# Bayesian Simultaneous Partial Envelope Model with Application to an Imaging Genetics Analysis

Yanbo SHEN, Yeonhee PARK, Saptarshi CHAKRABORTY, and Chunming ZHANG*

## Abstract

As a prominent dimension reduction method for multivariate linear regression, the envelope model has received increased attention over the past decade due to its modeling flexibility and success in enhancing estimation and prediction efficiencies. Several enveloping approaches have been proposed in the literature; among these, the partial response envelope model [57] that focuses on only enveloping the coefficients for predictors of interest, and the simultaneous envelope model [14] that combines the predictor and the response envelope models within a unified modeling framework, are noteworthy. In this article we incorporate these two approaches within a Bayesian framework, and propose a novel Bayesian simultaneous partial envelope model that generalizes and addresses some limitations of the two approaches. Our method offers the flexibility of incorporating prior information if available, and aids coherent quantification of all modeling uncertainty through the posterior distribution of model parameters. A block Metropolis-within-Gibbs algorithm for Markov chain Monte Carlo (MCMC) sampling from the posterior is developed. The utility of our model is corroborated by theoretical results, comprehensive simulations, and a real imaging genetics data application for the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

KEYWORDS AND PHRASES: Bayesian envelope model, Multivariate regression, Reducing subspace, Simultaneous envelope, Partial envelope, Imaging genetics.

## 1. INTRODUCTION

Consider the standard multivariate linear regression

$$\boldsymbol{Y} = \boldsymbol{\mu_Y} + \boldsymbol{\beta}^T \boldsymbol{X} + \boldsymbol{\epsilon_{Y|X}}, \qquad (1.1)$$

where the responses $\boldsymbol{Y} \in \mathbb{R}^r$ and the centered predictors $\boldsymbol{X} \in \mathbb{R}^p$ are both multivariate, the means of responses $\boldsymbol{\mu_Y} \in \mathbb{R}^r$, the regression coefficients $\boldsymbol{\beta} \in \mathbb{R}^{p \times r}$, and the random error vector $\boldsymbol{\epsilon_{Y|X}} \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{\Sigma_{Y|X}})$. It has applications in broad fields of scientific research, including economics [6], chemistry [35], agriculture [51], engineering [44], bioinformatics [34], etc. Specifically, we consider the application to the imaging genetics problem (IGP).

Thanks to the rapid technological development in medical imaging and genome sequencing, the diagnosis, prevention and treatment of mental illness have been greatly improved. Compared with traditional medical imaging analysis and genome-wide association studies (GWAS) that make use of only imaging or only genome data and were studied to a large extent [67, 63], another more promising yet more challenging study is to directly relate the imaging phenotypes (IPs) as responses to genotypes (GPs) as predictors of interest, while also considering the effects of some demographic covariates or other risk factors. This is called IGP and is less explored in the statistical community [42]. Because of

the Central Dogma that is often stated as *DNA makes RNA and RNA makes protein*, genetic markers should be promising in revealing the abnormal protein expression in cortical or sub-cortical structures of brain, whose volumes should be more informative quantitative traits than simple disease status.

Some early practices in studying IGP focus on univariate [54] or voxel-wise type of methods [25], which explores the relationship between every (voxel, SNP) pair (SNP: Single-nucleotide polymorphism), or between each individual voxel and all genetic markers together, respectively. These types of methods enjoy simplicity and can handle high dimensional imaging or genomic datasets, but cannot leverage spatial dependence among voxels. To this end, several multivariate regression models have been proposed to jointly relate tens of brain summary measures, like the volumes of Regions of Interest (ROIs), to hundreds of SNPs [65, 68, 23, 64, 46]. Among them, [46] is the only work that employs the envelope model to IGP, to the best of our knowledge.

As a dimension reduction method developed in the past decade, the first envelope model [15] aims at providing an efficient estimator of the regression coefficients $\boldsymbol{\beta}$ in the multivariate linear regression (1.1), by assuming that there are linear combinations of the responses $\boldsymbol{Y}$ that are not related to the regression, and removing its variation as immaterial information (see Section 2.2 for more details). With a great

*Corresponding author.

success in achieving efficient estimation and modeling flexibility, the original envelope model has been adapted or extended to predictor envelope [12], groupwise envelope [46], partial response envelope [57], simultaneous envelope [14], sparsity learning [56], matrix-valued [18] or tensor-valued [39] responses or predictors, quantile regression [19], generalized linear models [11], etc. The envelope methods were recently reviewed by [37] and more details were introduced in [13]. The aforementioned envelope methods are all developed from the frequentist perspective. However, the literature addressing Bayesian envelope methods is rare, but still interesting for its capabilities of incorporating prior information into inference, and implementing posterior uncertainty quantification, compared with the unnatural prior information incorporation and the bootstrap or asymptotic variance in the frequentist setting.

[33] introduced the first framework for the Bayesian envelope model. However, it is hard to apply this framework to other envelope contexts, since it requires the specific formulation of the response envelope to construct the conjugacy. Furthermore, sampling from the generalized matrix Bingham distributions and the truncated inverse gamma distributions is required in this framework, which makes the MCMC algorithm computationally expensive. Later, [5] proposed a new flexible and handy Bayesian framework for the envelope model, which bypasses the manifold restriction on the basis matrix of the envelope space by an unconstrained matrix re-parameterization. Since all parameters under this new framework are either vectors, or unconstrained, or positive definite matrices, the computation is much more efficient without sampling from the generalized matrix Bingham distributions and the truncated inverse gamma distributions.

The rapid development in the envelope family requires a model to unify various perspectives, and a more clear choice to practitioners on which envelope model to apply, especially when we have a particular interest for a subset of predictors that could be either continuous or discrete or a mix of them. The existing simultaneous envelope model [14] integrates both the response and predictor envelopes within a unified modeling framework. It simultaneously enjoys the benefits from two envelope components, i.e., more efficient estimation of the regression coefficients offered by the response envelope, and improved prediction efficiency of the responses obtained by the predictor envelope, but it cannot focus on the predictors of main interest, and therefore cannot leverage the idea from the partial response envelope [57] to obtain further estimation efficiency. More importantly, without additional assumptions imposed, it is guaranteed to outperform the predictor or the response envelope only for the Normal predictors ([14], Propositions 2 and 5). However, the discrete predictors will not follow the Normal distribution and including them in the simultaneous envelope by implicitly regarding them as Normal will lead to biased estimates and invalid inference. The partial response envelope

[57] focuses on enveloping only the coefficients of the predictors of main interest $X_1$ (where in (1.1), $X$ is partitioned into $(X_1^T, X_2^T)^T$, and $X_1$ denotes the predictor of main interest to us and $X_2$ are nuisance covariates), for example the genetic markers in IGP. Compared with full response envelope [15] that indistinguishably envelopes the coefficients of all predictors $X$, enhanced estimation efficiency is possible by this model since the envelope space could be shrunk if only concentrating on the coefficients of $X_1$, but it could not leverage the benefits of the predictor envelope [12], which is known to have the potential to increase prediction efficiency. The envelope model that is proposed in this paper could address these limitations while still providing a valid inference procedure.

To address the aforementioned limitations, and improve the efficiencies for the estimation of the regression coefficients of main interest (the coefficients of $X_1$, which we denote as $\beta_1$ in (3.2)) and the prediction of the responses ($Y$ in (3.2)), this paper proposes a new envelope model, called the simultaneous partial envelope model (3.7)–(3.10), in Section 3. The proposed envelope model combines the partial response envelope model [57] and the simultaneous envelope model [14] within a more general modeling framework, by further partitioning the predictors of main interest $X_1$ into $X_{1C}$ and $X_{1D}$, the continuous part and the discrete part. Our proposed envelope model simultaneously imposes the partial response envelope structure on the coefficients of $X_1$, and the partial predictor envelope only on the coefficients of $X_{1C}$, instead of the coefficients of whole $X_1$. By this construction, our proposed model could address the aforementioned limitations for the simultaneous and the partial response envelopes. Firstly, two envelope components that are simultaneously imposed on our model are partial envelopes, which both focus on the predictors of interest only. Secondly, it avoids the Normal requirement mentioned above for the discrete predictors in the simultaneous envelope model while still providing a valid inference, by only imposing the partial response envelope (i.e. no partial predictor envelope) on the coefficients of $X_{1D}$. It is also worthy to note that $X_{1C}$ is Normal with means depending on $X_{1D}$ in our model. Lastly, to further improve the estimation and prediction efficiencies upon the partial response envelope, the simultaneous envelope structure is considered in our model for the coefficients of $X_{1C}$, which turns out to enhance not only the estimation of the coefficients of $X_{1C}$, but also that of the coefficients of $X_{1D}$ in our simulation (see the comparisons on the estimation of the coefficents of $X_{1D}$ between our method and the Bayesian partial response envelope in Table 3 and Figure 2), as a possible synergetic effect.

Our proposed model includes a number of important models in the envelope family as special cases, including the (partial) response envelope, the (partial) predictor envelope and the simultaneous envelope (see Table 1 for their relationship). In Section 4, we develop an efficient Bayesian inference procedure based on [5], which allows convenient

prior information incorporation and variability quantification, compared with the frequentist envelope methods.

The contribution of this paper is fivefold.

1. We propose a new simultaneous partial envelope model (3.7)–(3.10) that unifies several existing envelope models as special cases. Our proposed model allows both predictors of interest $\boldsymbol{X}_1$ and nuisance covariates $\boldsymbol{X}_2$ to be either continuous or discrete or a mix of them. Compared with any envelopes that our model degenerates to, our method provides a more efficient while still valid estimator for the regression coefficients of $\boldsymbol{X}_1$ and improved prediction for $\boldsymbol{Y}$ in (3.2). These advantages address the limitations that are mentioned above for the simultaneous envelope model [14] and the partial response envelope model [57]. The corresponding block Metropolis-within-Gibbs algorithm is developed for the posterior inference.

2. We establish the theoretical properties including the posterior propriety for our model and the Harris ergodicity for our algorithm.

3. We are the first to investigate the performance of several popular dimension selection methods together, among the Bayesian envelope literature, to the best of our knowledge.

4. We apply our method to an imaging genetics analysis for the ADNI study. We show the improved prediction accuracy of our method compared with all other competitors, and obtain some scientifically meaningful findings from posterior selection. By incorporating the weak imaging genetics relationship as prior information, the prediction accuracy is further improved and the associated shrinkage effect is also investigated.

5. We build the R package SIMP that implements our algorithm, which is available at https://github.com/yanbowisc/SIMP.git.

Some notations frequently used in this paper are summarized here. For any $a = 1, 2, \ldots$, let $\mathbb{P}_{\mathcal{S}}$ be a projection matrix onto a subspace $\mathcal{S} \subseteq \mathbb{R}^a$ and $\mathbb{P}_{\mathcal{S}^\perp} = \mathbf{I}_a - \mathbb{P}_{\mathcal{S}}$ be the projection matrix onto $\mathcal{S}^\perp$, where $\mathcal{S}^\perp$ denotes the orthogonal complement of $\mathcal{S}$ and $\mathbf{I}_a$ denotes the $a$-dimensional identity matrix. Let $\mathbb{S}_+^{a \times a}$ denote the class of all $a \times a$ real valued positive definite matrices. Let $\mathbf{1}_n$ denote a vector of all ones with length $n$. For a matrix $\mathbf{A} \in \mathbb{R}^{a \times a}$ and a subspace $\mathcal{E} \subseteq \mathbb{R}^a$, $\mathbf{A}\mathcal{E}$ is defined as $\mathbf{A}\mathcal{E} = \{\mathbf{A}\boldsymbol{\epsilon} : \boldsymbol{\epsilon} \in \mathcal{E}\}$. Notations $|\mathbf{A}|$, $\|\mathbf{A}\|$, $\mathrm{vec}(\mathbf{A})$ and $\mathrm{span}(\mathbf{A})$ denote the determinant, the spectral norm, the vectorization by columns and the column space of the matrix $\mathbf{A}$ respectively. For $a_1, a_2, \ldots, a_n \in \mathbb{R}$ and a square matrix $\mathbf{S} \in \mathbb{R}^{a \times a}$, we describe a diagonal matrix by either specifying its diagonal elements by the notation $\mathrm{diag}(a_1, \ldots, a_n)$, or using the notation $\mathrm{diag}\{\mathbf{S}\}$ to indicate that the diagonal elements of the square matrix $\mathbf{S}$ will be taken out to constitute this diagonal matrix. Moreover, $\mathbf{S}_{ii}$ denotes the $i$-th diagonal element of the square matrix $\mathbf{S}$. For a random vector $\boldsymbol{Y}$, the distribution of $\boldsymbol{Y} \mid \boldsymbol{X}$ is interpreted

as the distribution of $\boldsymbol{Y}$ given the fixed value of $\boldsymbol{X}$ for a non-stochastic vector $\boldsymbol{X}$, or the distribution of $\boldsymbol{Y}$ conditional on $\boldsymbol{X}$ for a random vector $\boldsymbol{X}$. For $a, b, c = 1, 2, \ldots$, and random vectors $\boldsymbol{X} \in \mathbb{R}^a$, $\boldsymbol{Y} \in \mathbb{R}^b$, we use $\mathrm{Cov}((\boldsymbol{X}, \boldsymbol{Y}) \mid \boldsymbol{Z})$ to denote the $a \times b$ dimensional covariance matrix of $\boldsymbol{X}$ and $\boldsymbol{Y}$ conditional on a random (or given a non-stochastic) vector $\boldsymbol{Z} \in \mathbb{R}^c$. If multiple random (or non-stochastic) vectors are conditional on (or given), they are bracketed together after the vertical line, like $\boldsymbol{Y} \mid (\cdot)$ and $\mathrm{Cov}((\boldsymbol{X}, \boldsymbol{Y}) \mid (\cdot))$. Also, $:=$ and $\boldsymbol{X} \overset{\mathcal{D}}{=} \boldsymbol{Y}$ indicate equating by definition and random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ are equally distributed respectively. Lastly, for any $x, y \in \mathbb{R}$, $x \ll y$ indicates that $x$ is much less than $y$, while $x \gg y$ means that $x$ is much greater than $y$.

The rest of this paper is organized as follows. Section 2 reviews two existing envelope models, which are prototypes of two envelope components of our proposed model. Section 3 introduces the definition and the formulation of our proposed simultaneous partial envelope model. Section 4 develops a Bayesian inference procedure for our proposed model, with detailed procedure of MCMC algorithm left to Appendix A. Section 5 establishes the theoretical properties of our model and algorithm. Section 6 investigates the issue of model selection for our Bayesian envelope model. Sections 7 and 8 conduct comprehensive simulation studies and investigate a real data application respectively. Section 9 concludes our paper by further discussions.

## 2. REVIEW OF THE RESPONSE AND THE PREDICTOR ENVELOPE MODELS

### 2.1 Preliminaries

In this section, we introduce two definitions and one proposition, which are necessary for the formal introduction of the response envelope model [15] and the predictor envelope model [12]. For convenience, the definition of *reducing subspace* [10] is firstly introduced.

**Definition 1** ([10], Chapter II, Definition 3.5). *A subspace $\mathcal{E} \subseteq \mathbb{R}^a$ is called a reducing subspace of $\boldsymbol{\Sigma} \in \mathbb{R}^{a \times a}$ for some integer $a = 1, 2, \ldots$, if $\boldsymbol{\Sigma}\mathcal{E} \subseteq \mathcal{E}$ and $\boldsymbol{\Sigma}\mathcal{E}^\perp \subseteq \mathcal{E}^\perp$.*

If $\mathcal{E}$ is a reducing subspace of $\boldsymbol{\Sigma}$, there is a crucial decomposition of $\boldsymbol{\Sigma}$ as illustrated in the following proposition.

**Proposition 1** ([15], Proposition 2.1). *$\mathcal{E}$ is a reducing subspace of $\boldsymbol{\Sigma}$ if and only if $\boldsymbol{\Sigma}$ could be decomposed as $\boldsymbol{\Sigma} = \mathbb{P}_{\mathcal{E}}\boldsymbol{\Sigma}\mathbb{P}_{\mathcal{E}} + \mathbb{P}_{\mathcal{E}^\perp}\boldsymbol{\Sigma}\mathbb{P}_{\mathcal{E}^\perp}$, where $\mathbb{P}_{\mathcal{E}}$ and $\mathbb{P}_{\mathcal{E}^\perp}$ are the projection matrices onto subspaces $\mathcal{E}$ and $\mathcal{E}^\perp$.*

Next, for a positive definite matrix $\mathbf{M}$, we introduce the general definition of *the $\mathbf{M}$-envelope of a subspace $\mathcal{S}$*. It will be used in defining the response and the predictor envelope models in Section 2.2 and Section 2.3 and our proposed model in Section 3.

**Definition 2** ([15], Definition 2.1). *Let a matrix $\mathbf{M} \in \mathbb{S}_+^{a \times a}$ and a subspace $\mathcal{S} \subset \mathrm{span}(\mathbf{M})$. The $\mathbf{M}$-envelope of $\mathcal{S}$ is the*

*intersection of all reducing subspaces of* $\mathbf{M}$ *that contain* $\mathcal{S}$, *and is denoted by* $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$.

For convenience, we will frequently use the notation $\mathcal{E}_{\mathbf{M}}(\mathcal{S})$ to define envelope structures in the rest of this paper. Below we briefly review the response envelope model [15] in Section 2.2, and mention how this idea was adapted to define the predictor envelope model [12] in Section 2.3.

