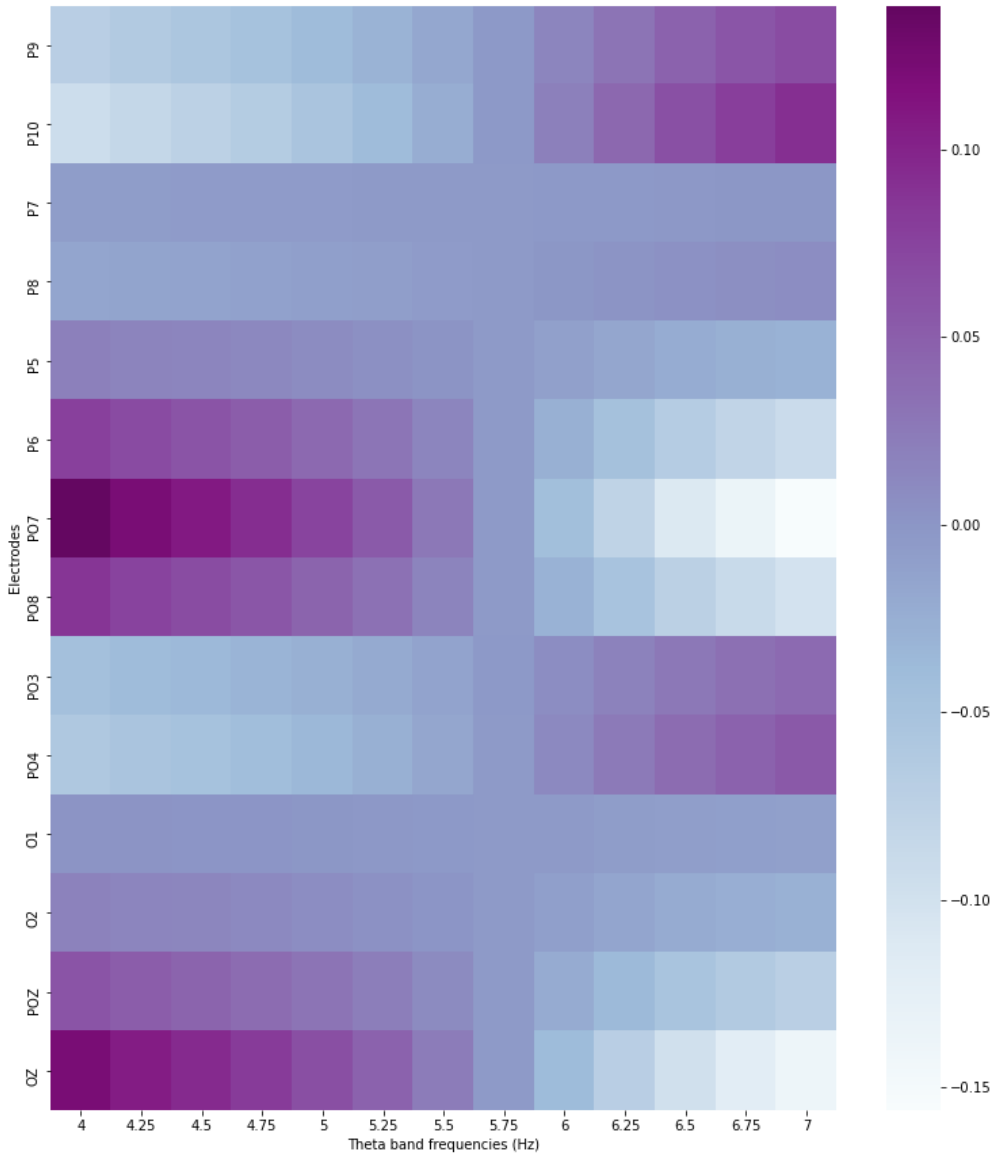

Figure 5.3: Estimation and significance results for naiveJM with option 2, when $(p_0, q_0) = (2, 2)$

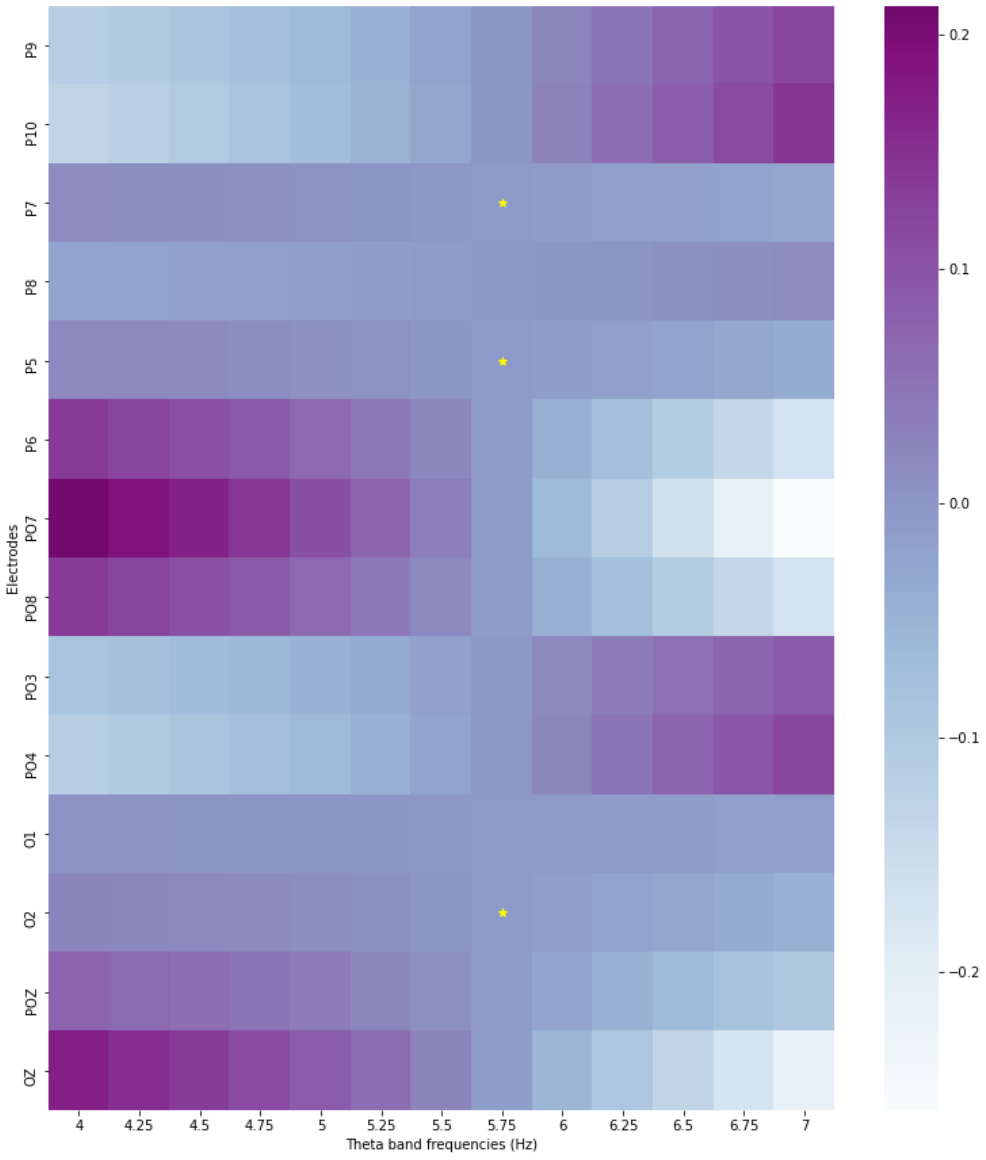

Figure 5.4: Estimation and significance results for naiveJM with option 3, when $(p_0, q_0) = (2, 2)$

The two methods without measurement error corrections estimated the coefficients for CSD-EEG very close to zero and can hardly infer the significant association between EEG data and patients' responses to SSRI antidepressants. However, our proposed method is able to successfully correct for potential measurement errors and recover meaningful EEG data. The corresponding results from our framework also corroborate the literature that baseline CSD-EEG at theta frequency band can provide feasible information on antidepressant response prediction.

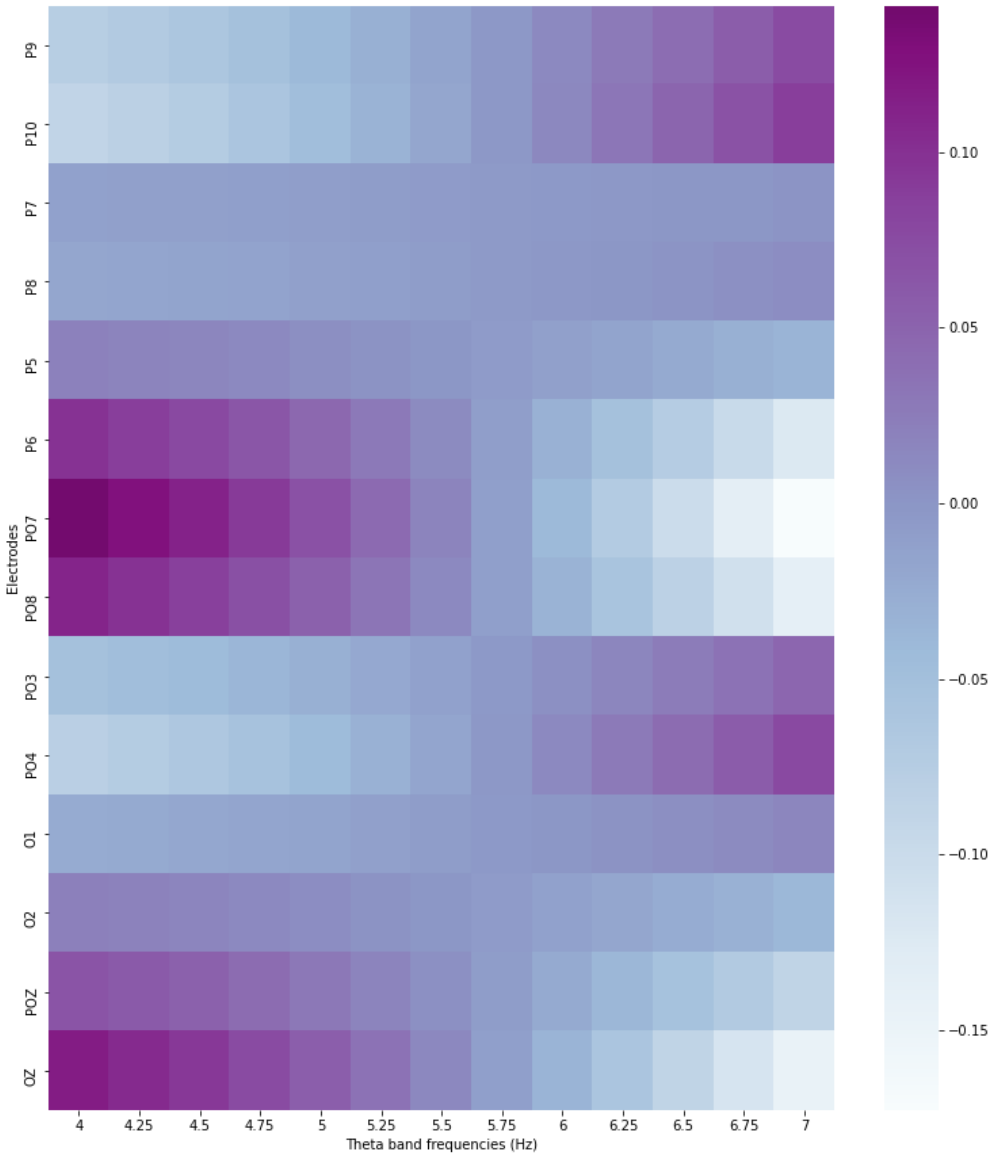

Figure 5.5: Estimation and Significance results for naiveTSM with option 1, when $(p_0, q_0) = (2, 2)$

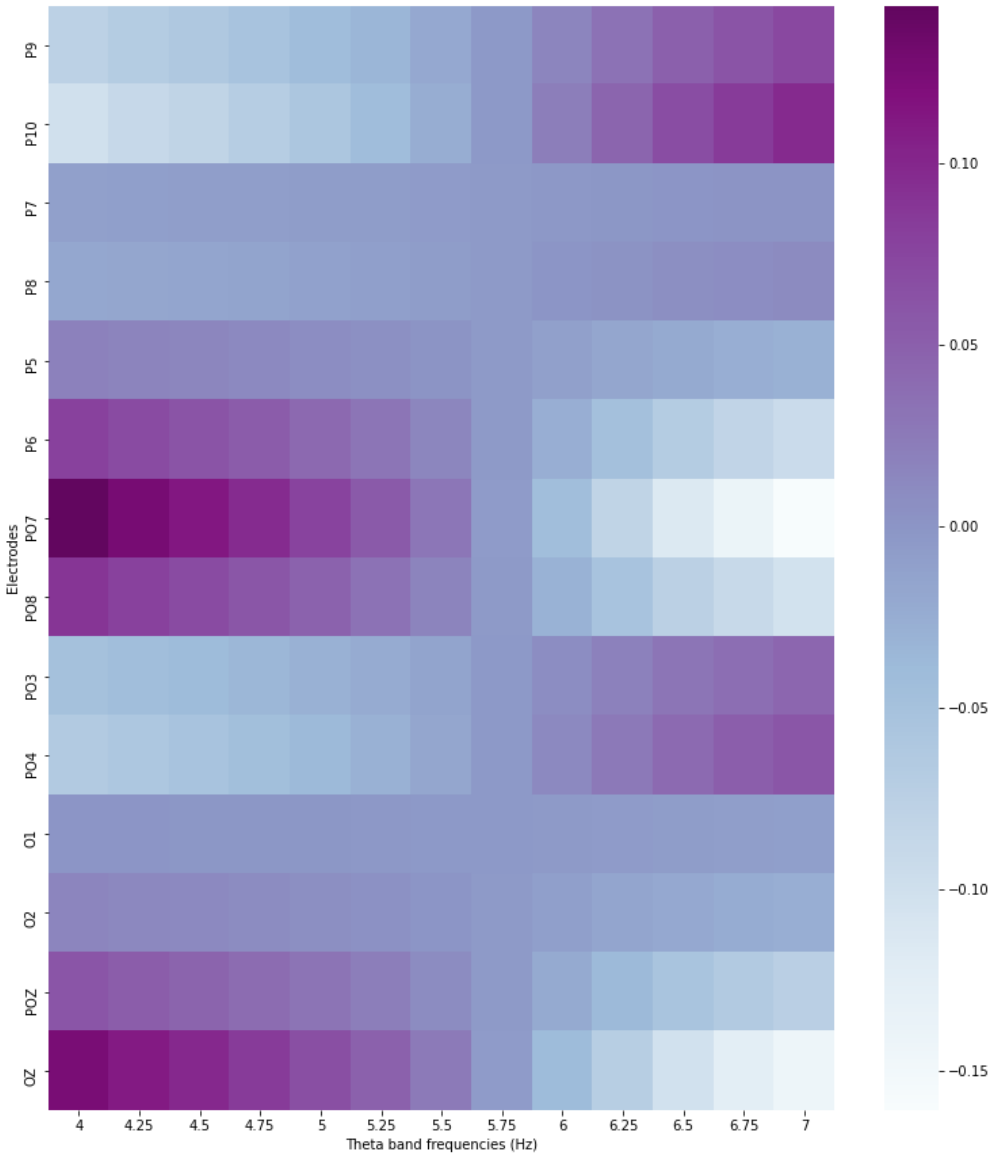

Figure 5.6: Estimation and Significance results for naiveTSM with option 2, when $(p_0, q_0) = (2, 2)$

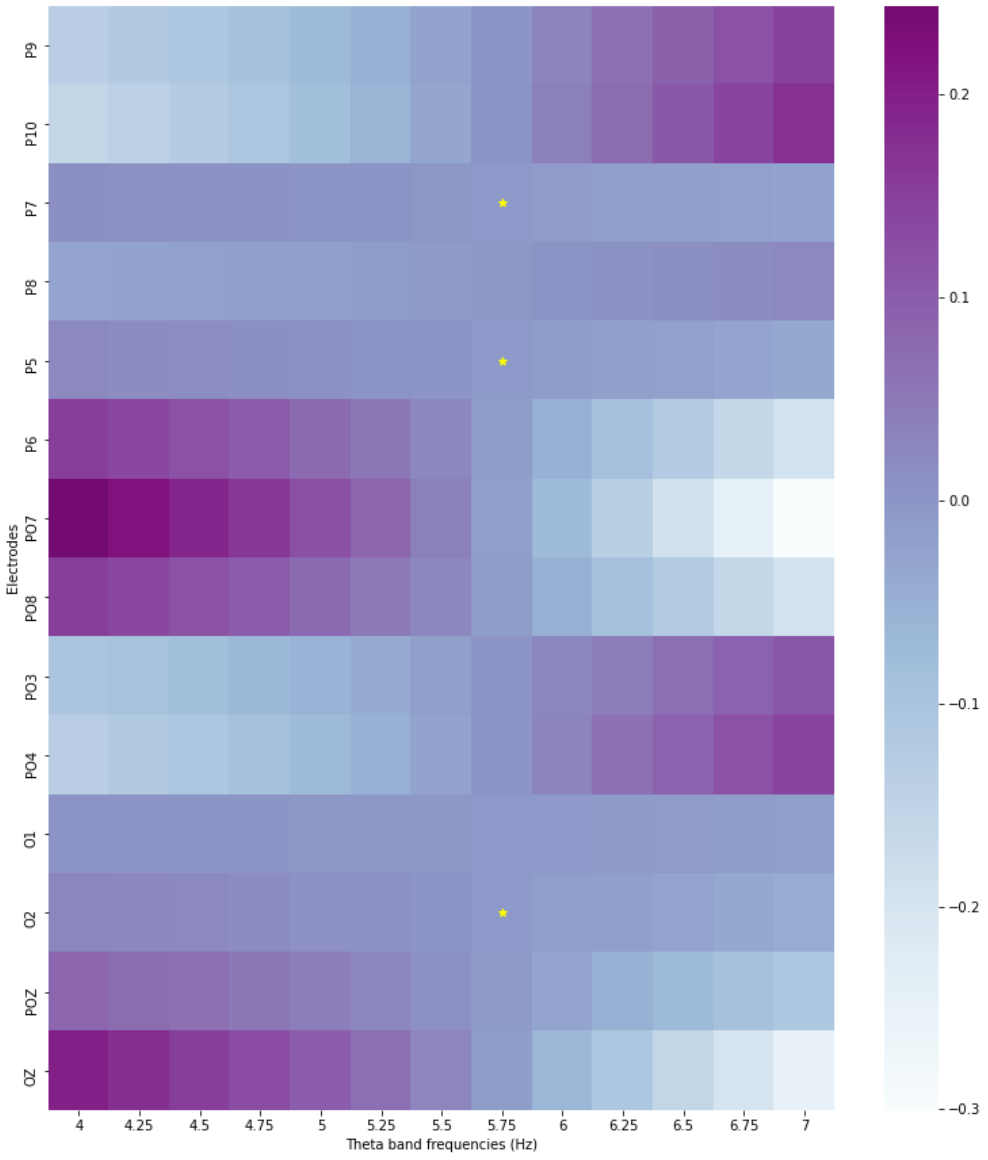

Figure 5.7: Estimation and Significance results for naiveTSM with option 3, when $(p_0, q_0) = (2, 2)$

Chapter 6

Discussion and Conclusion

In this thesis, in order to adjust for the measurement errors in the repeatedly observed matrix-valued surrogates and examine the underneath association between the latent matrix-valued covariates and the binary outcome, we have proposed a Bayesian joint model framework, JMME. Specifically, JMME consists of defining (**a**) a probabilistic measurement error sub-model $\mathcal{M}_{ME}$ assuming additive measurement errors in the repeatedly observed matrix covariates $\{\boldsymbol{x}_{im}\}$ and imposing a low-rank structure on the covariance matrix of measurement errors; (**b**) a probabilistic MPCA submodel $\mathcal{M}_{MPCA}$ extracting low dimensional features ${}_i \in \mathcal{R}^{p_0 \times q_0}$ to approximate the true latent matrix covariates $\boldsymbol{x}_i^* \in \mathcal{R}^{p \times q}$; and (**c**) a outcome regression model $\mathcal{M}_{out}$ relating the low dimensional features with binary outcome o_i. Our joint model framework is proved to be identifiable such that the coefficients of latent matrix covariates in the original $p \times q$ space can be uniquely determined. As demonstrated in the simulation study, our proposed joint model achieved the least RMSE under all described scenarios, compared to the naiveJM and naiveTSM methods. Especially when the sample sizes are less than the dimensionality of the matrix covariates, our proposed model outperforms the other two methods. The simulation results show that the joint model we proposed is able to efficiently correct for additive measurement errors and regulate the matrix-valued covari-

ates, leading to promising estimation of the relationship between outcomes and covariates. Moreover, by the advantage of the Bayesian method, the MCMC chains obtained from Gibbs sampling allow the quantification of the uncertainties among estimates of model parameters, so that the inferences for model parameters can be easily made.