## 2.2 The Response Envelope Model

In this section, we review the definitions and the coordinate form of the response envelope model [15], which aims to improve the estimation efficiency of $\boldsymbol{\beta}$ in (1.1) by assuming some linear combinations of $\boldsymbol{Y}$ do not depend on $\boldsymbol{X}$. Here, we assume $\boldsymbol{X}$ in (1.1) to be non-stochastic. The efficiency gains come from removing those redundant variation in $\boldsymbol{Y}$ by estimating the associated linear combination coefficients. Formally, the response envelope model is defined by the smallest subspace $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$ that satisfies Condition 1.

**Condition 1** ([15], p. 928). *Assume in* (1.1), *the subspace* $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subset \mathbb{R}^r$ *satisfies*

(a) $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y} \mid \boldsymbol{X} \stackrel{\mathcal{D}}{=} \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$,
(b) $\mathrm{Cov}((\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}, \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}) \mid \boldsymbol{X}) = \boldsymbol{0}$.

Condition 1(a) assumes the distribution of $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$ does not depend on the value of $\boldsymbol{X}$. Meanwhile, Condition 1(b) excludes the indirect effect of $\boldsymbol{X}$ on $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$ via $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}$, by assuming $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}$ and $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$ are uncorrelated given $\boldsymbol{X}$. Therefore, $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$ only contains immaterial information, and $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}$ carries all the material information and possibly some extra immaterial information for the regression. Instead of using any satisfied $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$, the definition of the response envelope model ensures the uniqueness of $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$ by taking it to be the smallest one, i.e. the intersection of all satisfied subspaces. At the same time, by using $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$ rather than any $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$, immaterial information is guaranteed to be not contained in $\mathbb{P}_{\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}$ anymore, and could be removed to the largest degree. So we call $\mathbb{P}_{\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{Y}$ the *material part* and $\mathbb{P}_{\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{Y}$ the *immaterial part* of $\boldsymbol{Y}$ respectively. Although intuitive, it is not easy to formulate the response envelope model under Condition 1. [15] derived the following Condition 2, which is equivalent to Condition 1. The response envelope model could be equivalently defined by Condition 2 with subspace $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$, for (1.1).

**Condition 2** ([15], Definition 2.1). *Assume in* (1.1), *the subspace* $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}} \subset \mathbb{R}^r$ *satisfies*

$(a')$ $\mathrm{span}(\boldsymbol{\beta}^T) \subseteq \mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$,
$(b')$ $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} = \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}} + \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}} \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}}$, *i.e.* $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$ *is a reducing subspace of* $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$.

Condition 2 assumes that $\mathcal{S}_{\boldsymbol{Y}|\boldsymbol{X}}$ is a reducing subspace of $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$ that contains $\mathrm{span}(\boldsymbol{\beta}^T)$. Recalling Definition 2,

$\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$ is exactly $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}}(\mathrm{span}(\boldsymbol{\beta}^T))$, i.e. the $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$-envelope of $\mathrm{span}(\boldsymbol{\beta}^T)$. Therefore, the response envelope model could be defined directly by $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}}(\mathrm{span}(\boldsymbol{\beta}^T))$ for (1.1).

Suppose that the orthonormal basis matrices of $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$ and $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}^{\perp}$ are $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ respectively. Then according to Condition 2 with subspace $\mathcal{E}_{\boldsymbol{Y}|\boldsymbol{X}}$, the coordinate form of the response envelope model is

$$Y = \boldsymbol{\mu}_{\boldsymbol{Y}} + \boldsymbol{\Gamma}\boldsymbol{\eta}X + \boldsymbol{\epsilon}_{\boldsymbol{Y}|\boldsymbol{X}}, \tag{2.1}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} = \boldsymbol{\Gamma}\boldsymbol{\Delta}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Delta}_0\boldsymbol{\Gamma}_0^T, \tag{2.2}$$

where $\boldsymbol{\eta}$, $\boldsymbol{\Delta} = \boldsymbol{\Gamma}^T\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\boldsymbol{\Gamma}$ and $\boldsymbol{\Delta}_0 = \boldsymbol{\Gamma}_0^T\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\boldsymbol{\Gamma}_0$ carry the coordinates of $\boldsymbol{\beta}^T$ relative to $\boldsymbol{\Gamma}$, $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$ relative to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$ respectively. The response envelope model could provide significant efficiency gains when $\|\boldsymbol{\Delta}\| \ll \|\boldsymbol{\Delta}_0\|$, since there will be a large amount of immaterial information for removal under this scenario.

## 2.3 The Predictor Envelope Model

In this section, we review the predictor envelope model [12] by showing how the idea of the response envelope could be adapted to reduce the predictor space. The predictor envelope model has the potential to improve the prediction of $\boldsymbol{Y}$ in (1.1).

To define the predictor envelope, we assume $\boldsymbol{X}$ to be stochastic following $\mathcal{N}(0, \boldsymbol{\Sigma}_{\boldsymbol{X}})$. Then (1.1) is called the predictor envelope model if the envelope structure $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\mathrm{span}(\boldsymbol{\beta}))$ is imposed. Unlike the response envelope, $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\mathrm{span}(\boldsymbol{\beta}))$ is the intersection of all reducing subspaces of $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ that contain $\mathrm{span}(\boldsymbol{\beta})$. Similarly, the material and immaterial parts of $\boldsymbol{X}$ could be defined. The immaterial part of $\boldsymbol{X}$ is assumed not to carry the information for the regression. Suppose the orthonormal basis matrices of $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{X}}}(\mathrm{span}(\boldsymbol{\beta}))$ and its orthogonal complement subspace are $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Upsilon}_0$, then the coordinate form of the predictor envelope model is

$$Y = \boldsymbol{\mu}_{\boldsymbol{Y}} + \boldsymbol{\psi}^T\boldsymbol{\Upsilon}^T X + \boldsymbol{\epsilon}_{\boldsymbol{Y}|\boldsymbol{X}}, \tag{2.3}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{X}} = \boldsymbol{\Upsilon}\boldsymbol{\Xi}\boldsymbol{\Upsilon}^T + \boldsymbol{\Upsilon}_0\boldsymbol{\Xi}_0\boldsymbol{\Upsilon}_0^T, \tag{2.4}$$

where $\boldsymbol{\psi}$, $\boldsymbol{\Xi} = \boldsymbol{\Upsilon}^T\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Upsilon}$ and $\boldsymbol{\Xi}_0 = \boldsymbol{\Upsilon}_0^T\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{\Upsilon}_0$ carry the coordinates of $\boldsymbol{\beta}$ relative to $\boldsymbol{\Upsilon}$, $\boldsymbol{\Sigma}_{\boldsymbol{X}}$ relative to $\boldsymbol{\Upsilon}$ and $\boldsymbol{\Upsilon}_0$ respectively. The large efficiency gains could be obtained when $\|\boldsymbol{\Xi}\| \gg \|\boldsymbol{\Xi}_0\|$.

# 3. PROPOSED SIMULTANEOUS PARTIAL ENVELOPE MODEL

As mentioned in Section 1, suppose that our predictors $\boldsymbol{X} \in \mathbb{R}^p$ could be partitioned into three parts

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}_{1C} \\ \boldsymbol{X}_{1D} \\ \boldsymbol{X}_2 \end{pmatrix},$$

where $\boldsymbol{X}_1 = (\boldsymbol{X}_{1C}^T, \boldsymbol{X}_{1D}^T)^T \in \mathbb{R}^{p_1}$ is the predictors of main interest to us in the multivariate linear regression, with its continuous and discrete parts being $\boldsymbol{X}_{1C} \in \mathbb{R}^{p_C}$ and $\boldsymbol{X}_{1D} \in \mathbb{R}^{p_D}$ respectively ($p_1 = p_C + p_D$), while $\boldsymbol{X}_2 \in \mathbb{R}^{p_2}$ denotes the set of predictors that is not of main interest ($p = p_1 + p_2$). By separating $\boldsymbol{X}_{1D}$ with $\boldsymbol{X}_{1C}$, we could avoid the Normality restrictions for $\boldsymbol{X}_{1D}$ and hence whole $\boldsymbol{X}_1$, while still providing a valid simultaneous envelope estimator for the coefficients of $\boldsymbol{X}_{1C}$. Assuming predictors are not centered, the standard multivariate linear model can be written as

$$\boldsymbol{Y} = \boldsymbol{\mu_Y} + \boldsymbol{\beta}_{1C}^T(\boldsymbol{X}_{1C} - \boldsymbol{\mu}_{1C}) + \boldsymbol{\beta}_{1D}^T(\boldsymbol{X}_{1D} - \boldsymbol{\mu}_{1D}) +$$
$$\boldsymbol{\beta}_2^T(\boldsymbol{X}_2 - \boldsymbol{\mu}_2) + \boldsymbol{\epsilon_{Y|X}}, \qquad (3.1)$$

where $\boldsymbol{\mu_Y} \in \mathbb{R}^r$, $\boldsymbol{\mu}_{1C} \in \mathbb{R}^{p_C}$, $\boldsymbol{\mu}_{1D} \in \mathbb{R}^{p_D}$ and $\boldsymbol{\mu}_2 \in \mathbb{R}^{p_2}$ denote the unknown means of $\boldsymbol{Y}$, $\boldsymbol{X}_{1C}$, $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$ respectively, $\boldsymbol{\beta}_{1C} \in \mathbb{R}^{p_C \times r}$, $\boldsymbol{\beta}_{1D} \in \mathbb{R}^{p_D \times r}$ and $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2 \times r}$ denote the unknown regression coefficients of $\boldsymbol{X}_{1C}$, $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$, and $\boldsymbol{\epsilon_{Y|X}}$ is the random error vector. Specifically, combining $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$, $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_{1C}^T, \boldsymbol{\beta}_{1D}^T)^T$ denotes the coefficients of main interest to us. Assume $\boldsymbol{X}_{1C}$ to be stochastic, and $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$ to be non-stochastic, with sample means $\overline{\boldsymbol{X}}_{1D}$ and $\overline{\boldsymbol{X}}_2$. Replacing $\boldsymbol{\mu}_{1D}$ and $\boldsymbol{\mu}_2$ with $\overline{\boldsymbol{X}}_{1D}$ and $\overline{\boldsymbol{X}}_2$ respectively, (3.1) is equal to (3.2)

$$\boldsymbol{Y} = \boldsymbol{\mu_Y} + \boldsymbol{\beta}_{1C}^T(\boldsymbol{X}_{1C} - \boldsymbol{\mu}_{1C}) + \boldsymbol{\beta}_{1D}^T(\boldsymbol{X}_{1D} - \overline{\boldsymbol{X}}_{1D}) +$$
$$\boldsymbol{\beta}_2^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) + \boldsymbol{\epsilon_{Y|X}}. \qquad (3.2)$$

**Remark.** *For the purpose of expositional simplicity, we implicitly assume the discrete predictors in $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$ to be quantitative variables in (3.1), (3.2) and later in (3.7). However, it is worthy to note that categorical predictors, either nominal or ordinal, are still applicable to our model by including their dummy variables into $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$ instead. In either case, the centering for these two predictors serves to ensure that $\boldsymbol{\mu_Y}$ could be interpreted as the expectation (conditional on model parameters) of the sample means of $\boldsymbol{Y}$.*

We firstly impose some distributional assumptions on (3.2). Assume $\boldsymbol{\epsilon_{Y|X}} \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{\Sigma_{Y|X}})$. Given $\boldsymbol{X}_{1D}$, assume that $\boldsymbol{X}_{1C} = \boldsymbol{\mu}_{1C} + \boldsymbol{\gamma}^T(\boldsymbol{X}_{1D} - \overline{\boldsymbol{X}}_{1D}) + \boldsymbol{\epsilon}_{C|D}$ for an unknown $\boldsymbol{\gamma} \in \mathbb{R}^{p_D \times p_C}$, where $\boldsymbol{\epsilon}_{C|D} \sim \mathcal{N}_{p_C}(\boldsymbol{0}, \boldsymbol{\Sigma}_{C|D})$ and is independent of $\boldsymbol{\epsilon_{Y|X}}$. Rigorously speaking, $\boldsymbol{\mu}_{1C}$ in (3.2) should only be interpreted as the expectation of $\overline{\boldsymbol{X}}_{1C}$ (rather than the expectation of $\boldsymbol{X}_{1C}$) conditional on model parameters under this assumption, due to the non-stochasticity of $\overline{\boldsymbol{X}}_{1D}$. However, we will keep using $\boldsymbol{\mu}_{1C}$ to avoid abusing notations.

To provide the efficiency gains on estimating $\boldsymbol{\beta}_1$ and predicting $\boldsymbol{Y}$, we combine the advantages of the response envelope on the estimation and the predictor envelope on the prediction, by further imposing the following two partial envelope structures on (3.2) (by Definition 2) simultaneously:

(*i*) Partial predictor envelope: $\mathcal{E}_{\boldsymbol{\Sigma}_{C|D}}(\mathcal{L})$, i.e. the $\boldsymbol{\Sigma}_{C|D}$-envelope of $\mathcal{L}$ and is shortened to $\mathcal{E}_{C|D}$, with dimension $d_{\boldsymbol{X}}$ ($d \le d_{\boldsymbol{X}} \le p_C$),

(*ii*) Partial response envelope: $\mathcal{E}_{\boldsymbol{\Sigma_{Y|X}}}(\mathcal{R})$, i.e. the $\boldsymbol{\Sigma_{Y|X}}$-envelope of $\mathcal{R}$ and is shortened to $\mathcal{E}_{\boldsymbol{Y}|1}$, with dimension $d_{\boldsymbol{Y}}$ ($d_1 \le d_{\boldsymbol{Y}} \le r$),

where $\mathcal{L}$ and $\mathcal{R}$ denote $\text{span}(\boldsymbol{\beta}_{1C})$ and $\text{span}(\boldsymbol{\beta}_1^T)$, and $d$ and $d_1$ denote $\text{rank}(\boldsymbol{\beta}_{1C})$ and $\text{rank}(\boldsymbol{\beta}_1)$ respectively. (3.2) is called the **Sim**ultaneous **P**artial Envelope Model (SIMP), if these distributional assumptions and envelope structures are imposed. Note that SIMP avoids the Normality requirement for the whole $\boldsymbol{X}_1$ (i.e., no other assumptions on $\boldsymbol{X}_{1D}$ besides the non-stochasticity, and $\boldsymbol{X}_{1C}$ is Normal with means depending on $\boldsymbol{X}_{1D}$, which relaxes the identical Normal distribution assumption of each observation of predictors in [14]) by imposing $\mathcal{E}_{\boldsymbol{Y}|1}$ on the row space of whole $\boldsymbol{\beta}_1$ while $\mathcal{E}_{C|D}$ on the column space of only $\boldsymbol{\beta}_{1C}$.

It is worthy to note that SIMP is a broad class of envelopes that degenerates to several popular envelope models under certain conditions as listed in Table 1. Below, we introduce Conditions 3 and 4 in Section 3.1, which could equivalently define SIMP and will offer more intuitions on where the efficiency gains come from by assuming these two envelope components. For the convenience of developing the Bayesian inference procedure in Section 4, we give the coordinate form of SIMP in (3.7)–(3.10) in Section 3.3. For the preparation of Section 3.3, Section 3.2 will introduce a re-parameterization trick, to avoid the direct Bayesian inference on matrix parameters living in the Stiefel manifold.

Table 1. *Relationship between* SIMP (3.7)–(3.10) *and several other envelope models. The meanings of $r$, $p_C$, $p_D$, $p_1$ and $p_2$ are introduced in (3.1), and $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$ are dimensions of $\mathcal{E}_{C|D}$ and $\mathcal{E}_{\boldsymbol{Y}|1}$ respectively. The effective number of parameters for each model is also listed. Note for the response and the partial response envelopes, $\boldsymbol{X}_{1D}$ is required to be non-stochastic only and not necessary to be discrete.*

| | Conditions for (3.7)–(3.10) | Effective number of parameters |
|---|---|---|
| SIMP | N.A. | $d_{\boldsymbol{Y}}(d_{\boldsymbol{X}} + p_D) + r(r + 2p_2 + 3)/2 + p_C(p_C + 3)/2$ |
| response envelope [15] | $p_C = p_2 = 0$ | $d_{\boldsymbol{Y}}p + r(r + 3)/2$ |
| predictor envelope [12] | $p_D = p_2 = 0$ and $d_{\boldsymbol{Y}} = r$ | $r(r + 2d_{\boldsymbol{X}} + 3)/2 + p(p + 3)/2$ |
| simultaneous envelope [14] | $p_D = p_2 = 0$ | $d_{\boldsymbol{Y}}d_{\boldsymbol{X}} + r(r + 3)/2 + p(p + 3)/2$ |
| partial response envelope [57] | $p_C = 0$ | $d_{\boldsymbol{Y}}p_1 + r(r + 2p_2 + 3)/2$ |
| partial predictor envelope [45] | $p_2 = 0$ and $d_{\boldsymbol{Y}} = r$ | $r(r + 2d_{\boldsymbol{X}} + 2p_D + 3)/2 + p_C(p_C + 3)/2$ |

## 3.1 Equivalent Conditions for SIMP

In this section, we introduce Conditions 3 and 4, which can equivalently define SIMP for (3.2). These two conditions will not be used in the formulation of SIMP in Section 3.3, but will offer some intuitions on the source of efficiency gains. The equivalence between the conditions in this section and the envelope structures that we have given for SIMP (by Definition 2) shares the same reason with the equivalence between Conditions 1 and 2 for the response envelope model in Section 2.2.