When applying the proposed JMME framework to the motivating EEG dataset, our proposed joint modeling approach successfully estimated the most significant effects in the posterior brain areas at theta band frequencies. Without correcting for measurement errors, the effects of EEG tend to be underestimated from both the naive joint model and the naive two-stage model. Indicating from the proposed model, for the true unobserved EEG data, the 9 electrodes (P10, P7, P6, PO7, PO8, PO3, PO4, POZ, OZ) at 4 - 5Hz and 6 -7 Hz frequencies jointly are associated with patients' response to antidepressants. Our finding matches the findings from recent studies that EEG measures recorded at posterior brain regions and theta band frequencies are persistently advised as correlated with antidepressant outcomes.

To the best of our knowledge, JMME framework is the first method to simultaneously correct for the measurement errors for repeated matrix-valued covariates through low-rank covariance structure, utilize a probabilistic MPCA model, and predict the binary outcome of interest. The advantage is that inferences of model parameters can be naturally established through Bayesian credible intervals. However, we have to admit that our method is still subject to limitations: the current formulation assumed independent additive measurement errors with specific low-rank covariance structure, which may not always be true in the reality. It is less flexible than the two-stage model, in the sense that the naiveTSM method could utilize other regression or classification models and correct for measurement errors at stage 2 to further improve the performance.

Our work can be extended in many aspects. Firstly, the assumption on measurement errors in observed matrix-valued surrogates can be extended to more relaxed structures. For example, consider the additive matrix-valued measurement errors with correlation among rows and columns, or consider the measurement errors correlated with scalar covariates. Furthermore, we could think of imposing low-rank structures directly for the unobserved matrix covariates, so that shrinkage on the rows and columns in the original high dimensional space can be achieved.

References

Alhaj, H., Wisniewski, G., & McAllister-Williams, R. H. (2011). The use of the eeg in measuring therapeutic drug action: Focus on depression and antidepressants. *Journal of Psychopharmacology*, *25*(9), 1175–1191.

Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports medicine*, *26*(4), 217–238.

Carroll, R. J., & Stefanski, L. A. (1990). Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association*, *85*(411), 652–663.

Carroll, R. J., & Wand, M. P. (1991). Semiparametric estimation in logistic measurement error models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *53*(3), 573–585.

Chen, X., Hong, H., & Nekipelov, D. (2007). Measurement error models. *Prepared for the Journal of Economic Literature. www. stanford. edu/˜ doubleh/eco273B/surveyjan27chenhandenis-07. pdf.*

Chib, S., & Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The american statistician*, *49*(4), 327–335.

Datta, A., Zou, H. Et al. (2017). Cocolasso for high-dimensional error-in-variables regression. *The Annals of Statistics*, *45*(6), 2400–2426.

Ding, S., & Cook, R. D. (2016). Matrix-variate regressions and envelope models. *arXiv preprint arXiv:1605.01485*.

Fang, J., & Yi, G. Y. (2020). Matrix-variate logistic regression with measurement error. *Biometrika*.

Fuller, W. A. (2009). *Measurement error models* (Vol. 305). John Wiley & Sons.

Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor economics*, *24*(3), 411–482.

Hellton, K. H., & Thoresen, M. (2014). The impact of measurement error on principal component analysis. *Scandinavian Journal of Statistics*, *41*(4), 1051–1063.

Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, *102*(478), 674–685.

Holsboer, F. (2008). How can we realize the promise of personalized antidepressant medicines? *Nature Reviews Neuroscience*, *9*(8), 638–646.

Hung, H., & Wang, C.-C. (2013). Matrix variate logistic regression model with application to eeg data. *Biostatistics*, *14*(1), 189–202.

Hung, H., Wu, P., Tu, I., & Huang, S. (2012). On multilinear principal component analysis of order-two tensors. *Biometrika*, *99*(3), 569–583.

Iosifescu, D. V., Greenwald, S., Devlin, P., Mischoulon, D., Denninger, J. W., Alpert, J. E., & Fava, M. (2009). Frontal eeg predictors of treatment outcome in major depressive disorder. *European Neuropsychopharmacology*, *19*(11), 772–777.

Jiang, B., Petkova, E., Tarpey, T., & Ogden, R. T. (2017). Latent class modeling using matrix covariates with application to identifying early placebo responders based on eeg signals. *The annals of applied statistics*, *11*(3), 1513.

Jiang, B., Petkova, E., Tarpey, T., & Ogden, R. T. (2020). A bayesian approach to joint modeling of matrix-valued imaging data and treatment outcome with applications to depression studies. *Biometrics*, *76*(1), 87–97.

Khosla, D., Don, M., & Kwong, B. (1999). Spatial mislocalization of eeg electrodes–effects on accuracy of dipole estimation. *Clinical neurophysiology*, *110*(2), 261–271.

Lai, Z., Xu, Y., Chen, Q., Yang, J., & Zhang, D. (2014). Multilinear sparse principal component analysis. *IEEE Transactions on Neural Networks and Learning Systems*, *25*(10), 1942–1950. https://doi.org/10.1109/TNNLS.2013.2297381.

Lee, T.-W., Wu, Y.-T., Yu, Y. W.-Y., Chen, M.-C., & Chen, T.-J. (2011). The implication of functional connectivity strength in predicting treatment response of major depressive disorder: A resting eeg study. *Psychiatry Research: Neuroimaging*, *194*(3), 372–377.

Li, J., Bien, J., & Wells, M. T. (2018). Rtensor: An r package for multidimensional array (tensor) unfolding, multiplication, and decomposition. *Journal of Statistical Software*, *87*(1), 1–31.

Liu, A. K., Dale, A. M., & Belliveau, J. W. (2002). Monte carlo simulation studies of eeg and meg localization accuracy. *Human brain mapping*, *16*(1), 47–62.

Lu, H., Plataniotis, K. N., & Venetsanopoulos, A. N. (2008). Mpca: Multilinear principal component analysis of tensor objects. *IEEE transactions on Neural Networks*, *19*(1), 18–39.

Olbrich, S., & Arns, M. (2013). Eeg biomarkers in major depressive disorder: Discriminative power and prediction of treatment response. *International Review of Psychiatry*, *25*(5), 604–618.

Panagakis, Y., Kotropoulos, C., & Arce, G. R. (2009). Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification. *IEEE Transactions on Audio, Speech, and Language Processing*, *18*(3), 576–588.

Richardson, S., & Gilks, W. R. (1993). Conditional independence models for epidemiological studies with covariate measurement error. *Statistics in Medicine*, *12*(18), 1703–1722.

Sanguinetti, G., Milo, M., Rattray, M., & Lawrence, N. D. (2005). Accounting for probe-level noise in principal component analysis of microarray data. *Bioinformatics*, *21*(19), 3748–3754.

Schofield, L. S. (2015). Correcting for measurement error in latent variables used as predictors. *The annals of applied statistics*, *9*(4), 2133.

Steiger, A., & Kimura, M. (2010). Wake and sleep eeg provide biomarkers in depression. *Journal of psychiatric research*, *44*(4), 242–252.

Tenke, C. E., & Kayser, J. (2012). Generator localization by current source density (csd): Implications of volume conduction and field closure at intracranial and scalp resolutions. *Clinical neurophysiology*, *123*(12), 2328–2345.

Tenke, C. E., Kayser, J., Pechtel, P., Webb, C. A., Dillon, D. G., Goer, F., Murray, L., Deldin, P., Kurian, B. T., McGrath, P. J., Et al. (2017). Demonstrating test-retest reliability of electrophysiological measures for healthy adults in a multisite study of biomarkers of antidepressant treatment response. *Psychophysiology*, *54*(1), 34–50.

van Driel, J., Ridderinkhof, K. R., & Cohen, M. X. (2012). Not all errors are alike: Theta and alpha eeg dynamics relate to differences in error-processing dynamics. *Journal of Neuroscience*, *32*(47), 16795–16806.

Wade, E. C., & Iosifescu, D. V. (2016). Using electroencephalography for treatment guidance in major depressive disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *1*(5), 411–422.

Wentzell, P. D., Andrews, D. T., Hamilton, D. C., Faber, K., & Kowalski, B. R. (1997). Maximum likelihood principal component analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *11*(4), 339–366.

Yang, J., Zhang, D., Frangi, A. F., & Yang, J.-y. (2004). Two-dimensional pca: A new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence*, *26*(1), 131–137.

Zhang, D., & Zhou, Z.-H. (2005). (2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition. *Neurocomputing*, *69*(1-3), 224–231.

Zhou, H., & Li, L. (2014). Regularized matrix regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(2), 463–483.

Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, *108*(502), 540–552.

Appendix A

Derivation for model parameters' full conditional posterior distributions

This appendix provides the computation of model parameters' full conditional posterior distributions.