**Condition 3.** *Assume in* (3.2), *the subspace* $\mathcal{S}_{C|D} \subset \mathbb{R}^{p_C}$ *satisfies,*

(i) $\mathrm{Cov}((\boldsymbol{Y}, \mathbb{P}_{\mathcal{S}_{C|D}^\perp} \boldsymbol{X}_{1C}) \mid (\mathbb{P}_{\mathcal{S}_{C|D}} \boldsymbol{X}_{1C}, \boldsymbol{X}_{1D}, \boldsymbol{X}_2)) = \boldsymbol{0}$,

(ii) $\mathrm{Cov}((\mathbb{P}_{\mathcal{S}_{C|D}} \boldsymbol{X}_{1C}, \mathbb{P}_{\mathcal{S}_{C|D}^\perp} \boldsymbol{X}_{1C}) \mid \boldsymbol{X}_{1D}) = \boldsymbol{0}$.

**Condition 4.** *Assume in* (3.2), *the subspace* $\mathcal{S}_{\boldsymbol{Y}|1} \subset \mathbb{R}^r$ *satisfies,*

(I) $\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|1}^\perp} \boldsymbol{Y} \mid (\boldsymbol{X}_1, \boldsymbol{X}_2) \overset{\mathcal{D}}{=} \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|1}^\perp} \boldsymbol{Y} \mid \boldsymbol{X}_2$,

(II) $\mathrm{Cov}((\mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|1}} \boldsymbol{Y}, \mathbb{P}_{\mathcal{S}_{\boldsymbol{Y}|1}^\perp} \boldsymbol{Y}) \mid (\boldsymbol{X}_1, \boldsymbol{X}_2)) = \boldsymbol{0}$.

Condition 3 excludes the direct and indirect partial effects (via $\mathbb{P}_{\mathcal{S}_{C|D}} \boldsymbol{X}_{1C}$) of $\mathbb{P}_{\mathcal{S}_{C|D}^\perp} \boldsymbol{X}_{1C}$ on $\boldsymbol{Y}$. The intersection of all subspaces that satisfy Condition 3 gives the same partial predictor envelope space $\mathcal{E}_{C|D}$ as defined previously by $\mathcal{E}_{\boldsymbol{\Sigma}_{C|D}}(\mathcal{L})$. We call $\mathbb{P}_{\mathcal{E}_{C|D}} \boldsymbol{X}_{1C}$ and $\mathbb{P}_{\mathcal{E}_{C|D}^\perp} \boldsymbol{X}_{1C}$ the *material part* and *immaterial part* of $\boldsymbol{X}_{1C}$. Given $\boldsymbol{X}_{1D}$ and $\boldsymbol{X}_2$, the immaterial part of $\boldsymbol{X}_{1C}$ is assumed not to affect $\boldsymbol{Y}$ by Condition 3. Similarly, the partial response envelope space $\mathcal{E}_{\boldsymbol{Y}|1}$ previously defined by $\mathcal{E}_{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}}(\mathcal{R})$ is also the intersection of all subspaces that satisfy Condition 4. We call $\mathbb{P}_{\mathcal{E}_{\boldsymbol{Y}|1}} \boldsymbol{Y}$ and $\mathbb{P}_{\mathcal{E}_{\boldsymbol{Y}|1}^\perp} \boldsymbol{Y}$ the *material part* and *immaterial part* of $\boldsymbol{Y}$. Given $\boldsymbol{X}_2$, the immaterial part of $\boldsymbol{Y}$ is assumed not to be affected by $\boldsymbol{X}_1$.

## 3.2 Re-Parameterization of the Basis Matrices

We only consider the case of $0 < d_{\boldsymbol{X}} < p_C$ and $0 < d_{\boldsymbol{Y}} < r$ in this section. When $d_{\boldsymbol{X}} \in \{0, p_C\}$ or $d_{\boldsymbol{Y}} \in \{0, r\}$, we can simply take the orthonormal bases of envelope space $\mathcal{E}_{C|D}$ or $\mathcal{E}_{\boldsymbol{Y}|1}$ and the orthogonal complement subspace to be either null or the identity matrix. When $0 < d_{\boldsymbol{X}} < p_C$ and $0 < d_{\boldsymbol{Y}} < r$, let $\mathbf{L} \in \mathbb{R}^{p_C \times d_{\mathbf{X}}}$, $\mathbf{L}_0 \in \mathbb{R}^{p_C \times (p_C - d_{\mathbf{X}})}$, $\mathbf{R} \in \mathbb{R}^{r \times d_{\boldsymbol{Y}}}$ and $\mathbf{R}_0 \in \mathbb{R}^{r \times (r - d_{\boldsymbol{Y}})}$ denote the orthonormal bases of $\mathcal{E}_{C|D}$, $\mathcal{E}_{C|D}^\perp$, $\mathcal{E}_{\boldsymbol{Y}|1}$ and $\mathcal{E}_{\boldsymbol{Y}|1}^\perp$, respectively.

To avoid the difficulty of direct Bayesian inference on $\mathbf{L}$, $\mathbf{L}_0$, $\mathbf{R}$ and $\mathbf{R}_0$, which all live in the Stiefel manifold, re-parameterization of $\mathbf{L}$, $\mathbf{L}_0$, $\mathbf{R}$ and $\mathbf{R}_0$ is considered in the coordinate form of SIMP in Section 3.3. We illustrate the re-parameterization of $\mathbf{L}$ and $\mathbf{L}_0$ here, and this idea can be similarly exploited to re-parameterize $\mathbf{R}$ and $\mathbf{R}_0$. Let $\mathbf{L}_1$ be the matrix formed by the upper $d_{\boldsymbol{X}}$ rows of $\mathbf{L}$, and $\mathbf{L}_2$ be the matrix that contains the remaining rows. Without loss of generality, we assume that $\mathbf{L}_1$ is nonsingular. Otherwise,

we could reorder the rows of $\mathbf{L}$ to make $\mathbf{L}_1$ nonsingular. Then,

$$\mathbf{L} = \begin{pmatrix} \mathbf{L}_1 \\ \mathbf{L}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{I}_{d_{\boldsymbol{X}}} \\ \mathbf{A} \end{pmatrix} \mathbf{L}_1 := \mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A}) \mathbf{L}_1, \qquad (3.3)$$

where $\mathbf{A} = \mathbf{L}_2 \mathbf{L}_1^{-1}$ and $\mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A}) = \begin{pmatrix} \mathbf{I}_{d_{\mathbf{X}}} \\ \mathbf{A} \end{pmatrix}$. [56] shows that $\mathbf{A}$ depends on $\mathbf{L}$ only through $\mathrm{span}(\mathbf{L})$ and there is a one-to-one correspondence between $\mathcal{E}_{C|D}$ and $\mathbf{A}$. [8] further shows that if $\mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A})$ is a basis of $\mathcal{E}_{C|D}$, then

$$\mathbf{D}_{p_C - d_{\mathbf{X}}}(\mathbf{A}) = \begin{pmatrix} -\mathbf{A}^T \\ \mathbf{I}_{p_C - d_{\mathbf{X}}} \end{pmatrix} \qquad (3.4)$$

is a basis of $\mathcal{E}_{C|D}^\perp$. Therefore, after normalization,

$$\mathbf{L}(\mathbf{A}) = \mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A}) \big( \mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A})^T \mathbf{C}_{d_{\mathbf{X}}}(\mathbf{A}) \big)^{-1/2}$$

and

$$\mathbf{L}_0(\mathbf{A}) = \mathbf{D}_{p_C - d_{\mathbf{X}}}(\mathbf{A}) \big( \mathbf{D}_{p_C - d_{\mathbf{X}}}(\mathbf{A})^T \mathbf{D}_{p_C - d_{\mathbf{X}}}(\mathbf{A}) \big)^{-1/2}$$

are a pair of orthonormal bases of $\mathcal{E}_{C|D}$ and $\mathcal{E}_{C|D}^\perp$ respectively. The re-parameterization

$$\mathbf{R}(\mathbf{B}) = \mathbf{C}_{d_{\mathbf{Y}}}(\mathbf{B}) \big( \mathbf{C}_{d_{\mathbf{Y}}}(\mathbf{B})^T \mathbf{C}_{d_{\mathbf{Y}}}(\mathbf{B}) \big)^{-1/2}$$

and

$$\mathbf{R}_0(\mathbf{B}) = \mathbf{D}_{r - d_{\mathbf{Y}}}(\mathbf{B}) \big( \mathbf{D}_{r - d_{\mathbf{Y}}}(\mathbf{B})^T \mathbf{D}_{r - d_{\mathbf{Y}}}(\mathbf{B}) \big)^{-1/2}$$

by an unconstrained matrix $\mathbf{B} \in \mathbb{R}^{(r - d_{\mathbf{Y}}) \times d_{\mathbf{Y}}}$ can be similarly constructed.

## 3.3 Coordinate Form of SIMP

Like (2.1)–(2.2) for the response envelope model and (2.3)–(2.4) for the predictor envelope model, we intend to give the coordinate form of SIMP explicitly as in (3.7)–(3.10), i.e., formulate two envelope structures $\mathcal{E}_{C|D}$ and $\mathcal{E}_{\boldsymbol{Y}|1}$ by analytical expressions for $\boldsymbol{\beta}_{1C}$, $\boldsymbol{\beta}_{1D}$, $\boldsymbol{\Sigma}_{C|D}$ and $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$, and correspondingly rewrite (3.2), for the preparation of the Bayesian inference in Section 4.

Let $\boldsymbol{\beta}_{1C} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ be the singular value decomposition, where $\mathbf{U} \in \mathbb{R}^{p_C \times d}$ and $\mathbf{V} \in \mathbb{R}^{r \times d}$ are semi-orthogonal matrices, and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$, with $\lambda_i \geq 0$, $i = 1, \ldots, d$, being $d$ singular values of $\boldsymbol{\beta}_{1C}$ where $d = \mathrm{rank}(\boldsymbol{\beta}_{1C})$. Since $\mathcal{L} = \mathrm{span}(\boldsymbol{\beta}_{1C}) = \mathrm{span}(\mathbf{U})$ and $\mathcal{R} = \mathrm{span}(\boldsymbol{\beta}_{1C}^T) = \mathrm{span}(\boldsymbol{\beta}_{1C}^T, \boldsymbol{\beta}_{1D}^T) = \mathrm{span}(\mathbf{V}, \boldsymbol{\beta}_{1D}^T)$, according to the definitions of $\mathcal{E}_{C|D}$ and $\mathcal{E}_{\boldsymbol{Y}|1}$ by Definition 2 at the start of Section 3, $\mathbf{U} = \mathbf{L}\mathbf{O}$ and $\mathbf{V} = \mathbf{R}\mathbf{P}$ for some $\mathbf{O} \in \mathbb{R}^{d_{\mathbf{X}} \times d}$ and $\mathbf{P} \in \mathbb{R}^{d_{\mathbf{Y}} \times d}$. Hence the coordinate form of $\boldsymbol{\beta}_{1C}$ is

$$\boldsymbol{\beta}_{1C} = (\mathbf{L}\mathbf{O})\mathbf{D}(\mathbf{R}\mathbf{P})^T := \mathbf{L}\boldsymbol{\eta}_C^T \mathbf{R}^T, \qquad (3.5)$$

where $\boldsymbol{\eta}_C = \mathbf{PDO}^T \in \mathbb{R}^{d_{\boldsymbol{Y}} \times d_{\boldsymbol{X}}}$ is the coordinate of $\boldsymbol{\beta}_{1C}^T$ relative to $\mathbf{R}$ and $\mathbf{L}^T$. Also,

$$\boldsymbol{\beta}_{1D}^T = \mathbf{R}\boldsymbol{\eta}_D, \qquad (3.6)$$

for some $\boldsymbol{\eta}_D \in \mathbb{R}^{d_{\boldsymbol{Y}} \times p_D}$, which carries the coordinate of $\boldsymbol{\beta}_{1D}^T$ relative to $\mathbf{R}$. Then the coordinate form of SIMP is given by

$$\boldsymbol{Y} = \boldsymbol{\mu}_{\boldsymbol{Y}} + \mathbf{R}(\mathbf{B})\boldsymbol{\eta}_C \mathbf{L}(\mathbf{A})^T(\boldsymbol{X}_{1C} - \boldsymbol{\mu}_{1C}) + \mathbf{R}(\mathbf{B})\boldsymbol{\eta}_D$$
$$(\boldsymbol{X}_{1D} - \overline{\boldsymbol{X}}_{1D}) + \boldsymbol{\beta}_2^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) + \boldsymbol{\epsilon}_{\boldsymbol{Y}|\boldsymbol{X}}, \qquad (3.7)$$

$$\boldsymbol{X}_{1C} = \boldsymbol{\mu}_{1C} + \boldsymbol{\gamma}^T(\boldsymbol{X}_{1D} - \overline{\boldsymbol{X}}_{1D}) + \boldsymbol{\epsilon}_{C|D}, \qquad (3.8)$$

$$\boldsymbol{\Sigma}_{C|D} = \mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T, \qquad (3.9)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}} = \mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0\mathbf{R}_0(\mathbf{B})^T, \qquad (3.10)$$

where $\boldsymbol{\beta}_{1C} = \mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T$ and $\boldsymbol{\beta}_{1D} = \boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T$, as illustrated in (3.5) and (3.6). This implies that the rows and columns of $\boldsymbol{\beta}_{1C}$ depend on the partial predictor and partial response envelopes only, and the columns of $\boldsymbol{\beta}_{1D}$ depend on the partial response envelope only. Note that $\boldsymbol{\Omega} = \mathbf{L}(\mathbf{A})^T\boldsymbol{\Sigma}_{C|D}\mathbf{L}(\mathbf{A})$, $\boldsymbol{\Omega}_0 = \mathbf{L}_0(\mathbf{A})^T\boldsymbol{\Sigma}_{C|D}\mathbf{L}_0(\mathbf{A})$, $\boldsymbol{\Phi} = \mathbf{R}(\mathbf{B})^T\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\mathbf{R}(\mathbf{B})$ and $\boldsymbol{\Phi}_0 = \mathbf{R}_0(\mathbf{B})^T\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\mathbf{R}_0(\mathbf{B})$ carry the coordinates of $\boldsymbol{\Sigma}_{C|D}$ relative to $\mathbf{L}(\mathbf{A})$ and $\mathbf{L}_0(\mathbf{A})$, and $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$ relative to $\mathbf{R}(\mathbf{B})$ and $\mathbf{R}_0(\mathbf{B})$ respectively. Multiply $\mathbf{R}(\mathbf{B})^T$ or $\mathbf{R}_0(\mathbf{B})^T$ on both sides of (3.7),

$$\begin{cases} \mathbf{R}(\mathbf{B})^T\boldsymbol{Y} = \mathbf{R}(\mathbf{B})^T\boldsymbol{\mu}_{\boldsymbol{Y}} + \boldsymbol{\eta}_C(\mathbf{L}(\mathbf{A})^T(\boldsymbol{X}_{1C} - \boldsymbol{\mu}_{1C})) + \\ \qquad \boldsymbol{\eta}_D(\boldsymbol{X}_{1D} - \overline{\boldsymbol{X}}_{1D}) + \mathbf{R}(\mathbf{B})^T\boldsymbol{\beta}_2^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) + \\ \qquad \mathbf{R}(\mathbf{B})^T\boldsymbol{\epsilon}_{\boldsymbol{Y}|\boldsymbol{X}}, \\ \mathbf{R}_0(\mathbf{B})^T\boldsymbol{Y} = \mathbf{R}_0(\mathbf{B})^T\boldsymbol{\mu}_{\boldsymbol{Y}} + \mathbf{R}_0(\mathbf{B})^T\boldsymbol{\beta}_2^T(\boldsymbol{X}_2 - \overline{\boldsymbol{X}}_2) + \\ \qquad \mathbf{R}_0(\mathbf{B})^T\boldsymbol{\epsilon}_{\boldsymbol{Y}|\boldsymbol{X}}. \end{cases}$$

This pair of equations shows that $\boldsymbol{X}_2$ affects both material and immaterial parts of $\boldsymbol{Y}$, $\boldsymbol{X}_{1D}$ affects only the material part of $\boldsymbol{Y}$, and only the material part of $\boldsymbol{X}_{1C}$ affects the material part of $\boldsymbol{Y}$ only. The estimation of $\boldsymbol{\beta}_{1C}$ is benefitted from removing the redundant variations of both $\mathbf{R}_0(\mathbf{B})^T\boldsymbol{Y}$ and $\mathbf{L}_0(\mathbf{A})^T\boldsymbol{X}_{1C}$, whereas the efficiency gains for estimating $\boldsymbol{\beta}_{1D}$ only come from removing the redundant variation of $\mathbf{R}_0(\mathbf{B})^T\boldsymbol{Y}$.

Inherited from the (partial) predictor and the (partial) response envelope models, the partial predictor envelope component of SIMP offers large efficiency gains when $\|\boldsymbol{\Omega}\| \gg \|\boldsymbol{\Omega}_0\|$, and significant advantages of the partial response envelope component require $\|\boldsymbol{\Phi}\| \ll \|\boldsymbol{\Phi}_0\|$.

## 4. BAYESIAN INFERENCE

In this section, we develop a Bayesian procedure for the statistical inference of SIMP, hence our method is called the Bayesian simultaneous partial envelope model. It is implemented through sampling from the posterior distribution of parameters. The determination of the posterior distribution requires both the likelihood function and prior distributions.

Suppose that we have $n$ independent observations $\mathcal{D} = \{\boldsymbol{Y}_i, \boldsymbol{X}_{1C,i}, \boldsymbol{X}_{1D,i}, \boldsymbol{X}_{2,i}\}_{i=1}^n$ from SIMP. Let $\mathbb{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)^T \in \mathbb{R}^{n \times r}$, $\mathbb{X}_{1C} = (\boldsymbol{X}_{1C,1}, \ldots, \boldsymbol{X}_{1C,n})^T \in \mathbb{R}^{n \times p_C}$, $\mathbb{X}_{1D} = (\boldsymbol{X}_{1D,1}, \ldots, \boldsymbol{X}_{1D,n})^T \in \mathbb{R}^{n \times p_D}$ and $\mathbb{X}_2 = (\boldsymbol{X}_{2,1}, \ldots, \boldsymbol{X}_{2,n})^T \in \mathbb{R}^{n \times p_2}$ be data matrices. For notational simplicity, we consider standardized data matrices $\widetilde{\mathbb{Y}} = \mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{Y}}^T$, $\widetilde{\mathbb{X}}_{1C} = \mathbb{X}_{1C} - \mathbf{1}_n\boldsymbol{\mu}_{1C}^T$, $\widetilde{\mathbb{X}}_{1D} = \mathbb{X}_{1D} - \mathbf{1}_n\overline{\boldsymbol{X}}_{1D}^T$ and $\widetilde{\mathbb{X}}_2 = \mathbb{X}_2 - \mathbf{1}_n\overline{\boldsymbol{X}}_2^T$ for $\mathbb{Y}$, $\mathbb{X}_{1C}$, $\mathbb{X}_{1D}$ and $\mathbb{X}_2$, respectively. Then, for fixed envelope dimensions $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$, the log-likelihood function for SIMP (3.7)–(3.10) is given by

$$l(\boldsymbol{\Theta}) = const - \frac{n}{2}\log(|\boldsymbol{\Omega}|) - \frac{n}{2}\log(|\boldsymbol{\Omega}_0|) - \frac{1}{2}\mathrm{tr}\big\{(\widetilde{\mathbb{X}}_{1C} -$$
$$\widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T\big)^{-1}$$
$$(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T\big\} - \frac{n}{2}\log(|\boldsymbol{\Phi}|) - \frac{n}{2}\log(|\boldsymbol{\Phi}_0|) -$$
$$\frac{1}{2}\mathrm{tr}\big\{\big(\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T -$$
$$\widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0\mathbf{R}_0(\mathbf{B})^T\big)^{-1}$$
$$\big(\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T -$$
$$\widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)^T\big\}, \qquad (4.1)$$

where *const* denotes a constant which does not depend on parameters of the model, $\boldsymbol{\Theta} = \{\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D, \mathbf{A}, \mathbf{B}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0\}$.

To perform Bayesian inference on $\boldsymbol{\Theta}$ related to SIMP, we put the following prior distributions on $\boldsymbol{\Theta}$:

$$\pi(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_{\boldsymbol{Y}}) \propto 1,$$
$$\boldsymbol{\beta}_2 \mid \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{B} \sim \mathcal{MN}_{p_2,r}\big(\mathbf{M}^{-1}\mathbf{Z}, \mathbf{M}^{-1}, \boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}\big),$$
$$\boldsymbol{\gamma} \mid \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \mathbf{A} \sim \mathcal{MN}_{p_D,p_C}\big(\boldsymbol{\Lambda}^{-1}\mathbf{F}, \boldsymbol{\Lambda}^{-1}, \boldsymbol{\Sigma}_{C|D}\big),$$
$$\boldsymbol{\Omega} \sim \mathcal{IW}_{d_{\boldsymbol{X}}}(\boldsymbol{\Psi}_{\boldsymbol{X}}, w_{\boldsymbol{X}}),$$
$$\boldsymbol{\Omega}_0 \sim \mathcal{IW}_{p_C-d_{\boldsymbol{X}}}(\boldsymbol{\Psi}_{\boldsymbol{X}_0}, w_{\boldsymbol{X}_0}),$$
$$\boldsymbol{\Phi} \sim \mathcal{IW}_{d_{\boldsymbol{Y}}}(\boldsymbol{\Psi}_{\boldsymbol{Y}}, w_{\boldsymbol{Y}}),$$
$$\boldsymbol{\Phi}_0 \sim \mathcal{IW}_{r-d_{\boldsymbol{Y}}}(\boldsymbol{\Psi}_{\boldsymbol{Y}_0}, w_{\boldsymbol{Y}_0}),$$
$$\mathbf{A} \sim \mathcal{MN}_{p_C-d_{\boldsymbol{X}},d_{\boldsymbol{X}}}(\mathbf{A}_0, \mathbf{K}_{\mathbf{A}}, \boldsymbol{\Sigma}_{\mathbf{A}}),$$
$$\mathbf{B} \sim \mathcal{MN}_{r-d_{\boldsymbol{Y}},d_{\boldsymbol{Y}}}(\mathbf{B}_0, \mathbf{K}_{\mathbf{B}}, \boldsymbol{\Sigma}_{\mathbf{B}}),$$
$$\boldsymbol{\eta}_C \mid \boldsymbol{\Phi} \sim \mathcal{MN}_{d_{\boldsymbol{Y}},d_{\boldsymbol{X}}}\big(\mathbf{W}\mathbf{E}^{-1}, \boldsymbol{\Phi}, \mathbf{E}^{-1}\big),$$
$$\boldsymbol{\eta}_D \mid \boldsymbol{\Phi} \sim \mathcal{MN}_{d_{\boldsymbol{Y}},p_D}\big(\mathbf{T}\mathbf{Q}^{-1}, \boldsymbol{\Phi}, \mathbf{Q}^{-1}\big),$$

where $\mathcal{MN}$ and $\mathcal{IW}$ denote Matrix normal and Inverse-Wishart distributions. See Appendix E for the detailed introduction of these two distributions. Here we implicitly assume $\mathbf{A}_0 \in \mathbb{R}^{(p_C-d_{\boldsymbol{X}}) \times d_{\boldsymbol{X}}}$, $\mathbf{B}_0 \in \mathbb{R}^{(r-d_{\boldsymbol{Y}}) \times d_{\boldsymbol{Y}}}$, $\mathbf{Z} \in \mathbb{R}^{p_2 \times r}$, $\mathbf{F} \in \mathbb{R}^{p_D \times p_C}$, $\mathbf{W} \in \mathbb{R}^{d_{\boldsymbol{Y}} \times d_{\boldsymbol{X}}}$, $\mathbf{T} \in \mathbb{R}^{d_{\boldsymbol{Y}} \times p_D}$, $w_{\boldsymbol{X}} > d_{\boldsymbol{X}} - 1$, $w_{\boldsymbol{X}_0} > p_C - d_{\boldsymbol{X}} - 1$, $w_{\boldsymbol{Y}} > d_{\boldsymbol{Y}} - 1$, $w_{\boldsymbol{Y}_0} > r - d_{\boldsymbol{Y}} - 1$ and $\mathbf{M} \in \mathbb{S}_+^{p_2 \times p_2}$, $\boldsymbol{\Lambda}, \mathbf{Q} \in \mathbb{S}_+^{p_D \times p_D}$, $\boldsymbol{\Psi}_{\boldsymbol{X}}, \boldsymbol{\Sigma}_{\mathbf{A}}, \mathbf{E} \in \mathbb{S}_+^{d_{\boldsymbol{X}} \times d_{\boldsymbol{X}}}$, $\boldsymbol{\Psi}_{\boldsymbol{Y}}, \boldsymbol{\Sigma}_{\mathbf{B}} \in \mathbb{S}_+^{d_{\boldsymbol{Y}} \times d_{\boldsymbol{Y}}}$, $\boldsymbol{\Psi}_{\boldsymbol{X}_0}, \mathbf{K}_{\mathbf{A}} \in \mathbb{S}_+^{(p_C-d_{\boldsymbol{X}}) \times (p_C-d_{\boldsymbol{X}})}$, $\boldsymbol{\Psi}_{\boldsymbol{Y}_0}, \mathbf{K}_{\mathbf{B}} \in \mathbb{S}_+^{(r-d_{\boldsymbol{Y}}) \times (r-d_{\boldsymbol{Y}})}$. The prior information for the parameters could be easily incorporated through these hyperparameters. For example, if we know

a priori that the partial predictor envelope $\mathcal{E}_{C|D}$ is likely to be $\widehat{\mathcal{E}}_{C|D}^{prior}$, then we could compute $\widehat{\mathbf{A}}^{prior}$ from the basis of $\widehat{\mathcal{E}}_{C|D}^{prior}$ by (3.3), and set $\mathbf{A}_0$ as $\widehat{\mathbf{A}}^{prior}$. Our confidence about the prior of $\mathbf{A}$ is encoded in $\mathbf{K_A}$ and $\mathbf{\Sigma_A}$. Specifically, the prior covariance matrices for the $i$-th row and the $j$-th column of $\mathbf{A}$ are $\mathbf{K}_{\mathbf{A},ii}\mathbf{\Sigma_A}$ and $\mathbf{\Sigma}_{\mathbf{A},jj}\mathbf{K_A}$ respectively, where $\mathbf{K}_{\mathbf{A},ii}$ and $\mathbf{\Sigma}_{\mathbf{A},jj}$ are the $i$-th and $j$-th diagonal elements of $\mathbf{K_A}$ and $\mathbf{\Sigma_A}$, and the small $\mathbf{K}_{\mathbf{A},ii}$'s or $\mathbf{\Sigma}_{\mathbf{A},jj}$'s can reflect our strong prior belief of the specific row(s) or column(s) of $\widehat{\mathbf{A}}^{prior}$. The Bayesian inference is made via the Metropolis-within-Gibbs algorithm to generate the sample from the posterior distribution of the parameters $\mathbf{\Theta}$. A number of posterior samples at early iterations are discarded as burn-in. Suppose the retained posterior samples are $\mathcal{D}^* = \{\mathbf{\Theta}^{(s)}\}_{s=1}^S := \{\boldsymbol{\mu}_{1C}^{(s)}, \boldsymbol{\mu}_{\mathbf{Y}}^{(s)}, \boldsymbol{\beta}_2^{(s)}, \boldsymbol{\gamma}^{(s)}, \boldsymbol{\eta}_C^{(s)}, \boldsymbol{\eta}_D^{(s)},$ $\mathbf{A}^{(s)}, \mathbf{B}^{(s)}, \mathbf{\Omega}^{(s)}, \mathbf{\Omega}_0^{(s)}, \mathbf{\Phi}^{(s)}, \mathbf{\Phi}_0^{(s)}\}_{s=1}^S$ with sample size $S$. The algorithm is provided in Appendix A, and the warm start initial estimator that we propose for the MCMC algorithm is detailed in Appendix B.

## 5. THEORETICAL PROPERTIES

In this section, we establish two theoretical properties of our method. We first establish the propriety of the joint posterior density of all parameters in Theorem 1, although the priors of $\boldsymbol{\mu}_{1C}$ and $\boldsymbol{\mu}_{\mathbf{Y}}$ are improper. Proofs are left to Appendix D.

**Theorem 1.** *The posterior density of* $(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_{\mathbf{Y}}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\eta}_C,$ $\boldsymbol{\eta}_D, \mathbf{A}, \mathbf{B}, \mathbf{\Omega}, \mathbf{\Omega}_0, \mathbf{\Phi}, \mathbf{\Phi}_0)$ *with respect to the Lebesgue measure on* $\mathbb{R}^{P_C} \times \mathbb{R}^r \times \mathbb{R}^{p_2 \times r} \times \mathbb{R}^{p_D \times p_C} \times \mathbb{R}^{d_{\mathbf{Y}} \times d_{\mathbf{X}}} \times \mathbb{R}^{d_{\mathbf{Y}} \times p_D} \times$ $\mathbb{R}^{(p_C - d_{\mathbf{X}}) \times d_{\mathbf{X}}} \times \mathbb{R}^{(r - d_{\mathbf{Y}}) \times d_{\mathbf{Y}}} \times \mathbb{S}_+^{d_{\mathbf{X}} \times d_{\mathbf{X}}} \times \mathbb{S}_+^{(p_C - d_{\mathbf{X}}) \times (p_C - d_{\mathbf{X}})} \times$ $\mathbb{S}_+^{d_{\mathbf{Y}} \times d_{\mathbf{Y}}} \times \mathbb{S}_+^{(r - d_{\mathbf{Y}}) \times (r - d_{\mathbf{Y}})}$ *is proper.*

Theorem 2 establishes the Harris ergodicity of the Metropolis-within-Gibbs algorithm developed for the Bayesian inference of SIMP. This property ensures the asymptotic convergence of the Markov chain generated by our algorithm to the joint posterior distribution, irrespective of the starting points. This convergence may fail on a set of measure zero without the guarantee of this property. Despite measure zero, some choices for starting points in this pathological null set can arise naturally, as illustrated in [48] for example. Technical details including definitions for probabilistic terminologies and proofs are left to Appendix D.

**Theorem 2.** *Whenever* $0 \le d_{\mathbf{X}} \le p_C$ *and* $0 \le d_{\mathbf{Y}} \le r$, *the Markov chain generated by the Metropolis-within-Gibbs algorithm for the posterior sampling of* $\mathbf{\Theta}$ *in* SIMP, *as detailed in Appendix A, is Harris ergodic, i.e. (a) $\phi$-irreducible with respect to some measure $\phi$, (b) Aperiodic, and (c) Harris recurrent.*

## 6. SELECTION OF ENVELOPE DIMENSIONS

To effectively search the envelope dimensions, we first select the minimum possible dimension of two envelope components, which is the rank of $\boldsymbol{\beta}_{1C}$, and narrow down the range of dimensions to be searched to save the computational cost.

## 6.1 Selecting $d = \mathrm{rank}(\boldsymbol{\beta}_{1C})$

To determine $d$, we adapt the Bura-Cook estimator [4] to our model setting. This estimator includes a sequence of Chi-squared tests. For $k = 0, 1, \ldots, \min(p_C, r) - 1$, the test statistic is $T_k = n\sum_{j=k+1}^{\min(p_C,r)} \varphi_j$, where $\varphi_1 \ge \cdots \ge \varphi_{\min(p_C,r)}$ are eigenvalues of

$$\widehat{\boldsymbol{\beta}}_{1C,std} = \left\{(n - p_C - 1)/n\right\}^{1/2} \mathbf{S}_{1C}^{1/2} \widehat{\boldsymbol{\beta}}_{R_{\mathbf{Y}|1D,2}|1C} \mathbf{S}_{R_{\mathbf{Y}|1D,2}|1C}^{-1/2},$$

where $\mathbf{S}_{1C}$ is the sample covariance matrix of $\mathbf{X}_{1C}$. The residuals of the regression from $\mathbf{Y}$ on $\mathbf{X}_{1D}$ and $\mathbf{X}_2$ are regressed on $\mathbf{X}_{1C}$ again, and the estimated regression coefficients and residual sample covariance matrix are $\widehat{\boldsymbol{\beta}}_{R_{\mathbf{Y}|1D,2}|1C}$ and $\mathbf{S}_{R_{\mathbf{Y}|1D,2}|1C}$ respectively. Hereinafter, if we don't mention specifically for the method of a regression, it always refers to the frequentist ordinary least squares. [4] shows $T_k$ follows the $\chi^2_{(p_C-k)(r-k)}$ distribution under the null hypothesis that $\mathrm{rank}(\boldsymbol{\beta}_{1C})$ is equal to $k$. For each $k = 0, 1, \ldots, \min(p_C, r) - 1$, we compare $T_k$ with the upper $\alpha$ quantile of the $\chi^2_{(p_C-k)(r-k)}$ distribution, where $\alpha$ is the pre-specified significance level, and default to be 0.05 in our numerical studies. We choose $d$ as the first non-significant value of $k$ or $d = \min(p_C, r)$ otherwise. The Bura-Cook estimator exhibits excellent ability in selecting $d$, as illustrated in the third column of Table 2. Details of the simulation set-up and numerical performance are introduced in Section 7.1.

## 6.2 Selecting Envelope Dimensions

Dimensions $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$ should be specified before fitting SIMP. In Section 7.1, we investigate the performance of Bayesian cross-validation (CV) and four information criteria, Akaike information criterion (AIC-MCMC), Bayesian information criterion (BIC-MCMC), Deviance information criterion (DIC) and Watanabe-Akaike information criterion (WAIC) in selecting these two envelope dimensions for SIMP. They are selected for their popularity in the Bayesian literature. Dimension selection is critical in envelope modeling. To the best of our knowledge, we are the first to investigate the performance of these interesting methods together for the dimension selection of the Bayesian envelope model, while previous Bayesian envelope literature simply chooses one of them directly [5, 38]. For each fixed $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$,

$$\mathrm{AIC-MCMC} = -2l(\widehat{\mathbf{\Theta}}_{d_{\mathbf{X}},d_{\mathbf{Y}}}^{max}) + 2K(d_{\mathbf{X}}, d_{\mathbf{Y}}), \quad (6.1)$$

$$\mathrm{BIC-MCMC} = -2l(\widehat{\mathbf{\Theta}}_{d_{\mathbf{X}},d_{\mathbf{Y}}}^{max}) + \log(n)K(d_{\mathbf{X}}, d_{\mathbf{Y}}), \quad (6.2)$$

where $l(\widehat{\mathbf{\Theta}}_{d_{\mathbf{X}},d_{\mathbf{Y}}}^{max})$ is the largest log-likelihood value attained by MCMC samples after burn-in, with envelope dimensions fixed at $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$ for associated parameters, and $K(d_{\mathbf{X}}, d_{\mathbf{Y}}) = d_{\mathbf{Y}}(d_{\mathbf{X}}+p_D)+r(r+2p_2+3)/2+p_C(p_C+3)/2$ is the effective number of free parameters in SIMP. Note that the definitions of AIC-MCMC and BIC-MCMC here are slightly different from the ones that readers might be used to in the frequentist setting, where we replace the maximum likelihood estimates of the parameters by the empirical maximizers in the retained MCMC samples. Definitions of DIC, WAIC and Bayesian CV involve more notations, and hence are left to Appendix F. We choose $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$ as the minimizers of either the minus of the average out-of-sample estimate of the log predictive density for the Bayesian CV (see details in Appendix F), or one of these four information criteria, among each pair of $d_{\mathbf{X}} = \widehat{d}, \ldots, p_C$ and $d_{\mathbf{Y}} = \widehat{d}, \ldots, r$ by the grid search, where $\widehat{d}$ is the estimate of $d$ from the Bura-Cook estimator introduced in Section 6.1.