1. **full conditional posterior distribution of $\widetilde{\boldsymbol{\Sigma}}_i$**

   With $\widetilde{\boldsymbol{\Sigma}}_i = \begin{pmatrix} \sigma^2_{i11} & \cdots & \sigma^2_{i1q} \\ \vdots & \ddots & \vdots \\ \sigma^2_{ip1} & \cdots & \sigma^2_{ipq} \end{pmatrix}$ in $\mathcal{M}_{ME}$ (model 3.1), let σ^2_{ijk} denotes the element in the j^{th} row and k^{th} column of $\widetilde{\boldsymbol{\Sigma}}_i$. Then, the low rank formulation (model 3.2) results in the following structure for $\log(\sigma^2_{ijk})$:

   $$\log(\sigma^2_{ijk}) = \widetilde{u}_j \widetilde{v}_k + \widetilde{\epsilon}_{ijk},$$

   where $\widetilde{\epsilon}_{ijk} \sim \mathcal{N}(0, \delta^2)$. Moreover, model (3.1) also indicates that

   $$\boldsymbol{x}_{imjk} = \boldsymbol{x}^*_{imjk} + \epsilon^*_{imjk},$$

   where $\epsilon^*_{imjk} \sim \mathcal{N}(0, \sigma^2_{ijk})$. In order to make the notation consistent, we adopted the identity transformation of σ^2_{ijk}: $\epsilon^*_{imjk} \sim \mathcal{N}(0, \exp(\log(\sigma^2_{ijk})))$. Therefore, the full conditional distribution of $\log(\sigma^2_{ijk})$ given all other

model parameters is derived as follows:

$$
\begin{aligned}
& p(\log(\sigma_{ijk}^2)|\cdot) \\
& \propto \prod_{m=1}^{M} p(\boldsymbol{x}_{imjk}|\log(\sigma_{ijk}^2, \boldsymbol{x}_{imjk}^*)) \times p(\log(\sigma_{ijk}^2)|\widetilde{u}_j, \widetilde{v}_k, \delta^2) \\
& \propto \prod_{m=1}^{M} \left[\left(\exp(\log(\sigma_{ijk}^2))\right)^{-1/2} \exp\left\{ -\frac{1}{2} \frac{(\boldsymbol{x}_{imjk} - \boldsymbol{x}_{ijk}^*)^2}{\exp(\log(\sigma_{ijk}^2))} \right\} \right] \times \exp\left\{ -\frac{1}{2} \frac{(\log(\sigma_{ijk}^2) - \widetilde{u}_j \widetilde{v}_k)^2}{\delta^2} \right\}.
\end{aligned}
\tag{A.1}
$$

It is noticed that this kernel density of posterior distribution is not proposed to any named distributions, so that the direct sampling is difficult in this case. In order to draw the samples from an arbitrary probability density, we adopted the Metropolis-Hastings algorithm (Chib & Greenberg, 1995) to draw the posterior samples from a kernel function proportional to the probability density function. To apply Metropolis-Hastings algorithm, we let $f(\log(\sigma_{ijk}^2)$ denote this kernel of posterior density, assume a normal proposal distribution (i.e., the distribution that generates next candidate value), and set the rejection probability as uniformly distributed on $(0, 1)$. Specifically, the Metropolis-Hastings sampling procedure is implemented as follows:

(a) Randomly generate $\log(\sigma_{ijk}^2)^0$ as the initial and calculate the value of $f(\log(\sigma_{ijk}^2)^0|\cdot)$;

(b) At iteration $t+1$, Proposed the candidate value $\log(\sigma_{ijk}^2)^{cand} \sim \mathcal{N}(0, 10)$;

   i. if $f(\log(\sigma_{ijk}^2)^{cand}|\cdot) > f(\log(\sigma_{ijk}^2)^t|\cdot)$, set $\log(\sigma_{ijk}^2)^{t+1} = \log(\sigma_{ijk}^2)^{cand}$;

   ii. if not, generate $u \sim \text{uniform}(0, 1)$:

      A. if $u < \frac{f(\log(\sigma_{ijk}^2)^{cand}|\cdot)}{f(\log(\sigma_{ijk}^2)^t}$, set $\log(\sigma_{ijk}^2)^{t+1} = \log(\sigma_{ijk}^2)^{cand}$;

      B. if not, set $\log(\sigma_{ijk}^2)^{t+1} = \log(\sigma_{ijk}^2)^t$.

2. **full conditional posterior distribution of $\widetilde{\boldsymbol{U}}, \mu_u, \sigma_u^2$**

Recall that in $\mathcal{M}_{low}$ (model 3.2), the rank of the measurement error covariance matrix $\widetilde{\boldsymbol{\Sigma}}_i$ is determined as $r < \min(p, q)$. Here, for the ease of derivation, we developed the posterior distribution with $r = 1$. The

derivation can be easily generalized to $r > 1$.

(a) With $\widetilde{\boldsymbol{U}} = \begin{pmatrix} \widetilde{u}_1 \\ \vdots \\ \widetilde{u}_p \end{pmatrix}$ in $\mathcal{M}_{low}$ (model 3.2), let $\widetilde{u}_j$ stand for the j^{th} element of $\widetilde{\boldsymbol{U}}$. The full conditional distribution of $\widetilde{u}_j$ is of the form

$$
\begin{aligned}
& p(\widetilde{u}_j|\cdot) \\
& \propto \prod_{i=1}^{n} p(\log(\boldsymbol{\Sigma}_{i[j,]})|\widetilde{u}_j, \widetilde{V}, \delta^2) \times p(\widetilde{u}_j|\mu_u, \sigma_u^2) \\
& \propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\delta^2}\left[(\log\boldsymbol{\Sigma}_{i[j,]})^\top - \widetilde{u}_j\widetilde{\boldsymbol{V}}\right]^\top\left[(\log\boldsymbol{\Sigma}_{i[j,]})^\top - \widetilde{u}_j\widetilde{\boldsymbol{V}}\right]\right\} \times \exp\left\{-\frac{1}{2}\frac{(\widetilde{u}_j - \mu_u)^2}{\sigma_u^2}\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\widetilde{u}_j - m\right]^\top C^{-1}\left[\widetilde{u}_j - m\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.2}
$$

leading to a normal kernel with mean M and variance E, where

$$
\begin{aligned}
E &= \left(\frac{1}{\sigma_u^2} + \frac{\widetilde{\boldsymbol{V}}^\top\widetilde{\boldsymbol{V}}}{\delta^2}\right)^{-1}, \\
M &= E\left(\frac{\mu_u}{\sigma_u^2} + \sum_{i=1}^{n}\frac{\widetilde{\boldsymbol{V}}^\top\log(\boldsymbol{\Sigma}_{i[j,]})^\top}{\delta^2}\right).
\end{aligned}
$$

Therefore, $[\widetilde{u}_j|\cdot] \sim \mathcal{N}(M, E)$.

(b) With $\widetilde{\boldsymbol{U}} = \begin{pmatrix} \widetilde{u}_1 \\ \vdots \\ \widetilde{u}_p \end{pmatrix}$, the full conditional distribution of μ_u is

$$
\begin{aligned}
& p(\mu_u|\cdot) \\
& \propto \prod_{j=1}^{p} p(\widetilde{u}_j|\mu_u, \sigma_u^2) \times p(\mu_u|\sigma_0^2) \\
& \propto \prod_{j=1}^{p} \exp\left\{-\frac{1}{2}\frac{(\widetilde{u}_j - \mu_u)^2}{\sigma_u^2}\right\} \times \exp\left(-\frac{\mu_u^2}{2\sigma_0^2}\right) \\
& \propto \exp\left\{-\frac{1}{2}\left[\mu_u - m\right]^\top C^{-1}\left[\mu_u - m\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.3}
$$

corresponding to a normal kernel with mean M and variance E, where

$$E = (\frac{1}{\sigma_0^2} + \frac{p}{\sigma_u^2})^{-1},$$
$$M = E\left(\sum_{j=1}^{p} \frac{\widetilde{u}_j}{\sigma_u^2}\right).$$

Therefore, $[\mu_u|\cdot] \sim \mathcal{N}(M, E)$.

(c) The full conditional distribution of σ_u^2 is

$$\begin{aligned} & p(\sigma_u^2|\cdot) \\ & \propto \prod_{j=1}^{p} p(\widetilde{u}_j|\mu_u, \sigma_u^2) \times p(\sigma_u^2|a_0, b_0) \\ & \propto \prod_{j=1}^{p} \exp\left\{-\frac{1}{2}\frac{(\widetilde{u}_j - \mu_u)^2}{\sigma_u^2}\right\} \times (\sigma_u^2)^{(-a_0-1)} \exp\left(-\frac{b_0}{\sigma_u^2}\right) \\ & \propto (\sigma_u^2)^{(-a_1-1)} \exp\left(-\frac{b_1}{\sigma_u^2}\right), \end{aligned} \tag{A.4}$$

leading to an inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$a_1 = a_0 + p/2,$$
$$b_1 = b_0 + \frac{1}{2}\sum_{j=1}^{p}(\widetilde{u}_j - \mu_u)^2.$$

Thus, $[(\sigma_u^2)^{-1}|\cdot] \sim \text{Gamma}(a_1, b_1)$.

3. **full conditional posterior distribution of $\widetilde{\boldsymbol{V}}, \mu_v, \sigma_v^2$**

Similar as the derivation for $\widetilde{\boldsymbol{U}}$, here we derive the posterior distribution for $\widetilde{\boldsymbol{V}} \in \mathcal{R}^{q\times r}$ with $r = 1$.

(a) With $\widetilde{\boldsymbol{V}} = \begin{pmatrix} \widetilde{v}_1 \\ \vdots \\ \widetilde{v}_q \end{pmatrix}$, let $\widetilde{v}_k$ stand for the k^{th} element of $\widetilde{\boldsymbol{V}}$. The full

conditional distribution of $\widetilde{v}_k$ is of the form

$$
\begin{aligned}
&p(\widetilde{v}_k|\cdot) \\
&\propto \prod_{i=1}^{n} p(\log(\boldsymbol{\Sigma}_{i[,k]})|\widetilde{U}, \widetilde{v}_k, \delta^2) \times p(\widetilde{v}_k|\mu_v, \sigma_v^2) \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\delta^2}\left[(\log\boldsymbol{\Sigma}_{i[,k]}) - \widetilde{v}_k\widetilde{\boldsymbol{U}}\right]^\top \left[(\log\boldsymbol{\Sigma}_{i[,k]}) - \widetilde{v}_k\widetilde{\boldsymbol{U}}\right]\right\} \times \exp\left\{-\frac{1}{2}\frac{(\widetilde{v}_k - \mu_v)^2}{\sigma_v^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\widetilde{v}_k - m\right]^\top C^{-1}\left[\widetilde{v}_k - m\right]\right\} \text{ by completing the square.}
\end{aligned}
\tag{A.5}
$$

This is a normal kernel with mean M and variance E, where

$$
\begin{aligned}
E &= \left(\frac{1}{\sigma_v^2} + \frac{\widetilde{\boldsymbol{U}}^\top\widetilde{\boldsymbol{U}}}{\delta^2}\right)^{-1}, \\
M &= E\left(\frac{\mu_v}{\sigma_v^2} + \sum_{i=1}^{n}\frac{\widetilde{\boldsymbol{U}}\top\log(\boldsymbol{\Sigma}_{i[,k]})}{\delta^2}\right).
\end{aligned}
$$

Therefore, $[\widetilde{v}_k|\cdot] \sim \mathcal{N}(M, E)$.