## 7. SIMULATION STUDY

### 7.1 Performance of the Envelope Dimensions Selection

In this section, we investigate the numerical performance of the dimension selection methods for SIMP, which are introduced in Section 6.2. We fix $r = p_C = 8$, $p_D = p_2 = 2$, and we consider four cases where true dimensions $(d_{\mathbf{X}}, d_{\mathbf{Y}})$ are $(2,2)$, $(6,2)$, $(2,6)$ or $(6,6)$, with their corresponding effective numbers of parameters being 112, 120, 128 and 152. Three sample sizes, 150, 300 and 500, are considered, given the magnitude of the effective numbers of parameters. We generate $\mathbf{\Omega}$, $\mathbf{\Omega}_0$, $\mathbf{\Phi}$ and $\mathbf{\Phi}_0$ all as diagonal matrices, with their associated $L_{1,1}$ norms being 100, 0.5, 1 and 10, and the assignment proportions to each diagonal element are generated from Dirichlet distributions with shape vectors of

all 5, 5, 1 and 1 respectively. Entries of $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}_C$, $\boldsymbol{\eta}_D$ are generated independently from $Unif(-2,2)$, and those of $\mathbf{A}$ and $\mathbf{B}$ are from $Unif(-1,1)$ independently. All elements of $\boldsymbol{\mu}_{1C}$ and $\boldsymbol{\mu}_{\mathbf{Y}}$ are independently from $Unif(0,10)$. Covariates of $\mathbb{X}_{1D}$ are generated independently from discrete $Unif\{0,1,2\}$, and samples for $\mathbb{X}_2$ are independently from $\mathcal{N}_{p_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$ with $\boldsymbol{\mu}_2 = (2,5)^T$, and $\boldsymbol{\Sigma}_2$ as a realization from $\mathcal{IW}_{p_2}(\mathbf{I}_{p_2}, p_2)$.

For our method, we specify the hyperparameters $\mathbf{Z}$, $\mathbf{F}$, $\mathbf{W}$, $\mathbf{T}$, $\mathbf{A}_0$ and $\mathbf{B}_0$ to be zero matrices, $\mathbf{M}$, $\mathbf{\Lambda}$, $\mathbf{E}$ and $\mathbf{Q}$ to be $10^{-6}$ times the identity matrices, $\mathbf{K_A}$, $\mathbf{K_B}$, $\boldsymbol{\Sigma_A}$ and $\boldsymbol{\Sigma_B}$ to be $10^6$ times the identity matrices, $w_{\mathbf{X}}$, $w_{\mathbf{X}_0}$, $w_{\mathbf{Y}}$ and $w_{\mathbf{Y}_0}$ to be the row numbers of $\mathbf{\Omega}$, $\mathbf{\Omega}_0$, $\mathbf{\Phi}$ and $\mathbf{\Phi}_0$ respectively, and $\boldsymbol{\Psi}_{\mathbf{X}}$, $\boldsymbol{\Psi}_{\mathbf{X}_0}$, $\boldsymbol{\Psi}_{\mathbf{Y}}$ and $\boldsymbol{\Psi}_{\mathbf{Y}_0}$ to be identity matrices multiplied by $10^{-6}$ and their respective degrees of freedom (correspondingly, $w_{\mathbf{X}}$, $w_{\mathbf{X}_0}$, $w_{\mathbf{Y}}$ and $w_{\mathbf{Y}_0}$). The MCMC algorithm for our method is run for 20,000 iterations with the first 50% iterations as burn-in. For each case, their performance is averaged over 500 repetitions. Percentages that the Bura-Cook estimator correctly identifies $d$ and each of five methods that correctly identifies true $d_{\mathbf{X}}$ or $d_{\mathbf{Y}}$ individually or true $(d_{\mathbf{X}}, d_{\mathbf{Y}})$ jointly are listed in Table 2.

Results in Table 2 reveal the advantage of BIC-MCMC in selecting $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$ for SIMP, among various settings of true $d_{\mathbf{X}}$ and $d_{\mathbf{Y}}$ and sample sizes, except when $d_{\mathbf{X}} = 6$, $d_{\mathbf{Y}} = 2$ and the sample size is small. It suggests that the excellent performance of BIC-MCMC may need at least moderate sample sizes. AIC-MCMC and WAIC also perform relatively well in the large $d_{\mathbf{X}}$ and large $d_{\mathbf{Y}}$ setting. These observations for BIC-MCMC that needs large sample size to respond, and the better performance of AIC-MCMC under large envelope dimensions are consistent with the previous findings for envelope models ([58]). Compared with underestimation, a little over-estimating of these two envelope dimensions is acceptable, since only some estimation efficiency, rather than the material information, is lost [5]. The

Table 2. Percentages that the Bura-Cook estimator correctly identifies $d$ (Column 3), and each of five methods correctly identifies $d_{\mathbf{X}}$ or $d_{\mathbf{Y}}$ individually or the $(d_{\mathbf{X}}, d_{\mathbf{Y}})$ pair (Columns 4-8) out of 500 repetitions. Among five methods, the best one for selecting $(d_{\mathbf{X}}, d_{\mathbf{Y}})$ pair under each scenario is highlighted in bold face.

| $(d_{\mathbf{X}}, d_{\mathbf{Y}})$ | n | B-C | AIC-MCMC | BIC-MCMC | DIC | WAIC | 5-fold CV |
|---|---|---|---|---|---|---|---|
| | 150 | 0.99 | 0.37/0.65/0.29 | **0.85/0.98/0.84** | 0.00/0.20/0.00 | 0.00/0.24/0.00 | 0.58/0.56/0.26 |
| (2, 2) | 300 | 1.00 | 0.35/0.63/0.25 | **0.92/0.99/0.91** | 0.00/0.19/0.00 | 0.00/0.20/0.00 | 0.54/0.61/0.27 |
| | 500 | 1.00 | 0.44/0.68/0.35 | **0.96/0.99/0.95** | 0.00/0.15/0.00 | 0.00/0.24/0.00 | 0.53/0.63/0.30 |
| | 150 | 0.98 | **0.51/0.79/0.41** | 0.42/0.98/0.40 | 0.07/0.15/0.01 | 0.54/0.30/0.18 | 0.08/0.09/0.01 |
| (6, 2) | 300 | 0.99 | 0.59/0.84/0.49 | **0.66/0.99/0.66** | 0.03/0.16/0.00 | 0.55/0.28/0.15 | 0.08/0.06/0.01 |
| | 500 | 0.99 | 0.61/0.87/0.53 | **0.75/0.99/0.75** | 0.02/0.13/0.00 | 0.53/0.27/0.15 | 0.15/0.05/0.01 |
| | 150 | 0.99 | 0.50/0.79/0.39 | **0.99/0.99/0.98** | 0.00/0.36/0.00 | 0.04/0.71/0.03 | 0.41/0.01/0.01 |
| (2, 6) | 300 | 1.00 | 0.55/0.81/0.45 | **0.99/1.00/0.99** | 0.00/0.31/0.00 | 0.02/0.72/0.01 | 0.40/0.08/0.04 |
| | 500 | 1.00 | 0.63/0.81/0.51 | **1.00/1.00/0.99** | 0.00/0.31/0.00 | 0.01/0.69/0.01 | 0.44/0.16/0.08 |
| | 150 | 0.99 | 0.90/0.92/0.83 | **0.99/0.99/0.99** | 0.12/0.44/0.05 | 0.89/0.91/0.82 | 0.64/0.82/0.52 |
| (6, 6) | 300 | 1.00 | 0.91/0.94/0.86 | **1.00/1.00/1.00** | 0.08/0.51/0.05 | 0.90/0.90/0.81 | 0.61/0.76/0.44 |
| | 500 | 0.99 | 0.89/0.95/0.85 | **0.99/0.99/0.99** | 0.05/0.46/0.01 | 0.87/0.88/0.78 | 0.57/0.71/0.41 |
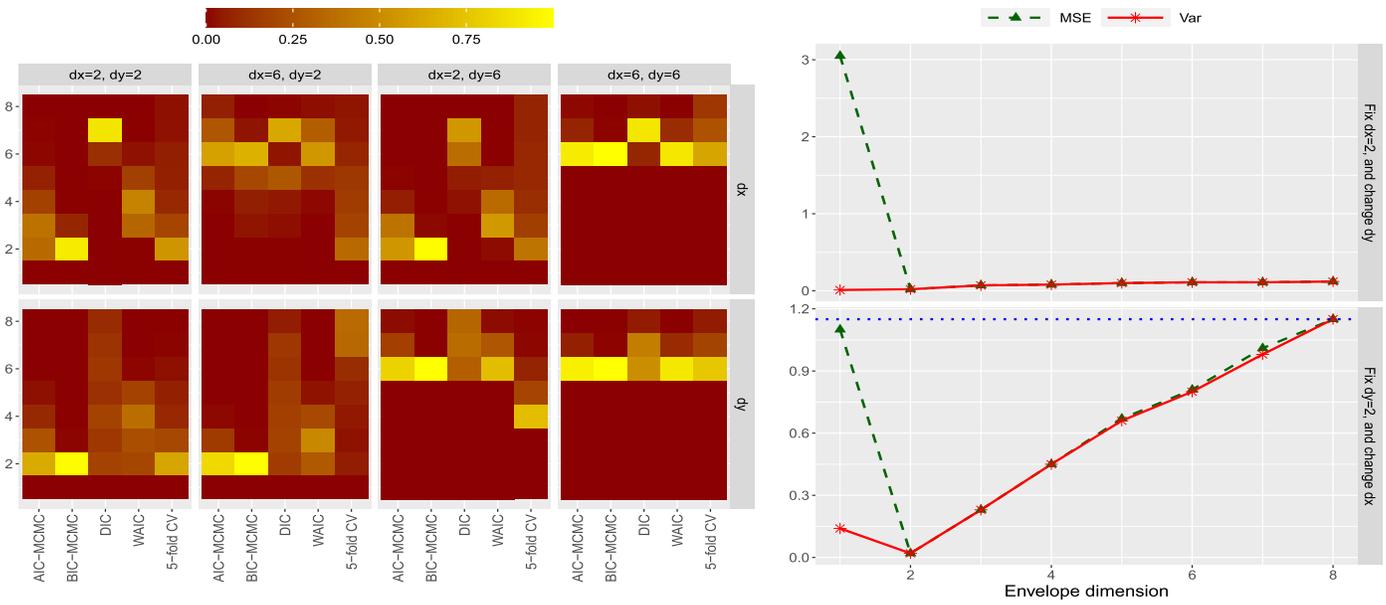
Figure 1: Left: Selection frequencies of $d_X$ and $d_Y$ (Rows 1 and 2) of five criteria, for different settings of true $(d_X, d_Y)$ (Columns 1–4). Colors represent the selection frequency; Right: MSE and estimated variances (Var) of $\widehat{\beta}_1$ from SIMP across all possible choices of $d_Y$ (Row 1) or $d_X$ (Row 2) with $d_X$ or $d_Y$ fixed at 2 (the true value) respectively. The horizontal blue dotted line in the figure that is at the bottom of the right panel shows the MSE of $\widehat{\beta}_1$ from the frequentist partial response envelope with the envelope dimension as 2. Here the results of two panels are based on the sample size as $n = 300$ over 500 repetitions. Results with $n = 150$ and $n = 500$ are similar.

right panel of Figure 1 illustrates this phenomenon, since it is observed that the estimation biases (i.e., the square root of the vertical gaps between the green and the red lines, due to the Bias-variance decomposition) are large for underestimated envelope dimensions, whereas with correct or overestimated envelope dimensions, the biases are nearly zero and the estimated variances only increase linearly. Although all of AIC-MCMC, DIC and WAIC show some tendency in over-estimating $d_X$ and $d_Y$ in Table 2, the most significant one is DIC (see the left panel of Figure 1). Hence, AIC-MCMC and WAIC may also be considered in practice, especially for small sample sizes or the large $d_X$ and large $d_Y$ setting. The Bayesian CV chooses envelope dimensions based on the prediction performance. Although its performance in selecting the true envelope dimensions is far from the best in our simulation, it is still recommended in practice if the prediction performance is our utmost concern and the computational burden is manageable.

## 7.2 Estimation Performance

### 7.2.1 Setting

In this section, we compare the estimation performance of SIMP with the Bayesian partial predictor envelope (PX-env), the Bayesian partial response envelope (PY-env) and several other classical envelope methods or multivariate regression approaches, including the frequentist predictor

envelope (FX-env) [12], the frequentist response envelope (FY-env) [15], the frequentist simultaneous envelope (FS-env) [14], the frequentist partial response envelope (FPY-env) [57], the frequentist ordinary least squares (FOLS), the principal component regression (PCR) [32], the partial least squares regression (PLSR) [16], the canonical correlation analysis (CCA) [26] and the reduced rank regression (RRR) [29]. Note in Sections 7.2 and 8, PX-env and PY-env simply refer to SIMP with $d_Y$ and $d_X$ fixed at $r$ and $p_C$ respectively, rather than the Bayesian versions of the partial predictor and partial response envelopes as exactly indicated in Table 1, however their differences are small (see the simulation results between FPY-env and PY-env in Sections 7.2.2–7.2.5) but will facilitate the comparison. We choose the posterior mean as the point estimator here and later in Section 8 for all the Bayesian (envelope) approaches. It is a decision-theoretic estimator and could be obtained easily once the posterior samples are available. In comparison, the posterior median and the maximum a posteriori (MAP) estimators are not considered in our paper, since they are expected to perform similarly (see Table C.4 in Appendix C for a numerical evidence of the posterior median estimator) due to the bell-shaped posterior distribution that is observed in both simulation studies (see Appendix G) and the real data application (see Appendix H.2), and the MAP estimator requires additional computation for the optimization.

In this section, we keep almost all set-ups from Section 7.1 except the following. We test on nine cases, for each combination of $r = 3, 8$ or 15 with $n = 300, 500$ or 1000. The covariates of $\mathbb{X}_{1D}$ are generated independently from discrete $Unif\{0, 1\}$. Dimensions $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$ are both fixed at 2. The specification of the hyperparameters (if still existed), the number of iterations and the burn-in proportion of the MCMC algorithm for PX-env, PY-env and SIMP in this section are the same as those for SIMP in Section 7.1.

To make comparison fair, for competitors that cannot account for the partial structures (i.e., all methods excluding SIMP, PX-env, PY-env and FPY-env), $\boldsymbol{Y}$ is regressed on $\boldsymbol{X}_2$ first, and the fitted residuals are the actual responses $\check{\boldsymbol{Y}}$ for the regression on our interested predictors $\boldsymbol{X}_1$. This is a common way to adjust the effects of prognostic factors that are not of main interest for IGP [68]. Tuning parameters or envelope dimensions for all methods are assumed to be unknown. The envelope dimensions for all Bayesian envelope methods (SIMP, PX-env, PY-env) are chosen by BIC-MCMC, and those for the frequentist envelope methods (FPY-env, FX-env, FY-env and FS-env) are chosen by the traditional BIC. The numbers of components for PCR and PLSR are selected based on minimizing the Mean squared prediction error (MSPE) from 5-fold CV. For $k = 1, 2, \ldots, 5$, let $\mathcal{C}_k$ represent the $k$-th index set among five partition sets of $n$ samples. For $\ell = 1, 2, \ldots, p_1$, we determine the number of components as the $\ell$ that minimizes

$$\text{MSPE}(\ell) = \frac{1}{n} \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \sum_{j=1}^{r} \left( \check{y}_{ij} - \widehat{\widetilde{y}}_{ij}^{-k}(\ell) \right)^2. \quad (7.1)$$

For the sample $i$ in $\mathcal{C}_k$, $\check{y}_{ij}$ represents the $j$-th response adjusted by $\boldsymbol{X}_2$, while $\widehat{\widetilde{y}}_{ij}^{-k}(\ell)$ is the prediction of $\check{y}_{ij}$ by $\boldsymbol{X}_{1C,i}$ and $\boldsymbol{X}_{1D,i}$, for either PCR or PLSR trained on all samples except those in $\mathcal{C}_k$, and with number of components as $\ell$. The rank for RRR is determined by the original Bura-Cook estimator [4]. As the estimation performance of $\boldsymbol{\mu}_{1C}$, $\boldsymbol{\mu}_{\boldsymbol{Y}}$ and $\boldsymbol{\beta}_2$ is very close among the competitors that we choose, and not of main interest to us, their results are not reported for compactness. For the true data generating mechanisms, we consider three models (M1-M3) from SIMP, and one model (M4) from RRR which is to investigate the model mis-specification.