(b) The full conditional distribution of μ_v is

$$
\begin{aligned}
&p(\mu_v|\cdot) \\
&\propto \prod_{k=1}^{q} p(\widetilde{v}_k|\mu_v, \sigma_v^2) \times p(\mu_v|\sigma_0^2) \\
&\propto \prod_{k=1}^{q} \exp\left\{-\frac{1}{2}\frac{(\widetilde{v}_k - \mu_v)^2}{\sigma_v^2}\right\} \times \exp\left(-\frac{\mu_v^2}{2\sigma_0^2}\right) \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu_v - m\right]^\top C^{-1}\left[\mu_v - m\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.6}
$$

leading to a normal kernel with mean M and variance E, where

$$
\begin{aligned}
E &= (\frac{1}{\sigma_0^2} + \frac{q}{\sigma_v^2})^{-1}, \\
M &= E\left(\sum_{k=1}^{q}\frac{\widetilde{v}_k}{\sigma_v^2}\right).
\end{aligned}
$$

Therefore, $[\mu_v|\cdot] \sim \mathcal{N}(M, E)$.

(c) The full conditional distribution of σ_v^2 is

$$\begin{aligned} & p(\sigma_v^2|\cdot) \\ & \propto \prod_{k=1}^{q} p(\widetilde{v}_k|\mu_v, \sigma_v^2) \times p(\sigma_v^2|a_0, b_0) \\ & \propto \prod_{k=1}^{q} \exp\left\{-\frac{1}{2}\frac{(\widetilde{v}_k - \mu_v)^2}{\sigma_v^2}\right\} \times (\sigma_v^2)^{(-a_0-1)} \exp\left(-\frac{b_0}{\sigma_v^2}\right) \\ & \propto (\sigma_v^2)^{(-a_1-1)} \exp\left(-\frac{b_1}{\sigma_v^2}\right), \end{aligned} \tag{A.7}$$

corresponding to an inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$\begin{aligned} a_1 &= a_0 + q/2, \\ b_1 &= b_0 + \frac{1}{2}\sum_{k=1}^{q} (\widetilde{v}_k - \mu_v)^2. \end{aligned}$$

Thus, $[(\sigma_v^2)^{-1}|\cdot] \sim \text{Gamma}(a_1, b_1)$.

4. **full conditional posterior distribution of δ^2**

The full conditional distribution of δ^2 is

$$\begin{aligned} & p(\delta^2|\cdot) \\ & \propto \prod_{i=1}^{n} p(\log\boldsymbol{\Sigma}_i|\widetilde{\boldsymbol{U}}, \widetilde{\boldsymbol{V}}, \delta^2) \times p(\delta^2|a_0, b_0) \\ & \propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2\delta^2}\|\log\boldsymbol{\Sigma}_i - \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^\top\|_F^2\right\} \times (\delta^2)^{(-a_0-1)} \exp\left(-\frac{b_0}{\delta^2}\right) \\ & \propto (\delta^2)^{(-a_1-1)} \exp\left(-\frac{b_1}{\sigma_v^2}\right), \end{aligned} \tag{A.8}$$

corresponding to an inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$\begin{aligned} a_1 &= a_0 + npq/2, \\ b_1 &= b_0 + \frac{1}{2}\sum_{i=1}^{n} \|\log\boldsymbol{\Sigma}_i - \widetilde{\boldsymbol{U}}\widetilde{\boldsymbol{V}}^\top\|_F^2. \end{aligned}$$

Thus, $[(\delta)^{-1}|\cdot] \sim \text{Gamma}(a_1, b_1)$.

5. **full conditional posterior distribution of $\boldsymbol{\eta}$**

The full conditional distribution of $\text{vec}(\boldsymbol{\eta})$ is of the form

$$
\begin{aligned}
&p(\text{vec}(\boldsymbol{\eta})|\cdot) \\
&\propto \prod_{i=1}^{n} p(\text{vec}(\boldsymbol{x_i^*})|\text{vec}(\boldsymbol{\eta}), \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \tilde{\boldsymbol{u}}_i, \phi) \times p(\text{vec}(\boldsymbol{\eta}|\sigma_0^2) \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\left[\boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\right]\boldsymbol{B}^\top\|_F^2\right\} \times \exp\left\{-\frac{1}{2\sigma_0^2}\|\boldsymbol{\eta}\|_F^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\text{vec}(\boldsymbol{\eta}) - \boldsymbol{m}\right]^\top \boldsymbol{C}^{-1}\left[\text{vec}(\boldsymbol{\eta}) - \boldsymbol{m}\right]\right\} \text{ by completing the square.}
\end{aligned}
\tag{A.9}
$$

This is a normal kernel with mean $\boldsymbol{M}$ and variance $\boldsymbol{E}$, where

$$
\begin{aligned}
\boldsymbol{E} &= (\frac{1}{\sigma_0^2} + n\phi)^{-1}\boldsymbol{I}, \\
\boldsymbol{M} &= \boldsymbol{E}\left(\sum_{i=1}^{n} \phi\left[(\boldsymbol{B}\otimes\boldsymbol{A})^\top \text{vec}(\boldsymbol{x}_i^*) - (\boldsymbol{\xi}\otimes\boldsymbol{\lambda})^{1/2}\text{vec}(\tilde{\boldsymbol{u}}_i)\right]\right).
\end{aligned}
$$

Therefore, $[\text{vec}(\boldsymbol{\eta})|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$.

6. **full conditional posterior distribution of $\boldsymbol{\lambda}^{1/2}, \mu_\lambda, \sigma_\lambda^2$**

   (a) With $\boldsymbol{\lambda}^{1/2} = \text{diag}(\lambda_1^{1/2}, \ldots, \lambda_{p_0}^{1/2})$, define $\lambda_l^{1/2}$ as the element at the l^{th} row and the l^{th} column of $\boldsymbol{\lambda}^{1/2}$. The full conditional distribution of $\lambda_l^{1/2}$ is

$$
\begin{aligned}
&p(\lambda_l^{1/2}|\cdot) \\
&\propto \prod_{i=1}^{n} p(\text{vec}(\boldsymbol{x_i^*})|\text{vec}(\boldsymbol{\eta}), \boldsymbol{A}, \boldsymbol{B}, \lambda_l^{1/2}, \boldsymbol{\xi}, \tilde{\boldsymbol{u}}_i, \phi) \times p(\lambda_l^{1/2}|\mu_\lambda, \sigma_\lambda^2) \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{\phi}{2}\|\boldsymbol{A}_{[,l]}\boldsymbol{x}_i^*\boldsymbol{B} - \boldsymbol{\eta}_{[l,]} - \lambda_l^{1/2}\tilde{\boldsymbol{u}}_{i[l,]}\boldsymbol{\xi}\|_F^2\right\} \times \exp\left\{-\frac{1}{2\sigma_\lambda^2}(\lambda_l^{1/2} - \mu_\lambda)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\lambda_l^{1/2} - m\right]^\top C^{-1}\left[\lambda_l^{1/2} - m\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.10}
$$

leading to a normal kernel with mean$M$ and variance C, where

$$E = \left(\frac{1}{\sigma_\lambda^2} + \phi \sum_{i=1}^{n} \tilde{\boldsymbol{u}}_{i[l,]} \boldsymbol{\xi} \tilde{\boldsymbol{u}}_{i[l,]}^\top \right)^{-1},$$
$$M = E \left(\frac{\mu_\lambda}{\sigma_\lambda^2} + \phi \sum_{i=1}^{n} \tilde{\boldsymbol{u}}_{i[l,]} \boldsymbol{\xi}^{1/2} \left[\boldsymbol{A}_{[,l]}^\top \boldsymbol{x}_i^* \boldsymbol{B} - \boldsymbol{\eta}_{[l,]} \right]^\top \right).$$

Therefore, $\left[\lambda_l^{1/2} | \cdot \right] \sim \mathcal{N}(M, E)$.

(b) The full conditional posterior distribution of μ_λ is derived as

$$\begin{aligned}
&p(\mu_\lambda | \cdot) \\
&\propto \prod_{l=1}^{p_0} p(\lambda_l^{1/2} | \mu_\lambda \sigma_\lambda^2) \times p(\mu_\lambda | \sigma_0^2) \\
&\propto \prod_{l=1}^{p_0} \exp\left\{ -\frac{1}{2\sigma_\lambda^2} (\lambda_l^{1/2} - \mu_\lambda)^2 \right\} \times \exp\left\{ -\frac{\mu_\lambda^2}{2\sigma_\lambda^2} \right\} \\
&\propto \exp\left\{ -\frac{1}{2} \left[\mu_\lambda - m \right]^\top C^{-1} \left[\mu_\lambda - m \right] \right\} \text{ by completing the square,}
\end{aligned} \tag{A.11}$$

corresponding to a normal kernel with mean M and variance E, where

$$E = (\frac{1}{\sigma_0^2} + \frac{p_0}{\sigma_\lambda^2})^{-1},$$
$$M = E \frac{\sum_{l=1}^{p_0} \lambda_l^{1/2}}{\sigma_\lambda^2}.$$

Thus, $[\mu_\lambda | \cdot] \sim \mathcal{N}(M, E)$.

(c) The posterior distribution of σ_λ^2 is given as

$$\begin{aligned}
&p(\sigma_\lambda^2 | \cdot) \\
&\propto \prod_{l=1}^{p_0} p(\lambda_l^{1/2} | \mu_\lambda \sigma_\lambda^2) \times p(\sigma_\lambda^2 | a_0, b_0) \\
&\propto \prod_{l=1}^{p_0} \exp\left\{ -\frac{(\lambda_l^{1/2} - \mu_\lambda)^2}{2\sigma_\lambda^2} \right\} \times (\sigma_\lambda^2)^{(-a_0-1)} \exp\left(-\frac{b_0}{\sigma_\lambda^2} \right) \\
&\propto (\sigma_\lambda^2)^{(-a_1-1)} \exp\left(-\frac{b_1}{\sigma_\lambda^2} \right),
\end{aligned} \tag{A.12}$$

corresponding to an inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$a_1 = a_0 + p_0/2,$$
$$b_1 = b_0 + \frac{1}{2}\sum_{l=1}^{p_0}(\lambda_l^{1/2} - \mu_\lambda)^2.$$

Thus, $[(\sigma_\lambda^2)^{-1}|\cdot] \sim \text{Gamma}(a_1, b_1)$.

7. **full conditional posterior distribution of $\boldsymbol{\xi}^{1/2}, \mu_\xi, \sigma_\xi^2$**

(a) With $\boldsymbol{\xi}^{1/2} = \text{diag}(\xi^{1/2}, \ldots, \xi_{q_0}^{1/2})$, define $\xi_s^{1/2}$ as the element at the s^{th} row and the s^{th} column of $\boldsymbol{\xi}^{1/2}$, $s = 1, \ldots, q_0$. The full conditional distribution of $\xi_s^{1/2}$ is

$$\begin{aligned}
&p(\xi_l^{1/2}|\cdot) \\
&\propto \prod_{i=1}^{n} p(\text{vec}(\boldsymbol{x_i^*})|\text{vec}(\boldsymbol{\eta}), \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\lambda}, \xi_s^{1/2}, \tilde{\boldsymbol{u}}_i, \phi) \times p(\xi_s^{1/2}|\mu_\xi, \sigma_\xi^2) \\
&\propto \prod_{i=1}^{n} \exp\left\{-\frac{\phi}{2}\|\boldsymbol{A}\top \boldsymbol{x}_i^* \boldsymbol{B}_{[,s]} - \boldsymbol{\eta}_{[,s]} - \boldsymbol{\lambda}\tilde{\boldsymbol{u}}_{i[l,]}\xi_s^{1/2}\|_F^2\right\} \times \exp\left\{-\frac{1}{2\sigma_\xi^2}(\xi_s^{1/2} - \mu_\xi)^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\xi_s^{1/2} - m\right]^\top C^{-1}\left[\xi_s^{1/2} - m\right]\right\} \text{ by completing the square,}
\end{aligned} \tag{A.13}$$

leading to a normal kernel with mean$M$ and variance C, where

$$E = \left(\frac{1}{\sigma_\xi^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[,s]}^\top \boldsymbol{\lambda}\tilde{\boldsymbol{u}}_{i[,s]}\right)^{-1},$$
$$M = E\left(\frac{\mu_\xi}{\sigma_\xi^2} + \phi\sum_{i=1}^{n}\tilde{\boldsymbol{u}}_{i[,s]}^\top \boldsymbol{\lambda}^{1/2}\left[\boldsymbol{A}^\top \boldsymbol{x}_i^* \boldsymbol{B}_{[,s]} - \boldsymbol{\eta}_{[,s]}\right]\right).$$

Therefore, $\left[\xi_s^{1/2}|\cdot\right] \sim \mathcal{N}(M, E)$.

(b) The full conditional posterior distribution of μ_ξ is derived as

$$\begin{aligned}
&p(\mu_\xi|\cdot) \\
&\propto \prod_{s=1}^{q_0} p(\xi_s^{1/2}|\mu_\xi\sigma_\xi^2) \times p(\mu_\xi|\sigma_0^2) \\
&\propto \prod_{s=1}^{q_0} \exp\left\{-\frac{1}{2\sigma_\xi^2}(\xi_l^{1/2}-\mu_\xi)^2\right\} \times \exp\left\{-\frac{\mu_\xi^2}{2\sigma_\xi^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\mu_\xi - m\right]^\top C^{-1}\left[\mu_\xi - m\right]\right\} \text{ by completing the square,}
\end{aligned} \tag{A.14}$$

corresponding to a normal kernel with mean M and variance E, where

$$\begin{aligned}
E &= (\frac{1}{\sigma_0^2} + \frac{q_0}{\sigma_\xi^2})^{-1}, \\
M &= E\frac{\sum_{s=1}^{q_0}\xi_s^{1/2}}{\sigma_\xi^2}
\end{aligned}$$

Thus, $[\mu_\xi|\cdot] \sim \mathcal{N}(M, E)$.

(c) The posterior distribution of σ_ξ^2 is given as

$$\begin{aligned}
&p(\sigma_\xi^2|\cdot) \\
&\propto \prod_{s=1}^{q_0} p(\xi_s^{1/2}|\mu_\xi\sigma_\xi^2) \times p(\sigma_\xi^2|a_0, b_0) \\
&\propto \prod_{s=1}^{q_0} \exp\left\{-\frac{(\xi_s^{1/2}-\mu_\xi)^2}{2\sigma_\xi^2}\right\} \times (\sigma_\xi^2)^{(-a_0-1)}\exp\left(-\frac{b_0}{\sigma_\xi^2}\right) \\
&\propto (\sigma_\xi^2)^{(-a_1-1)}\exp\left(-\frac{b_1}{\sigma_\xi^2}\right),
\end{aligned} \tag{A.15}$$

corresponding to an inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$\begin{aligned}
a_1 &= a_0 + q_0/2, \\
b_1 &= b_0 + \frac{1}{2}\sum_{s=1}^{q_0}(\xi_s^{1/2}-\mu_\xi)^2.
\end{aligned}$$

Thus, $\left[(\sigma_\xi^2)^{-1}|\cdot\right] \sim \text{Gamma}(a_1, b_1)$.

8. **full conditional posterior distribution of $\tilde{\boldsymbol{u}}_i, \boldsymbol{u}_i$**

(a) To derive the full conditional distribution of $\text{vec}(\tilde{\boldsymbol{u}}_i)$, recall that we have introduced a latent variable $w_i \sim \text{N}(\psi + \boldsymbol{\gamma}^\top \boldsymbol{z}_i + \boldsymbol{\theta}^\top \text{vec}(\boldsymbol{u}_i), 1)$ when justify the identifiability, and that $\tilde{w}_i = w_i - \psi - \boldsymbol{\gamma}^\top \boldsymbol{z}_i$. Then, the posterior distribution of $\tilde{\boldsymbol{u}}_i$ is of the form:

$$
\begin{aligned}
&p(\text{vec}(\tilde{\boldsymbol{u}}_i)|\cdot) \\
&\propto p(\text{vec}(\boldsymbol{x}_i^*)|\boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\lambda}, \boldsymbol{\xi}, \text{vec}(\tilde{\boldsymbol{u}}_i), \phi) \times p(w_i|\text{vec}(\tilde{\boldsymbol{u}}_i), \psi, \boldsymbol{\gamma}, \boldsymbol{\theta}) \times p(\text{vec}(\tilde{\boldsymbol{u}}_i)) \\
&\propto \exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\left[\boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2}\right]\boldsymbol{B}^\top\|_F^2\right\} \times \exp\left\{-\frac{1}{2}\left(\tilde{w}_i - \boldsymbol{\theta}^\top \text{vec}(\tilde{\boldsymbol{u}}_i)\right)^2\right\} \\
&\times \exp\left\{-\frac{1}{2}\|\tilde{\boldsymbol{u}}_i\|_F^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\text{vec}(\tilde{\boldsymbol{u}}_i) - \boldsymbol{m}\right]^\top \boldsymbol{C}^{-1}\left[\text{vec}(\tilde{\boldsymbol{u}}_i) - \boldsymbol{m}\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.16}
$$

corresponding to a normal kernel with mean $\boldsymbol{M}$ and variance $\boldsymbol{E}$, where

$$
\begin{aligned}
\boldsymbol{E} &= (\boldsymbol{I} + \phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda}) + \boldsymbol{\theta}\boldsymbol{\theta}^\top)^{-1}, \\
\boldsymbol{M} &= \boldsymbol{E}\left(\phi(\boldsymbol{\xi} \otimes \boldsymbol{\lambda})^{1/2}\left[\boldsymbol{B} \otimes \boldsymbol{A}\right)^\top \text{vec}(\boldsymbol{x}_i^*) - \text{vec}(\boldsymbol{\eta})\right] + \boldsymbol{\theta}\tilde{w}_i\right),
\end{aligned}
$$

Therefore, $[\text{vec}(\tilde{\boldsymbol{u}}_i)|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$.

(b) Recall that $\tilde{\boldsymbol{u}}_i$ is standardized from $\boldsymbol{u}_i$. Therefore, the posterior sample of $\boldsymbol{u}_i$ can be obtained through

$$
\boldsymbol{u}_i = \boldsymbol{\eta} + \boldsymbol{\lambda}^{1/2}\tilde{\boldsymbol{u}}_i\boldsymbol{\xi}^{1/2} \tag{A.17}
$$

9. **full conditional posterior distribution of ϕ**

The posterior distribution of ϕ is obtained as follows:

$$
\begin{aligned}
& p(\phi|\cdot) \\
& \propto \prod_{i=1}^{n} p(\boldsymbol{x_i^*}|\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{u}_i, \phi) \times p(\phi|a_0, b_0) \\
& \propto \prod_{i=1}^{n} \exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2\right\} \times ((1/\phi)^2)^{(-a_0-1)} \exp\left(-\frac{b_0}{(1/\phi)^2}\right) \\
& \propto ((1/\phi)^2)^{(-a_1-1)} \exp\left(-\frac{b_1}{(1/\phi)^2}\right), \qquad \text{(A.18)}
\end{aligned}
$$

leading to a inverse gamma kernel with $\alpha = a_1$ and $\beta = b_1$, where

$$
\begin{aligned}
a_1 &= a_0 + npq/2, \\
b_1 &=, b_0 + \frac{1}{2}\sum_{i=1}^{n} \|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2.
\end{aligned}
$$

Hence, $[\phi|\cdot] \sim \text{Gamma}(a_1, b_1)$

10. **full conditional posterior distribution of $\boldsymbol{A}$**

    As suggested by Hoff (2007), to derive the full posterior distribution of $\boldsymbol{A}$, we first rewrite the full likelihood based on the complete data (3.11) as function of $\boldsymbol{A}_{[,-j]}$ and $\boldsymbol{A}_{[,j]}$. Recall that $\boldsymbol{A}_{[,-j]}$ is the matrix $\boldsymbol{A}$ without its j^{th} column and $\boldsymbol{A}_{[,j]}$ is the removed j^{th} column of $\boldsymbol{A}$. With previously determined notations $\boldsymbol{B}_{[,k]}$ representing k^{th} column of matrix $\boldsymbol{B}$ and $\boldsymbol{B}_{[,-k]}$ referring to matrix $\boldsymbol{B}$ without the k^{th} column, We further define that

$$
\boldsymbol{x}_i^{*-j} = \boldsymbol{x}_i^* - \sum_{j' \neq j} \sum_{k=1}^{q_0} \boldsymbol{A}_{[,j']} \boldsymbol{u}_{i[j',k]} \boldsymbol{B}_{[,k]}^\top, \qquad \text{(A.19)}
$$

where $\boldsymbol{u}_{i[j,k]}$ denote the element in the j^{th} row and k^{th} column of $\boldsymbol{u}_i$