(M1) $\{\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0\} = \{10\mathbf{I}_{d_{\boldsymbol{X}}}, \mathbf{I}_{p_C - d_{\boldsymbol{X}}}, \mathbf{I}_{d_{\boldsymbol{Y}}}, 5\mathbf{I}_{r - d_{\boldsymbol{Y}}}\}$,
(M2) $\{\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0\} = \{\mathbf{I}_{d_{\boldsymbol{X}}}, 10\mathbf{I}_{p_C - d_{\boldsymbol{X}}}, \mathbf{I}_{d_{\boldsymbol{Y}}}, 5\mathbf{I}_{r - d_{\boldsymbol{Y}}}\}$,
(M3) $\{\boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0\} = \{10\mathbf{I}_{d_{\boldsymbol{X}}}, \mathbf{I}_{p_C - d_{\boldsymbol{X}}}, 5\mathbf{I}_{d_{\boldsymbol{Y}}}, \mathbf{I}_{r - d_{\boldsymbol{Y}}}\}$,
(M4) $\boldsymbol{\beta}_{1C} = \mathbf{G}_C \mathbf{H}$, $\boldsymbol{\beta}_{1D} = \mathbf{G}_D \mathbf{H}$, where $\mathbf{G}_C \in \mathbb{R}^{p_C \times d_1}$, $\mathbf{G}_D \in \mathbb{R}^{p_D \times d_1}$, $\mathbf{H} \in \mathbb{R}^{d_1 \times r}$, with $d_1 = 1$ and their elements are all generated independently from $Unif(-1, 1)$. Samples of $\mathbb{X}_{1C}$ are generated from $\mathcal{N}_{p_C}(\boldsymbol{\mu}_{1C}, \boldsymbol{\Sigma}_C)$, and $\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}$ and $\boldsymbol{\Sigma}_C$ are one realizations of $\mathcal{IW}(\mathbf{I}_r, r)$ and $\mathcal{IW}(\mathbf{I}_{p_C}, p_C)$ respectively.

Note that we allow the true data generating models to be SIMP with $\|\boldsymbol{\Omega}\| \gg \|\boldsymbol{\Omega}_0\|$ and $\|\boldsymbol{\Phi}\| \ll \|\boldsymbol{\Phi}_0\|$ under (M1) or $\|\boldsymbol{\Omega}\| \ll \|\boldsymbol{\Omega}_0\|$ and $\|\boldsymbol{\Phi}\| \ll \|\boldsymbol{\Phi}_0\|$ under (M2) or $\|\boldsymbol{\Omega}\| \gg \|\boldsymbol{\Omega}_0\|$ and $\|\boldsymbol{\Phi}\| \gg \|\boldsymbol{\Phi}_0\|$ under (M3). (M4) assumes the true model to be RRR and is designed to test the robustness of SIMP under model mis-specification. Results under all scenarios are displayed from Section 7.2.2 to Section 7.2.5.

### 7.2.2 Results on (M1): $\|\boldsymbol{\Omega}\| \gg \|\boldsymbol{\Omega}_0\|$ and $\|\boldsymbol{\Phi}\| \ll \|\boldsymbol{\Phi}_0\|$

Under (M1), the mean squared errors (MSE) for $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$, two parameters that we are most interested in, are reported in Table 3. For almost every combination of $r$ and $n$, SIMP always performs the best for the estimation of both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$. PX-env or PY-env, two special cases of SIMP by setting $d_{\boldsymbol{Y}} = r$ or $d_{\boldsymbol{X}} = p_C$, can achieve competitive performance for either only $\boldsymbol{\beta}_{1C}$, or whole $\boldsymbol{\beta}_1$ but is still less efficient than SIMP. Comparing these two models, we find the former obtains a slightly better performance for $\boldsymbol{\beta}_{1C}$ while the latter achieves a much better performance for $\boldsymbol{\beta}_{1D}$. This observation is consistent with our assumptions for SIMP in (3.7), where we have assumed both the partial predictor and partial response envelope structures for $\boldsymbol{\beta}_{1C}$ with a slightly stronger signal from the former component here ($\|\boldsymbol{\Omega}\|/\|\boldsymbol{\Omega}_0\| > \|\boldsymbol{\Phi}_0\|/\|\boldsymbol{\Phi}\|$), and only the partial response envelope structure is imposed on $\boldsymbol{\beta}_{1D}$. Although PY-env and SIMP shares the same partial response envelope structure for $\boldsymbol{\beta}_{1D}$, it is surprising that the latter outperforms the former for the estimation of $\boldsymbol{\beta}_{1D}$ when $r = 3$ and 8, which is possibly due to the reason that the more efficient estimator of $\boldsymbol{\beta}_{1C}$ improves the estimation of $\mathbf{R}(\mathbf{B})$, and hence the estimation of $\boldsymbol{\beta}_{1D}$ in SIMP. Meanwhile, it is noteworthy that PY-env and FPY-env show similar estimation performance for both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ under all cases, but the former one allows more convenient posterior variability quantification and utilization of the prior information. FX-env, FY-env, FS-env and other five non-envelope methods cannot outperform their partial envelope counterparts and any partial envelope methods respectively.

To further illustrate the advantage of SIMP from the perspective of posterior variability, the average of the posterior standard deviations (PSD) of each coordinate of $\text{vec}(\boldsymbol{\beta}_{1C})$ and $\text{vec}(\boldsymbol{\beta}_{1D})$ is calculated, as the average of empirical standard deviations of posterior samples of $\text{vec}(\boldsymbol{\beta}_{1C})$ and $\text{vec}(\boldsymbol{\beta}_{1D})$ across 500 repetitions. The performance on average PSD is compared between SIMP, Bayesian linear regression (BLR) by implementing SIMP with $d_{\boldsymbol{X}} = p_C$ and $d_{\boldsymbol{Y}} = r$, PX-env and PY-env under (M1). The envelope dimensions are assumed to be unknown and selected by BIC-MCMC for all three envelope methods. The results are reported in Figure 2. In terms of both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$, three envelope methods can always significantly outperform BLR under all cases. SIMP and PX-env obtain the lowest average PSD for $\boldsymbol{\beta}_{1C}$, and are better than PY-env especially when $r = 3$. This is as expected since we have imposed both partial predictor and partial response envelope structures on $\boldsymbol{\beta}_{1C}$ but the signal of the former is stronger. As $r$ becomes 8 or 15, the performance of PY-env is close to other

*Table 3.* MSE *comparison between* SIMP *and other 11 competitors for estimating* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M1). The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face.*

| $r$ | $n$ | | FOLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.16 | 0.07 | 0.16 | 0.17 | 3.95 | 0.10 | 0.05 | 0.05 | 0.05 | **0.02** | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.55 | 0.27 | 0.63 | 0.69 | 13.52 | 0.98 | 0.18 | 0.33 | 0.16 | 0.23 | 0.17 | **0.08** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.09 | 0.04 | 0.09 | 0.09 | 3.81 | 0.05 | 0.03 | 0.02 | 0.03 | **0.01** | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.34 | 0.16 | 0.34 | 0.36 | 12.87 | 0.38 | 0.10 | 0.14 | 0.09 | 0.14 | 0.10 | **0.05** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.04 | 0.02 | 0.04 | 0.04 | 3.80 | 0.02 | 0.01 | 0.01 | 0.01 | **0.00** | 0.01 | **0.00** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.16 | 0.07 | 0.16 | 0.17 | 13.68 | 0.17 | 0.05 | 0.06 | 0.05 | 0.07 | 0.05 | **0.02** |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.69 | 0.10 | 0.30 | 0.27 | 23.17 | 0.23 | 0.06 | 0.07 | 0.05 | 0.03 | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.58 | 0.39 | 3.23 | 3.14 | 86.62 | 2.77 | 0.19 | 0.52 | 0.18 | 0.93 | 0.18 | **0.11** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.40 | 0.05 | 0.28 | 0.27 | 22.80 | 0.15 | 0.03 | 0.03 | 0.03 | 0.02 | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.53 | 0.21 | 3.14 | 2.96 | 85.83 | 1.57 | 0.12 | 0.21 | 0.11 | 0.57 | 0.11 | **0.06** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.20 | 0.02 | 0.21 | 0.22 | 22.82 | 0.08 | 0.02 | **0.01** | **0.01** | **0.01** | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.74 | 0.09 | 1.5 | 1.57 | 84.69 | 0.89 | 0.05 | 0.07 | 0.05 | 0.27 | 0.05 | **0.03** |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | 1.45 | 0.16 | 0.16 | 0.11 | 30.38 | 0.10 | 0.09 | 0.13 | 0.08 | 0.06 | 0.08 | **0.05** |
| | | $\boldsymbol{\beta}_{1D}$ | 4.74 | 0.38 | 1.39 | 1.37 | 103.70 | 1.31 | 0.19 | 0.80 | **0.15** | 1.87 | **0.15** | 0.17 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.86 | 0.09 | 0.16 | 0.10 | 30.18 | 0.08 | 0.05 | 0.06 | 0.04 | 0.04 | 0.04 | **0.03** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.86 | 0.21 | 1.34 | 1.32 | 103.07 | 1.31 | 0.10 | 0.42 | **0.09** | 1.13 | **0.09** | **0.09** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.42 | 0.04 | 0.14 | 0.11 | 30.20 | 0.07 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.39 | 0.09 | 1.29 | 1.24 | 102.97 | 1.30 | 0.05 | 0.15 | **0.04** | 0.54 | **0.04** | **0.04** |
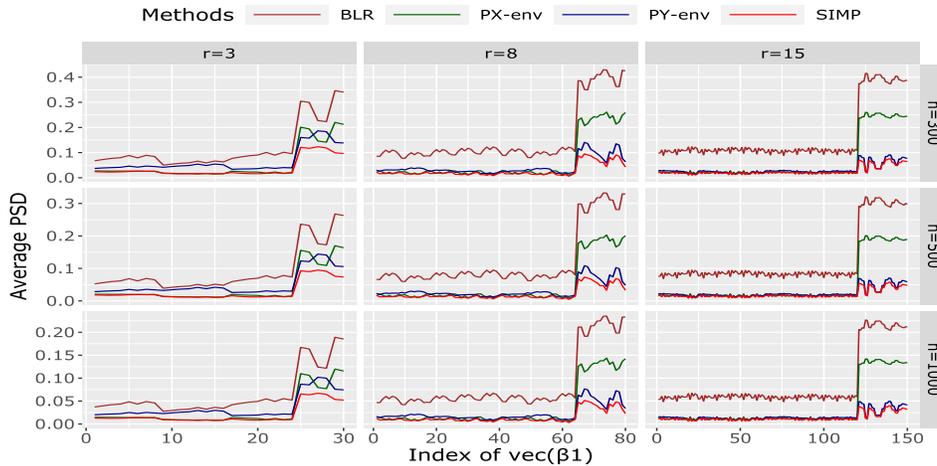


Figure 2: Average PSD across each coordinate of vec($\boldsymbol{\beta}_{1C}$) (1-8$r$), followed by vec($\boldsymbol{\beta}_{1D}$) ((8$r$+1)-10$r$), over 500 repetitions. Rows and columns indicate different sample sizes and dimensions of responses.

two envelope methods. In terms of $\boldsymbol{\beta}_{1D}$, SIMP achieves the lowest PSD under all scenarios. When $r = 8$ or 15, the performance of PY-env is close to SIMP, and is significantly better than PX-env. This is as expected as well, since the partial response envelope structure is imposed on $\boldsymbol{\beta}_{1D}$ for both SIMP and PY-env, and the slight advantage of SIMP for the estimation of $\boldsymbol{\beta}_{1D}$ is possibly due to the more efficient estimation of $\boldsymbol{\beta}_{1C}$, which is one synergetic effect of imposing two envelope components simultaneously. Although no envelope assumption is imposed on $\boldsymbol{\beta}_{1D}$ for PX-env, the posterior variability for the estimator of $\boldsymbol{\beta}_{1D}$ from PX-env is still smaller than that of BLR.

### 7.2.3 Results on (M2): $\|\boldsymbol{\Omega}\| \ll \|\boldsymbol{\Omega}_0\|$ and $\|\boldsymbol{\Phi}\| \ll \|\boldsymbol{\Phi}_0\|$

Table 4 shows the performance under (M2). Since $\|\boldsymbol{\Omega}\| \ll \|\boldsymbol{\Omega}_0\|$, the partial predictor envelope component cannot provide too many efficiency gains. Therefore, as expected, the best estimation performance comes from envelope models that contain the (partial) response envelope structure, including FY-env, FS-env, FPY-env, PY-env and SIMP. Although the best performance is mostly achieved by FY-env and FPY-env, the difference between these five methods is small. Since the partial response envelope structure is imposed on both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ in (3.7), large efficiency gains could be observed in estimating both of these two parame-

**Table 4**. MSE *comparison between* SIMP *and other 11 competitors for estimating* $\beta_{1C}$ *and* $\beta_{1D}$ *over* 500 *repetitions under the data generating mechanism (M2). The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face.*

| $r$ | $n$ | | FOLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\beta_{1C}$ | 0.06 | 0.05 | 0.06 | 0.06 | 1.31 | 0.05 | 0.04 | 0.02 | 0.04 | 0.06 | 0.04 | **0.03** |
| | | $\beta_{1D}$ | 0.32 | 0.46 | 0.34 | 0.35 | 7.11 | 0.33 | 0.47 | **0.12** | 0.45 | 0.31 | 0.48 | **0.12** |
| | 500 | $\beta_{1C}$ | 0.04 | 0.03 | 0.04 | 0.04 | 1.29 | 0.03 | **0.01** | **0.01** | **0.01** | 0.04 | **0.01** | 0.02 |
| | | $\beta_{1D}$ | 0.20 | 0.22 | 0.20 | 0.20 | 7.04 | 0.21 | 0.07 | 0.07 | 0.07 | 0.19 | 0.08 | **0.06** |
| | 1000 | $\beta_{1C}$ | 0.02 | 0.01 | 0.02 | 0.02 | 1.26 | 0.02 | **0.01** | **0.01** | **0.01** | **0.01** | **0.01** | **0.01** |
| | | $\beta_{1D}$ | 0.10 | 0.08 | 0.10 | 0.10 | 6.87 | 0.10 | **0.03** | 0.04 | **0.03** | 0.09 | **0.03** | **0.03** |
| 8 | 300 | $\beta_{1C}$ | 0.30 | 0.17 | 0.31 | 0.31 | 5.74 | 0.25 | **0.04** | 0.05 | **0.04** | 0.23 | **0.04** | **0.04** |
| | | $\beta_{1D}$ | 1.42 | 0.52 | 1.63 | 1.59 | 39.61 | 1.51 | **0.11** | 0.17 | **0.11** | 1.25 | 0.12 | 0.13 |
| | 500 | $\beta_{1C}$ | 0.17 | 0.09 | 0.18 | 0.18 | 5.63 | 0.14 | **0.02** | 0.03 | **0.02** | 0.13 | **0.02** | **0.02** |
| | | $\beta_{1D}$ | 0.86 | 0.29 | 1.00 | 1.00 | 39.68 | 0.91 | **0.07** | 0.10 | **0.07** | 0.76 | 0.08 | 0.08 |
| | 1000 | $\beta_{1C}$ | 0.08 | 0.04 | 0.08 | 0.08 | 5.42 | 0.07 | **0.01** | **0.01** | **0.01** | 0.06 | **0.01** | **0.01** |
| | | $\beta_{1D}$ | 0.41 | 0.13 | 0.41 | 0.43 | 38.19 | 0.41 | **0.03** | 0.04 | **0.03** | 0.37 | 0.04 | 0.04 |
| 15 | 300 | $\beta_{1C}$ | 0.60 | 0.42 | 0.59 | 0.55 | 5.19 | 0.43 | **0.10** | 0.16 | **0.10** | 0.46 | 0.12 | 0.13 |
| | | $\beta_{1D}$ | 2.32 | 0.29 | 1.83 | 1.19 | 49.09 | 1.36 | **0.09** | 0.40 | **0.09** | 2.04 | 0.11 | 0.17 |
| | 500 | $\beta_{1C}$ | 0.36 | 0.25 | 0.37 | 0.34 | 5.00 | 0.27 | **0.06** | 0.09 | **0.06** | 0.28 | **0.06** | 0.07 |
| | | $\beta_{1D}$ | 1.42 | 0.16 | 1.56 | 1.13 | 49.09 | 0.97 | **0.05** | 0.20 | 0.06 | 1.24 | 0.06 | 0.09 |
| | 1000 | $\beta_{1C}$ | 0.18 | 0.12 | 0.18 | 0.17 | 4.80 | 0.13 | **0.03** | **0.03** | **0.03** | 0.14 | **0.03** | **0.03** |
| | | $\beta_{1D}$ | 0.68 | 0.07 | 0.69 | 0.96 | 48.92 | 0.50 | **0.03** | 0.05 | **0.03** | 0.60 | **0.03** | 0.04 |