(Jiang et al., 2020). As a consequence,

$$
\begin{aligned}
&\sum_{i=1}^{n} \|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^* - \sum_{j=1}^{p_0}\sum_{k=1}^{q_0} \boldsymbol{A}_{[,j]}\boldsymbol{u}_{i[j,k]}\boldsymbol{B}_{[,k]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^* - \boldsymbol{A}_{[,j]}\sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]}\boldsymbol{B}_{[,k]}^\top - \sum_{j'\neq j}\sum_{k=1}^{q_0} \boldsymbol{A}_{[,j']}\boldsymbol{u}_{i[j',k]}\boldsymbol{B}_{[,k]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^{*-j} - \boldsymbol{A}_{[,j]}\sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]}\boldsymbol{B}_{[,k]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^{*-j}\|_F^2 - 2\boldsymbol{A}_{[,j]}^\top\sum_{i=1}^{n}\sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]}\boldsymbol{x}_i^{*-j}\boldsymbol{B}_{[,k]} + \sum_{i=1}^{n}\sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]}^2. \quad \text{(A.20)}
\end{aligned}
$$

It follows that the likelihood 3.11 can be rewritten as

$$
\begin{aligned}
&f(\boldsymbol{o}, \tilde{\boldsymbol{u}}, \boldsymbol{x}_m, \boldsymbol{x}^*, |\boldsymbol{z}, \boldsymbol{\nu}) \\
&\quad = \prod_{i=1}^{n} \left[p_i^{\mathcal{I}(o_i=1)} (1-p_i)^{\mathcal{I}(o_i=0)} \right] \\
&\quad \times \prod_{i=1}^{n} \prod_{m=1}^{M} \left[\frac{1}{(2\pi)^{pq/2} |\boldsymbol{\Sigma_i}|^{1/2}} \exp\left\{ -\frac{1}{2} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right)^\top \boldsymbol{\Sigma}_i^{-1} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right) \right\} \right] \\
&\quad \times \prod_{i=1}^{n} \left(\frac{\phi}{2\pi} \right)^{1/2} \exp\left\{ -\frac{\phi}{2} \| \boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i \boldsymbol{B}^\top \|_F^2 \right\} \\
&\quad \times \prod_{i=1}^{n} \frac{1}{(2\pi)^{p_0 q_0/2}} \exp\left\{ -\frac{1}{2} \mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top \mathrm{vec}(\tilde{\boldsymbol{u}}_i) \right\} \\
&\quad = \prod_{i=1}^{n} \left[p_i^{\mathcal{I}(o_i=1)} (1-p_i)^{\mathcal{I}(o_i=0)} \right] \\
&\quad \times \prod_{i=1}^{n} \prod_{m=1}^{M} \left[\frac{1}{(2\pi)^{pq/2} |\boldsymbol{\Sigma_i}|^{1/2}} \exp\left\{ -\frac{1}{2} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right)^\top \boldsymbol{\Sigma}_i^{-1} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right) \right\} \right] \\
&\quad \times \left(\frac{\phi}{2\pi} \right)^{n/2} \exp\left\{ -\frac{\phi}{2} \sum_{i=1}^{n} \| \boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i \boldsymbol{B}^\top \|_F^2 \right\} \\
&\quad \times \prod_{i=1}^{n} \frac{1}{(2\pi)^{p_0 q_0/2}} \exp\left\{ -\frac{1}{2} \mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top \mathrm{vec}(\tilde{\boldsymbol{u}}_i) \right\} \\
&\quad = \prod_{i=1}^{n} \left[p_i^{\mathcal{I}(o_i=1)} (1-p_i)^{\mathcal{I}(o_i=0)} \right] \\
&\quad \times \prod_{i=1}^{n} \prod_{m=1}^{M} \left[\frac{1}{(2\pi)^{pq/2} |\boldsymbol{\Sigma_i}|^{1/2}} \exp\left\{ -\frac{1}{2} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right)^\top \boldsymbol{\Sigma}_i^{-1} \left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*) \right) \right\} \right] \\
&\quad \times \left(\frac{\phi}{2\pi} \right)^{n/2} \exp\left\{ -\frac{\phi}{2} \left(\sum_{i=1}^{n} \| \boldsymbol{x}_i^{*-j} \|_F^2 - 2\boldsymbol{A}_{[,j]}^\top \sum_{i=1}^{n} \sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]} \boldsymbol{x}_i^{*-j} \boldsymbol{B}_{[,k]} + \sum_{i=1}^{n} \sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]}^2 \right) \right\} \\
&\quad \times \prod_{i=1}^{n} \frac{1}{(2\pi)^{p_0 q_0/2}} \exp\left\{ -\frac{1}{2} \mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top \mathrm{vec}(\tilde{\boldsymbol{u}}_i) \right\}. \qquad \text{(A.21)}
\end{aligned}
$$

From Hoff (2007), with the conditional distribution $p(\boldsymbol{A}_{[,j]}|\boldsymbol{A}_{[,-j]}) \overset{d}{=} \boldsymbol{N}_{A[,-j]}\boldsymbol{a}_j$ implied from a uniform prior, the full conditional posterior distribution of $\boldsymbol{a}_j$ is proportional to

$$
\exp\left(\phi \boldsymbol{a}_j^\top \boldsymbol{N}_{A[,-j]}^\top \sum_{i=1}^{n} \sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]} \boldsymbol{x}_i^{-j} \boldsymbol{B}_{[,k]} \right).
$$

This expression corresponds to a von Mises-Fisher distribution on the $(p-(p_0-1))$-dimensional unit sphere with notation vMF($\boldsymbol{\eta}^A$), where $\boldsymbol{\eta}^A \in \mathcal{R}^{p-p_0+1}$ and $\boldsymbol{\eta}^A = \phi \boldsymbol{N}_{A[,-j]}^\top \sum_{i=1}^n \sum_{k=1}^{q_0} \boldsymbol{u}_{i[j,k]} \boldsymbol{x}_i^{-j} \boldsymbol{B}_{[,k]}$. A uniform distribution on the sphere is a special case of von Mises-Fisher distribution (Hoff, 2007; Jiang et al., 2020). The general probability density function for a von Mises-Fisher distribution for a vector $\vec{\boldsymbol{u}}$ on the p-dimensional unit sphere is of the form:

$$f_p(\vec{\boldsymbol{u}}; \boldsymbol{\mu}) = c_p(||\boldsymbol{\mu}||)\exp\{\vec{\boldsymbol{u}}^\top \boldsymbol{\mu}\}$$

with parameter $\boldsymbol{\mu} \in \mathcal{R}^p$ and is denoted by vMF($\boldsymbol{\mu}$). Here, c_p is a normalizing constants for $f_p(\vec{\boldsymbol{u}}; \boldsymbol{\mu})$. Therefore, the density function for $\boldsymbol{A}$'s full conditional posterior distribution is

$$f_p(\boldsymbol{a}_j; \boldsymbol{\eta}^{\boldsymbol{A}}) = c_{p-p_0+1}(||\boldsymbol{\eta}^{\boldsymbol{A}}||)\exp\{\boldsymbol{a}_j^\top \boldsymbol{\eta}^{\boldsymbol{A}}\}.$$

Therefore, similar as Jiang et al. (2020), a posterior sample of $\boldsymbol{A}_{[,j]}$ can be drawn from the above posterior distributions in following ways:

(a) Draw a sample of $\boldsymbol{a}_j \sim$ vMF($\boldsymbol{\eta}^{\boldsymbol{A}}$);

(b) Find $\boldsymbol{N}_{A[,-j]}^\top$ and set $\boldsymbol{N}_{A[,-j]}^\top \boldsymbol{a}_j$ as a posterior sample of $\boldsymbol{A}_{[,j]}$.

11. **full conditional posterior distribution of $\boldsymbol{B}$**

Similar as the analysis for $\boldsymbol{A}$, with previously determined notations $\boldsymbol{B}_{[,k]}$ representing k^{th} column of matrix $\boldsymbol{B}$ and $\boldsymbol{B}_{[,-k]}$ referring to matrix $\boldsymbol{B}$ without the k^{th} column, we further let

$$\tilde{\boldsymbol{x}}_i^{-k} = \boldsymbol{x}_i^\top - \boldsymbol{\mu}^\top - \sum_{k' \neq k} \sum_{j=1}^{p_0} \boldsymbol{B}_{[,k']} \boldsymbol{u}_{i[j,k']} \boldsymbol{A}_{[,j]}^\top. \tag{A.22}$$

Recall that $\boldsymbol{u}_{i[j,k]}$ denote the element in the j^{th} row and k^{th} column of $\boldsymbol{u}_i$, $\boldsymbol{A}_{[,-j]}$ is the matrix $\boldsymbol{A}$ without its j^{th} column and $\boldsymbol{A}_{[,j]}$ is the removed j^{th} column of $\boldsymbol{A}$ (Jiang et al., 2020). As a consequence, the term with Frobenius norm in the complete likelihood (model 3.11) can

be transformed as:

$$
\begin{aligned}
&\sum_{i=1}^{n} \|\boldsymbol{x}_i^* - \boldsymbol{\mu} - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^{*\top} - \boldsymbol{\mu}^\top - \sum_{j=1}^{p_0}\sum_{k=1}^{q_0} \boldsymbol{B}_{[,k]}\boldsymbol{u}_{i[j,k]}\boldsymbol{A}_{[,j]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\boldsymbol{x}_i^{*\top} - \boldsymbol{\mu}^\top - \boldsymbol{B}_{[,k]}\sum_{j=1}^{p_0} \boldsymbol{u}_{i[j,k]}\boldsymbol{A}_{[,j]}^\top - \sum_{k'\neq k}\sum_{j=1}^{p_0} \boldsymbol{B}_{[,k']}\boldsymbol{u}_{i[j,k']}\boldsymbol{A}_{[,j]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\tilde{\boldsymbol{x}}_i^{*-k} - \boldsymbol{B}_{[,k]}\sum_{j=1}^{p_0} \boldsymbol{u}_{i[j,k]}\boldsymbol{A}_{[,j]}^\top\|_F^2 \\
=&\sum_{i=1}^{n} \|\tilde{\boldsymbol{x}}_i^{*-k}\|_F^2 - 2\boldsymbol{B}_{[,k]}^\top\sum_{i=1}^{n}\sum_{j=1}^{p_0} \boldsymbol{u}_{i[j,k]}\tilde{\boldsymbol{x}}_i^{*-k}\boldsymbol{A}_{[,j]} + \sum_{i=1}^{n}\sum_{j=1}^{p_0} \boldsymbol{u}_{i[j,k]}^2. \quad \text{(A.23)}
\end{aligned}
$$