**Table 5**. MSE *comparison between* SIMP *and other 11 competitors for estimating* $\beta_{1C}$ *and* $\beta_{1D}$ *over 500 repetitions under the data generating mechanism (M3). The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face.*

| $r$ | $n$ | | FOLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\beta_{1C}$ | 0.25 | 0.23 | 0.28 | 0.27 | 1.84 | 0.22 | 0.19 | 0.20 | 0.18 | **0.03** | 0.23 | 0.04 |
| | | $\beta_{1D}$ | 0.89 | 0.84 | 1.90 | 1.65 | 9.37 | 3.56 | 1.57 | 2.63 | 1.56 | **0.34** | 0.88 | 0.38 |
| | 500 | $\beta_{1C}$ | 0.14 | 0.13 | 0.15 | 0.15 | 1.44 | 0.14 | 0.13 | 0.11 | 0.13 | **0.01** | 0.13 | 0.02 |
| | | $\beta_{1D}$ | 0.52 | 0.48 | 0.68 | 0.69 | 8.69 | 2.34 | 0.71 | 1.52 | 0.69 | 0.21 | 0.46 | **0.20** |
| | 1000 | $\beta_{1C}$ | 0.07 | 0.07 | 0.07 | 0.07 | 1.21 | 0.05 | 0.06 | 0.04 | 0.06 | **0.01** | 0.06 | **0.01** |
| | | $\beta_{1D}$ | 0.25 | 0.23 | 0.26 | 0.27 | 8.18 | 0.78 | 0.23 | 0.39 | 0.23 | **0.10** | 0.23 | **0.10** |
| 8 | 300 | $\beta_{1C}$ | 0.35 | 0.23 | 0.29 | 0.29 | 5.25 | 0.06 | 0.22 | 0.18 | 0.22 | **0.02** | 0.22 | 0.04 |
| | | $\beta_{1D}$ | 1.28 | 0.84 | 2.86 | 2.71 | 19.95 | 3.55 | 0.81 | 2.21 | 0.80 | 0.47 | 0.80 | **0.34** |
| | 500 | $\beta_{1C}$ | 0.21 | 0.14 | 0.22 | 0.22 | 4.96 | 0.08 | 0.13 | 0.11 | 0.13 | **0.01** | 0.13 | 0.02 |
| | | $\beta_{1D}$ | 0.78 | 0.51 | 1.79 | 1.66 | 19.02 | 3.47 | 0.50 | 1.37 | 0.49 | 0.29 | 0.49 | **0.21** |
| | 1000 | $\beta_{1C}$ | 0.10 | 0.06 | 0.10 | 0.10 | 4.82 | 0.15 | 0.06 | 0.05 | 0.06 | **0.01** | 0.06 | **0.01** |
| | | $\beta_{1D}$ | 0.37 | 0.24 | 0.44 | 0.47 | 18.25 | 2.69 | 0.23 | 0.61 | 0.23 | 0.14 | 0.23 | **0.09** |
| 15 | 300 | $\beta_{1C}$ | 0.51 | 0.25 | 0.16 | 0.14 | 7.08 | 0.05 | 0.24 | 0.12 | 0.23 | **0.03** | 0.23 | 0.05 |
| | | $\beta_{1D}$ | 1.64 | 0.77 | 1.32 | 1.27 | 22.01 | 1.39 | 0.73 | 1.37 | 0.70 | 0.66 | 0.70 | **0.36** |
| | 500 | $\beta_{1C}$ | 0.30 | 0.14 | 0.14 | 0.13 | 6.67 | 0.04 | 0.14 | 0.08 | 0.13 | **0.02** | 0.13 | 0.03 |
| | | $\beta_{1D}$ | 0.98 | 0.45 | 1.29 | 1.23 | 21.46 | 1.38 | 0.43 | 1.16 | 0.41 | 0.40 | 0.41 | **0.21** |
| | 1000 | $\beta_{1C}$ | 0.15 | 0.07 | 0.12 | 0.11 | 6.39 | 0.03 | 0.07 | 0.05 | 0.07 | **0.01** | 0.07 | **0.01** |
| | | $\beta_{1D}$ | 0.48 | 0.22 | 1.09 | 1.09 | 21.05 | 1.38 | 0.21 | 0.66 | 0.21 | 0.20 | 0.21 | **0.10** |

ters for these five methods, as expected. Among these five methods, the slightly worse performance of SIMP under some cases is largely due to the model selection. See the performance with the true envelope dimensions known in Table C.3 in Appendix C.

### 7.2.4 Results on (M3): $\|\Omega\| \gg \|\Omega_0\|$ and $\|\Phi\| \gg \|\Phi_0\|$

The estimation performance under (M3) is shown in Table 5. Without surprise, the best performance is obtained by envelope models containing the partial predictor envelope structure, i.e. PX-env and SIMP. Similarly, their estimation performances for $\beta_{1C}$ are close, and the tiny difference between them mainly comes from the model selection. It is worthy to notice that when $r = 8$ or 15, some additional gains on estimating $\beta_{1D}$ could be offered by SIMP compared with PX-env, since the partial response envelope structure is assumed in SIMP for $\beta_{1D}$, although the signal may not be strong ($\|\Phi\| \gg \|\Phi_0\|$), however, the synergetic effect of

*Table 6.* MSE *comparison between* SIMP *and other 11 competitors for estimating* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M4). The lowest* MSE *under each combination of* $r$ *and* $n$ *among all methods and within envelope methods are in bold face and blue respectively.*

| $r$ | $n$ | | FOLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.92 | **0.46** | 0.98 | 0.77 | 27.61 | 0.57 | 0.90 | 0.59 | 0.91 | 0.51 | 0.90 | 0.53 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.12 | **0.06** | 0.12 | 0.11 | 3.58 | 0.06 | 0.12 | 0.07 | 0.12 | 0.12 | 0.12 | 0.12 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.53 | **0.26** | 0.60 | 0.45 | 27.60 | 0.43 | 0.53 | 0.43 | 0.53 | 0.38 | 0.53 | 0.39 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.07 | **0.04** | 0.08 | 0.07 | 3.72 | 0.05 | 0.07 | 0.05 | 0.07 | 0.07 | 0.07 | 0.07 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.27 | **0.13** | 0.34 | 0.25 | 28.57 | 0.28 | 0.27 | 0.28 | 0.27 | 0.26 | 0.27 | 0.26 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.03 | **0.01** | 0.03 | 0.03 | 3.37 | 0.02 | 0.03 | 0.02 | 0.03 | 0.03 | 0.03 | 0.03 |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.75 | **0.34** | 0.92 | 0.86 | 20.53 | 1.31 | 0.48 | 0.88 | 0.44 | 1.44 | 0.43 | 0.38 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.09 | **0.04** | 0.09 | 0.08 | 2.49 | 0.06 | 0.06 | 0.05 | 0.05 | 0.08 | 0.05 | 0.05 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.43 | **0.19** | 0.48 | 0.44 | 20.32 | 1.07 | 0.27 | 0.42 | 0.25 | 0.72 | 0.24 | 0.23 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.06 | **0.03** | 0.06 | 0.05 | 2.62 | 0.04 | 0.04 | **0.03** | **0.03** | 0.05 | **0.03** | 0.04 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.20 | **0.08** | 0.22 | 0.17 | 20.14 | 0.40 | 0.15 | 0.17 | 0.15 | 0.16 | 0.15 | 0.14 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.03 | **0.01** | 0.03 | 0.02 | 2.55 | 0.02 | 0.02 | 0.02 | 0.02 | 0.03 | 0.02 | 0.02 |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | 37.13 | 13.11 | 16.66 | 18.15 | 20.23 | 17.16 | **6.93** | 10.85 | 12.94 | 15.56 | 11.25 | 12.99 |
| | | $\boldsymbol{\beta}_{1D}$ | 4.74 | 1.55 | 2.50 | 2.01 | 2.10 | **0.09** | 0.70 | 0.11 | 1.67 | 4.74 | 1.31 | 2.73 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 22.73 | **5.75** | 13.16 | 11.24 | 20.95 | 10.32 | 7.59 | 12.67 | 17.04 | 9.21 | 15.42 | 14.60 |
| | | $\boldsymbol{\beta}_{1D}$ | 2.85 | 0.73 | 1.69 | 1.29 | 2.05 | **0.03** | 0.78 | 0.11 | 2.03 | 2.82 | 1.85 | 2.49 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 11.54 | **2.39** | 11.04 | 7.24 | 21.22 | 7.80 | 10.10 | 9.05 | 11.45 | 6.87 | 11.35 | 8.96 |
| | | $\boldsymbol{\beta}_{1D}$ | 1.46 | 0.29 | 0.91 | 0.63 | 2.08 | **0.01** | 1.24 | 0.10 | 1.44 | 1.46 | 1.43 | 1.43 |

SIMP that is mentioned above could possibly enhance this advantage.

### 7.2.5 Results on (M4): Model Mis-Specification

To test the robustness of our method, an additional simulation under the model mis-specification scenario (M4) is studied. From Table 6, RRR estimator almost always produces the lowest MSE when $r = 3$ or 8. Although under the mis-specified model, envelope methods still perform quite well. Within envelope methods, SIMP obtains a relatively better performance for both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ when $r$ is 8, and FX-env is better for $\boldsymbol{\beta}_{1D}$ when $r$ is 3 or 15. Note that when $r = 15$, several envelope methods could even outperform the RRR estimator. When we are more cautious for the results displayed here, we will find that the performance of all envelope methods especially SIMP is largely contaminated by the "bad" envelope dimensions that are selected under this model mis-specification scenario, since the likelihoods of them might significantly depart from the truth and hence are misleading in model selection. See Table C.5 in Appendix C for the estimation results if we rule out this factor by using the "optimal" tuning parameter or envelope dimension for each method, if we pretend to know the true values of $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$. Under such scenario, the envelope methods almost always dominate other methods (even RRR), and SIMP is almost always the best one among envelope methods except when $r = 3$. And when $r = 3$, the performance of SIMP is very close to the best one. Therefore, the robustness of our method and other envelope methods to a general low rank model RRR is observed in our simulation.

Besides, in Appendix C, we also display additional numerical results for the estimation performance of FPY-env, PX-env, PY-env and SIMP with the true envelope dimensions known under (M1)–(M3) or RRR with the true rank known under (M4), and the results for $p_2 = 4$ or $p_D = 4$ with other settings same as those in (M1).

## 8. REAL IMAGING GENETICS APPLICATION ON ADNI1

### 8.1 Background

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). Led by Principal Investigator Michael W. Weiner, MD, ADNI was launched in 2003 as a public-private partnership, and lasts for 4 stages (ADNI1, ADNI-GO, ADNI2 and ADNI3) till now. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), genetic or other biological markers, clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). This study has strict enrollment standards, follow-up and data checking protocols. All raw or preprocessed data is available through the Image and Data Archive (IDA) at https://ida.loni.usc.edu/, upon approval of the application. In this real data application, we want to study the genetic effects, including the effect of a well-known AD gene Apolipoprotein E (APOE) $\epsilon 4$, on the volumes of some cortical or sub-cortical structures in brain,

while also accounting for the effects of some other prognostic factors by applying SIMP, and compare with several other methods in terms of the prediction performance. Following previous practices on analyzing ADNI data [68, 46], we focus on Caucasian participants in the ADNI1 phase, to reduce the population stratification.

## 8.2 Data Pre-processing Procedure

Imaging responses, i.e. $\boldsymbol{Y}$ in (3.2), we adopted are the volume measurements on 12 ROIs that are listed in [42] (i.e., the Amygdala, the Cerebral cortex, the Cerebral white matter, the Hippocampus, the Inferior lateral ventricle and the Lateral ventricle on two hemispheres), which are obtained from 1.5T MRI scan at the screening visit of ADNI1. All imaging data is downloaded from IDA, and has been pre-processed by FreeSurfer [21] already. Among 845 participants who took MRI scans, we only retain 593 participants who passed the quality control (QC). The log transformation is applied to reduce skewness.

620,901 SNPs on all autosomes and sex chromosomes were genotyped for 757 participants in ADNI1, who are not contained in but are overlapped with the 845 participants who have MRI scans. Among 11,632 SNPs meta-analyzed by International Genomics of Alzheimer's Project (IGAP) [36], 852 autosome SNPs are selected as our candidate SNPs for model fitting by checking two conditions: (1) Genotyped by ADNI1; (2) P-values from IGAP are smaller than 0.01. Developed by Shaun Purcell and Christopher Chang, PLINK [7] is an open source software for performing QC and some routine analyses in GWAS in an efficient manner, and is available at https://www.cog-genomics.org/plink/1.9/. Following [68], we perform the following two lines of QC on the selected SNP data by PLINK. The first line of QC includes (1) call rate check per subject and per SNP marker, (2) gender check, (3) sibling pair identification, (4) Hardy-Weinberg equilibrium test, and (5) population stratification check. The second line of QC removes SNPs with (1) more than 5% missing values, (2) minor allele frequency smaller than 10%, and (3) Hardy-Weinberg equilibrium p-value $< 10^{-6}$. This leaves us with 732 SNPs that passed QC. A free genotype imputation and haplotype phasing program IMPUTE2 ([27], [28]) is used to impute SNPs with missingness. We choose HapMap 3 as our reference panel for the imputation, since it shares the same genome build b36 of National Center for Biotechnology Information (NCBI) with ADNI1. We delete 39 SNPs that are not shown in the reference panel,

which leads to 693 SNPs finally. To reduce dimensionality, the principal component analysis (PCA) is applied on these 693 SNPs. By thresholding the eigenvalues of its sample covariance matrix by 1, the first 186 principal components (PCs), which explains 87.19% of the total variation of these 693 SNP predictors, are selected as our $\boldsymbol{X}_{1C}$, the continuous part of our predictors of main interest in (3.2).

APOE $\epsilon 4$, taking values in $\{0, 1, 2\}$, is included in $\boldsymbol{X}_{1D}$ as the only discrete (quantitative) genetic predictor in (3.2). We have incorporated six other important prognostic factors in $\boldsymbol{X}_2$ in (3.2), including gender, marital status, handness, age, years of education and intracranial volume (ICV). The intersection of participants from all types of abovementioned datasets leads to 498 samples in our analysis finally. All predictors and responses are standardized before the model fitting.

## 8.3 Prediction Performance

We first compare the prediction performance of SIMP with other two partial envelope methods (PX-env, FPY-env), three envelope models that cannot account for the partial structures (FX-env, FY-env, FS-env; we call them non-partial envelope methods), and three popular multivariate linear regression methods that are not based on the envelope models (FOLS, RRR and PLSR; we call them non envelope methods). Similar with Section 7.2, for all methods except the partial envelope methods (PX-env, FPY-env and SIMP), $\boldsymbol{Y}$ is adjusted as the residuals from the regression on $\boldsymbol{X}_2$ first, and the model fitting will be done by the genetic markers $\boldsymbol{X}_1$ and the adjusted responses. For the best prediction performance, all associated envelope dimensions for envelope methods, or the number of components for PLSR, or the rank for RRR are selected by minimizing MSPE from 5-fold CV. The MCMC algorithm for SIMP and PX-env is run for 20,000 iterations with first 50% iterations as burn-in, and their specifications for the hyperparameters are the same as those in Section 7.

Table 7 reveals that the best prediction performance is achieved by SIMP. All of the envelope methods in the last two columns significantly outperform FOLS and RRR. FS-env and SIMP achieve the best prediction performance within the non-partial envelope methods and partial envelope methods we considered respectively. This illustrates the advantages of combining the predictor and response envelopes in both non-partial and partial envelope contexts. Comparing partial envelope methods with their non-partial

*Table 7.* MSPE *from 5-fold CV for various envelope and non envelope methods. Envelope dimensions, number of components (*ncomp*) or rank selected by 5-fold CV are mentioned inside parenthesis.*

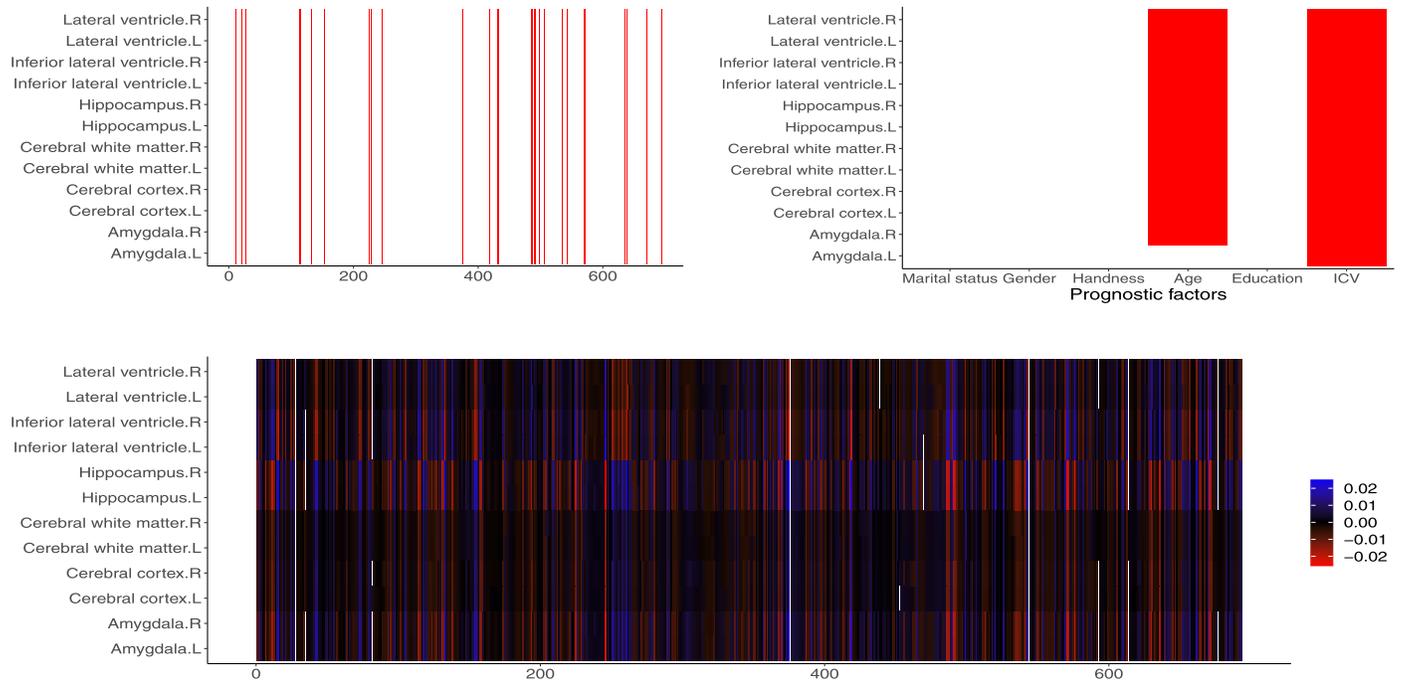| Non envelope methods | | Non-partial envelope methods | | Partial envelope methods | |
|---|---|---|---|---|---|
| FOLS | 14.02 | FX-env ($d_{\boldsymbol{X}} = 1$) | 8.34 | PX-env ($d_{\boldsymbol{X}} = 1$) | 7.94 |
| RRR (rank = 1) | 10.87 | FY-env ($d_{\boldsymbol{Y}} = 1$) | 9.77 | FPY-env ($d_{\boldsymbol{Y}} = 1$) | 9.82 |
| PLSR (ncomp = 1) | 8.20 | FS-env ($d_{\boldsymbol{X}} = 7, d_{\boldsymbol{Y}} = 1$) | 8.06 | SIMP($d_{\boldsymbol{X}} = 1, d_{\boldsymbol{Y}} = 3$) | 7.88 |