Accordingly, the complete likelihood can be then rewritten in the form:

$$
\begin{aligned}
&f(\boldsymbol{o}, \tilde{\boldsymbol{u}}, \boldsymbol{x}_m, \boldsymbol{x}^*, |\boldsymbol{z}, \boldsymbol{\nu}) \\
&\quad= \prod_{i=1}^{n}\left[p_i^{\mathcal{I}(o_i=1)}(1-p_i)^{\mathcal{I}(o_i=0)}\right] \\
&\quad\times \prod_{i=1}^{n}\prod_{m=1}^{M}\left[\frac{1}{(2\pi)^{pq/2}|\boldsymbol{\Sigma_i}|^{1/2}}\exp\left\{-\frac{1}{2}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)^\top \boldsymbol{\Sigma}_i^{-1}\left(\mathrm{vec}(\boldsymbol{x}_{im}) - \mathrm{vec}(\boldsymbol{x}_i^*)\right)\right\}\right] \\
&\quad\times \left(\frac{\phi}{2\pi}\right)^{n/2}\exp\left\{-\frac{\phi}{2}\left(\sum_{i=1}^{n}\|\tilde{\boldsymbol{x}}_i^{*-k}\|_F^2 - 2\boldsymbol{B}_{[,k]}^\top\sum_{i=1}^{n}\sum_{j=1}^{p_0}\boldsymbol{u}_{i[j,k]}\tilde{\boldsymbol{x}}_i^{*-k}\boldsymbol{A}_{[,j]} + \sum_{i=1}^{n}\sum_{j=1}^{p_0}\boldsymbol{u}_{i[j,k]}^2\right)\right\} \\
&\quad\times \prod_{i=1}^{n}\frac{1}{(2\pi)^{p_0q_0/2}}\exp\left\{-\frac{1}{2}\mathrm{vec}(\tilde{\boldsymbol{u}}_i)^\top\mathrm{vec}(\tilde{\boldsymbol{u}}_i)\right\}. \quad \text{(A.24)}
\end{aligned}
$$

Similarly, as clarified above, with the conditional distribution $p(\boldsymbol{B}_{[,k]}|\boldsymbol{B}_{[,-k]}) \overset{d}{=} \boldsymbol{N}_{B[,-k]}\boldsymbol{b}_k$ implied from a uniform prior, the full conditional posterior distribution of $\boldsymbol{b}_k$, is proportional to

$$\exp\{\boldsymbol{b}_k^\top(\phi\boldsymbol{N}_{B\{-k\}}^\top\sum_{i=1}^{n}\sum_{j=1}^{p_0}\boldsymbol{u}_{i[j,k]}\tilde{\boldsymbol{x}}_i^{-k}\boldsymbol{A}_{[,j]})\},$$

leading to a von Mises-Fisher distribution vMF($\boldsymbol{\eta}^B$), where $\boldsymbol{\eta}^B \in \mathcal{R}^{q-(q_0-1)}$ and $\boldsymbol{\eta}^B = \phi\boldsymbol{N}_{B\{-k\}}^\top\sum_{i=1}^{n}\sum_{j=1}^{p_0}\boldsymbol{u}_{i[j,k]}\tilde{\boldsymbol{x}}_i^{-k}\boldsymbol{A}_{[,j]}$ (Jiang et al., 2020). The corresponding density function for $\boldsymbol{B}$'s full conditional posterior distribution is

$$f_p(\boldsymbol{b}_k; \boldsymbol{\eta}^{\boldsymbol{B}}) = c_{q-q_0+1}(||\boldsymbol{\eta}^{\boldsymbol{B}}||)\exp\{\boldsymbol{b}_k^\top\boldsymbol{\eta}^{\boldsymbol{B}}\}.$$

Therefore, similar as Jiang et al. (2020), a posterior sample of $\boldsymbol{B}_{[,j]}$ can be drawn from the above posterior distributions through the following procedure:

(a) Draw a sample of $\boldsymbol{b}_k \sim \text{vMF}(\boldsymbol{\eta}^B)$;

(b) Find $\boldsymbol{N}^\top_{B[,-k]}$ and set $\boldsymbol{N}^\top_{B[,-k]}\boldsymbol{b}_k$ as a posterior sample of $\boldsymbol{B}_{[,k]}$.

12. **full conditional posterior distribution of $\boldsymbol{x}_i^*$**

The posterior distribution of $\boldsymbol{x}_i^*$ is derived as

$$
\begin{aligned}
&p(\text{vec}(\boldsymbol{x}_i^*)|\cdot) \\
&\propto \prod_{m=1}^{M} p(\boldsymbol{x}_{im}|\boldsymbol{x}_i^*, \widetilde{\boldsymbol{\Sigma}}_i) \times p(\text{vec}(\boldsymbol{x}_{\boldsymbol{i}}^{\boldsymbol{*}})|\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{u}_i, \phi) \\
&\propto \prod_{m=1}^{M} \exp\left\{-\frac{1}{2}(\text{vec}(\boldsymbol{x}_{im}) - \text{vec}(\boldsymbol{x}_i^*))^\top \left[\text{diag}(\text{vec}(\widetilde{\boldsymbol{\Sigma}}_i))\right]^{-1} (\text{vec}(\boldsymbol{x}_{im}) - \text{vec}(\boldsymbol{x}_i^*))\right\} \\
&\times \exp\left\{-\frac{\phi}{2}\|\boldsymbol{x}_i^* - \boldsymbol{A}\boldsymbol{u}_i\boldsymbol{B}^\top\|_F^2\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\text{vec}(\boldsymbol{x}_i^*) - \boldsymbol{m}\right]^\top \boldsymbol{C}^{-1}\left[\text{vec}(\boldsymbol{x}_i^*) - \boldsymbol{m}\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.25}
$$

corresponding to a normal kernel with mean $\boldsymbol{M}$ and variance $\boldsymbol{E}$, where

$$
\begin{aligned}
\boldsymbol{E} &= \left(\phi\boldsymbol{I} + \left[\text{diag}(\text{vec}(\widetilde{\boldsymbol{\Sigma}}_i))\right]^{-1}\right)^{-1}, \\
\boldsymbol{M} &= \boldsymbol{E}(\phi(\boldsymbol{B}\otimes\boldsymbol{A})\text{vec}(\boldsymbol{u}_i) + \sum_{m=1}^{2}\boldsymbol{\Sigma}_i^{-1}\text{vec}(\boldsymbol{x}_{im})).
\end{aligned}
$$

Therefore, $[\text{vec}(\boldsymbol{x}_i^*)|\cdot] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$.

13. **full conditional posterior distribution of w_i**

Recall that, in chapter 3, we have introduced $o_i = \mathcal{I}(w_i > 0)$ with an indicator function $\mathcal{I}(\cdot)$ and $w_i \sim \text{N}(\psi + \boldsymbol{\gamma}^\top\boldsymbol{z}_i + \boldsymbol{\theta}^\top\text{vec}(\boldsymbol{u}_i), 1)$. Therefore, the posterior sample for w_i is obtained as follows:

$$
\begin{aligned}
[w_i|\cdot] &\sim \mathcal{N}(\psi + \boldsymbol{\gamma}^T\boldsymbol{z}_i + \boldsymbol{\theta}^\top\text{vec}(\tilde{\boldsymbol{u}}_i), 1)\mathcal{I}(w_i > 0) \text{ if } o_i = 1 \\
[w_i|\cdot] &\sim \mathcal{N}(\psi + \boldsymbol{\gamma}^T\boldsymbol{z}_i + \boldsymbol{\theta}^\top\text{vec}(\tilde{\boldsymbol{u}}_i), 1)\mathcal{I}(w_i < 0) \text{ if } o_i = 0
\end{aligned}
$$

14. **full conditional posterior distribution of** $(\psi, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top$

To derive the posterior distribution of $(\psi, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top$, we let $\tilde{\boldsymbol{z}}_i$ represent $(1, \boldsymbol{z}_i^\top, \text{vec}(\tilde{\boldsymbol{u}}_i)^\top)^\top$ and let $\tilde{\boldsymbol{\theta}}$ denote $(\psi, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top$.It follows that

$$
\begin{aligned}
& p(\tilde{\boldsymbol{\theta}}|\cdot) \\
& \propto \prod_{i=1}^{n} p(w_i|\tilde{\boldsymbol{z}}_i, \tilde{\boldsymbol{\theta}}) \times p(\tilde{\boldsymbol{\theta}}|\sigma_0^2) \\
& \propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}\|w_i - \tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{\theta}}\|_F^2\right\} \times \exp\left\{-\frac{1}{2\sigma_0^2}\|\tilde{\boldsymbol{\theta}}\|_F^2\right\} \\
& \propto \exp\left\{-\frac{1}{2}\left[\tilde{\boldsymbol{\theta}} - \boldsymbol{m}\right]^\top \boldsymbol{C}^{-1}\left[\tilde{\boldsymbol{\theta}} - \boldsymbol{m}\right]\right\} \text{ by completing the square,}
\end{aligned}
\tag{A.26}
$$

leading to a normal kernel with mean $\boldsymbol{M}$ and variance $\boldsymbol{E}$, where

$$
\begin{aligned}
\boldsymbol{E} &= (\frac{1}{\sigma_0^2}\boldsymbol{I} + \sum_{i=1}^{n} \tilde{\boldsymbol{z}}_i\tilde{\boldsymbol{z}}_i^\top)^{-1}, \\
\boldsymbol{M} &= \boldsymbol{E}\sum_{i=1}^{n} w_i\tilde{\boldsymbol{z}}_i.
\end{aligned}
$$

Therefore, $\left[(\psi, \boldsymbol{\gamma}^\top, \boldsymbol{\theta}^\top)^\top|\cdot\right] \sim \mathcal{N}(\boldsymbol{M}, \boldsymbol{E})$