Figure 3: Upper left: Indicator of significance of each (SNP, IP) pair or APOE $\epsilon 4$ with each IP. APOE $\epsilon 4$ corresponds to the last vertical line in red; Upper right: Indicator of significance of each (prognostic factor, IP) pair. Red signifies significance in the figures of the upper row. Bottom: Heatmap of the estimated regression coefficients matrix $\widehat{\boldsymbol{\beta}}_{SNP}^T$ between 693 SNPs and 12 IPs.

counterparts, PX-env and SIMP both obtain lower MSPEs than FX-env and FS-env respectively. The finding that FPY-env cannot outperform FY-env seems surprising, but is consistent with the previous theoretical result that if the envelope spaces of FPY-env and FY-env are equal, then the asymptotic variance of the estimator of $\boldsymbol{\beta}_1$ from FPY-env could not be smaller than that from FY-env (see the proposition 2 of [57]), and we have verified the estimated envelope spaces by FPY-env and FY-env are very close in this analysis. It is also interesting to note that FX-env shares similar MSPE with PLSR. Their connection has been discussed in [12].

## 8.4 Posterior Estimation and Selection Performance

Both AIC-MCMC and BIC-MCMC select $d_{\boldsymbol{X}} = 87$ and $d_{\boldsymbol{Y}} = 1$ for SIMP. Using $d_{\boldsymbol{X}} = 87$ and $d_{\boldsymbol{Y}} = 1$, SIMP is fitted with the number of iterations, burn-in proportion and specification of the hyperparameters same as those in Section 7 and Section 8.3. It costs 14.82 hours for a run of 20,000 iterations under these envelope dimensions, for a Intel Xeon E5-2680 v3 processor (2.50 GHz CPU).

Aiming at interpreting our results with some scientifically meaningful findings and due to the large variation that is explained by the PCs that we choose, we intend to approximate the posterior samples of $\boldsymbol{\beta}_{SNP} \in \mathbb{R}^{693 \times 12}$, the

regression coefficients between the standardized version of the original 693 SNP predictors (i.e., the 693 SNP predictors before PCA) and 12 IPs. For $s = 1, 2, \ldots, 10,000$, the $s$-th retained posterior sample $\boldsymbol{\beta}_{1C}^{(s)} \in \mathbb{R}^{186 \times 12}$ is mapped to $\boldsymbol{\beta}_{SNP}^{(s)}$, the $s$-th approximate posterior sample of $\boldsymbol{\beta}_{SNP}$ by

$$\boldsymbol{\beta}_{SNP}^{(s)} = \left(\mathrm{diag}\{\mathbf{S}_{SNP}\}\right)^{1/2} \cdot \mathbf{LD} \cdot \left(\mathrm{diag}\{\mathbf{S}_{1C}\}\right)^{-1/2} \cdot \boldsymbol{\beta}_{1C}^{(s)}, \tag{8.1}$$

where $\mathrm{diag}\{\mathbf{S}_{SNP}\} \in \mathbb{S}_+^{693 \times 693}$ and $\mathrm{diag}\{\mathbf{S}_{1C}\} \in \mathbb{S}_+^{186 \times 186}$ are diagonal matrices with diagonal elements being sample variances of the original 693 SNPs and 186 PCs respectively, and $\mathbf{LD} \in \mathbb{R}^{693 \times 186}$ is the loading matrix from PCA.

The corresponding $(1 - 5.95 \times 10^{-6}) \times 100\%$ two-sided credible intervals for $\boldsymbol{\beta}_{SNP}$, $\boldsymbol{\beta}_{1D}$ and $\boldsymbol{\beta}_2$ are calculated, where the confidence level is determined by 0.95 with the Bonferroni correction for all 8,400 regression coefficients. Significance of any SNP, APOE $\epsilon 4$ or prognostic factor with any one of 12 IPs is determined by the exclusion of zero for the associated credible interval.

The upper-left panel of Figure 3 shows the significance of a few SNP and APOE $\epsilon 4$ with all 12 IPs. Besides the well-known AD gene APOE $\epsilon 4$, we identify another 37 significant SNPs that are possibly related to AD (see Appendix H.3 for the full list) under the control of the Family-wise error rate (FWER, the probability of reporting at least one false

*Table 8. Out of 37 significant SNPs from* SIMP *(besides APOE $\epsilon$4), 17 SNPs are reported here for the previously reported associations of them (or their RefSeq gene or their closest RefSeq gene with names indicated in Columns 2 and 5) with AD in the literature. In the literature, one previous study on the reported association is selected for each of these SNPs and displayed in Columns 3 and 6.*

| SNP | Gene | | SNP | Gene | |
|-----|------|-----|-----|------|-----|
| rs11685593 | BIN1 | [2] | rs10894473 | NTM | [66] |
| rs7561528 | BIN1 | [24] | rs11064498 | C1S | [61] |
| rs11706690 | CHL1 | [50] | rs757402 | OAS2 | [3] |
| rs10513391 | P2RY14 | [53] | rs2274736 | PTPN21 | [62] |
| rs2439538 | TBC1D7 | [17] | rs10498633 | SLC24A4/RIN3 | [60] |
| rs9381563 | CD2AP | [30] | rs2554389 | ADAMTSL3 | [55] |
| rs2280231 | NDUFS3 | [49] | rs4265771 | ADAMTSL3 | [55] |
| rs7120548 | MTCH2 | [31] | rs17809911 | CCDC102B | [41] |
| rs10501927 | CNTN5 | [24] | | | |

positives) at 0.05. Among them, 17 SNPs are reported in Table 8 for their appearance in the previous studies. Meanwhile, under the same control of FWER, the key role of the age for AD and the impact of ICV on the volumes of brain structures are confirmed at the same time, by the upper-right panel of Figure 3. The significance of SNPs seems to be consistent over all IPs but this is not observed for prognostic factors. This is expected since the rows of $\widehat{\boldsymbol{\beta}}_{1C}^T$ (and also $\widehat{\boldsymbol{\beta}}_{SNP}^T$, the posterior mean estimator of $\boldsymbol{\beta}_{SNP}^T$) depend on the estimated partial response envelope only, which happens to contribute much weaker effects than the estimated partial predictor envelope in this dataset. Such pattern is not observed in $\widehat{\boldsymbol{\beta}}_2^T$, since we have not imposed any envelope structure on $\boldsymbol{\beta}_2$.

The heatmap of the estimated regression coefficients matrix $\widehat{\boldsymbol{\beta}}_{SNP}^T$ is shown in the bottom panel of Figure 3. The elements of $\widehat{\boldsymbol{\beta}}_{1D}$ are much larger in scale than those in $\widehat{\boldsymbol{\beta}}_{SNP}^T$, hence are not displayed in this figure (The elements of $\widehat{\boldsymbol{\beta}}_{1D}$ are listed here instead. Amygdala.L: $-0.19$, Amygdala.R: $-0.19$, Cerebral cortex.L: $-0.12$, Cerebral cortex.R: $-0.13$, Cerebral white matter.L: $-0.11$, Cerebral white matter.R: $-0.11$, Hippocampus.L: $-0.22$, Hippocampus.R: $-0.22$, Inferior lateral ventricle.L: 0.18, Inferior lateral ventricle.R: 0.19, Lateral ventricle.L: 0.14, Lateral ventricle.R: 0.15). It is interesting that for each genetic predictor, the estimated coefficients related to each pair of brain measures on two hemispheres are close. This is reasonable since the bilateral correlations within all pairs of brain measures over two hemispheres are strong (see Appendix H.4 for the numerical evidence and [52] for another discussion on this bilateral correlation). The ROIs that have relatively large estimated effects with some SNPs are the Inferior lateral ventricle, the Hippocampus and the Amygdala (in both hemispheres). The Hippocampus and the Amygdala are related to memory and motor behavior respectively, and the importance of these three brain structures for AD have already been verified in the past studies [20, 47, 43].

## 8.5 Shrinkage Estimation by Incorporating the Prior Information of Weak Imaging Genetics Relationship

The past studies on AD have verified the weak effects of the SNP predictors on predicting AD outcomes, either the disease status [36] or the imaging phenotypes ([68, 46] and etc). This pattern of weak effects has also been verified by us from the bottom panel of Figure 3 in Section 8.4. Therefore, it is reasonable to incorporate the prior information of "weak signals in $\boldsymbol{\beta}_{1C}$" (and hence "weak signals in $\boldsymbol{\beta}_{SNP}$") into our analysis, and it should also improve the estimation performance from the standpoint of the shrinkage effect offered for this high dimensional problem, and help to illustrate the advantage of our Bayesian envelope method on prior information incorporation, compared with the frequentist envelope methods. The envelope dimensions are fixed at $d_{\boldsymbol{X}} = 87$ and $d_{\boldsymbol{Y}} = 1$. We adjust the (all equal) diagonal elements of the hyperparameter $\mathbf{E}$ (we set to be a diagonal matrix) in the prior of $\boldsymbol{\eta}_C$ from $10^{-6}$ to $10^6$ while keeping the specification for other hyperparameters same as previous sections. This adjustment strengthens our prior belief of $\boldsymbol{\eta}_C$ (and hence $\boldsymbol{\beta}_{1C}$) to be a zero matrix gradually, since the prior mean of $\boldsymbol{\eta}_C$ is fixed at the zero matrix by assigning $\mathbf{W}$ to be the zero matrix, and the prior covariance matrices of rows of $\boldsymbol{\eta}_C$ are proportional to $\mathbf{E}^{-1}$. From the frequentist perspective, the Bayesian estimator (strictly speaking, the MAP estimator, but here we use the posterior mean estimator for simplicity) with increasingly stronger prior belief in the weak signals of $\boldsymbol{\eta}_C$ corresponds to an estimator with gradually stronger shrinkage to the zero matrix.

The left panel of Figure 4 shows improved prediction performance as the diagonal elements of $\mathbf{E}$ increase from $10^{-6}$ to $10^3$, and the MSPE deteriorates slightly as the prior belief increases further beyond $10^3$, possibly due to the bias that is caused by the over-strong prior information of weak signals, or the over-penalization from the frequentist perspective. The middle and right panels of Figure 4 display the regularization paths for the row sums and the column
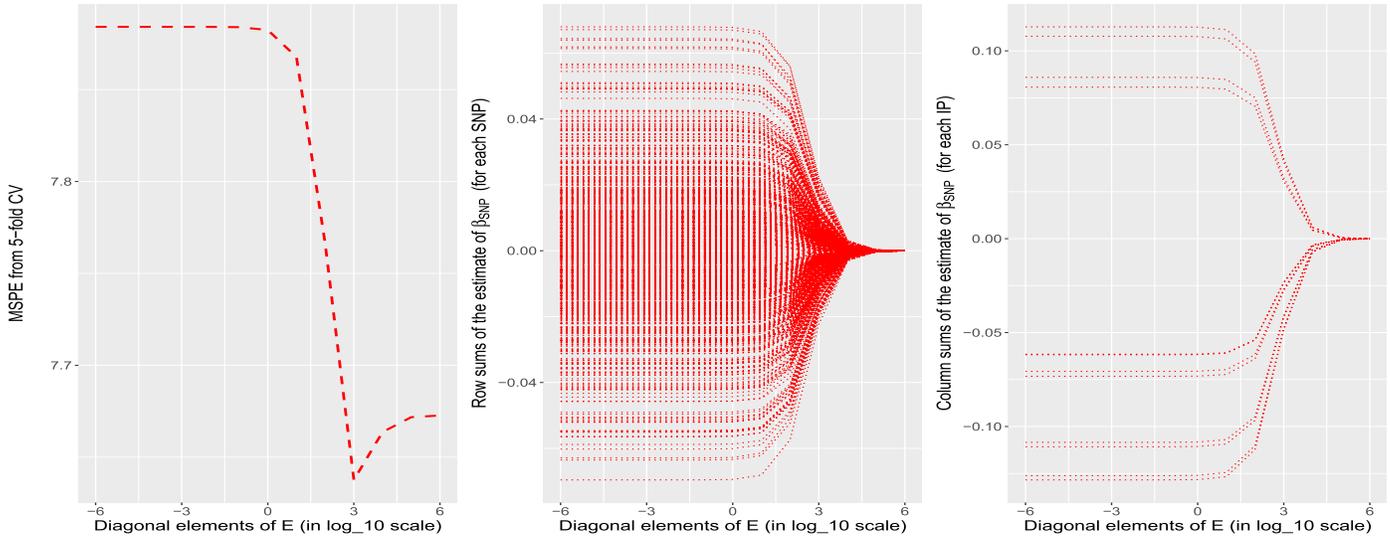
Figure 4: MSPE from the 5-fold CV (left), the row sums of $\widehat{\boldsymbol{\beta}}_{SNP}$ (middle; each line corresponds to one SNP) and the column sums of $\widehat{\boldsymbol{\beta}}_{SNP}$ (right; each line corresponds to one IP) with respect to the diagonal elements of $\mathbf{E}$ (in the $\log_{10}$ scale).

sums of $\widehat{\boldsymbol{\beta}}_{SNP}$ respectively. The more significant shrinkage effects could be observed for the stronger prior belief of the weak signals in $\boldsymbol{\beta}_{1C}$ (and hence in $\boldsymbol{\beta}_{SNP}$).

## 9. DISCUSSION

In this paper, we propose a new unified Bayesian envelope model by integrating the partial predictor envelope and the partial response envelope under a convenient Bayesian framework. The proposed model degenerates to several well established envelope models in the literature under specific conditions. It addresses the limitations mentioned for the simultaneous and the partial response envelopes. Specifically, our method improves the efficiencies for estimating $\boldsymbol{\beta}_1$ and predicting $\boldsymbol{Y}$ in (3.2), and has no restrictions on $\boldsymbol{X}_1$ to be continuous or Normal, by a subtle construction of separating the discrete part from the whole predictors of interest. Compared with the frequentist envelope approaches, our method is more flexible in incorporating prior information and quantifying uncertainty through posterior distribution of parameters. Overall, our method is inclusive and could be regarded as a building block for the future theoretical research of the envelope model, and an ideal solution for practitioners who seek dimension reduction and want to apply envelope methods to their application problems, including almost any problems that could be formulated by the multivariate linear regression with predictors either continuous or discrete or a mix of them and even containing nuisance covariates, but are worried about which specific envelope model to use. Meanwhile, we are the first to investigate the performance of several popular dimension selection methods together, among the Bayesian envelope literature, to the best of our knowledge.

To be less conservative and make more meaningful discoveries, a well-designed multiple comparisons procedure to control the False discovery rate, rather than FWER, could be considered in our real data application. A natural extension of our work is to generalize SIMP to the generalized linear model setting. However, for the multivariate probit model as pointed in [9], the identification issue exists for the error covariance matrix, which is suggested to be restricted to the correlation matrix. Hence, the extension of SIMP to the generalized linear model setting might require some special strategies, for example the parameter expansion for data augmentation technique ([40], [59]) under a marginally uniform prior for the error correlation matrix [1].

## APPENDIX A

### A.1 Metropolis-within-Gibbs MCMC Algorithm for SIMP

In this section, we show how to generate a MCMC chain of length $S$ on $\boldsymbol{\Theta}$ from SIMP, for a pre-specified pair of dimensions $(d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}) \in \{1, \ldots, p_C - 1\} \times \{1, \ldots, r - 1\}$. We could arbitrarily choose the initial value $\boldsymbol{\Theta}^{(0)}$, but a warm start initial estimator provided in Appendix B is recommended for faster convergence. For each $s = 1, \ldots, S$, we iterate the following steps to update $\boldsymbol{\Theta}$.

1. Update

$$
\begin{aligned}
\boldsymbol{\Sigma}_{C|D}^{(s-1)} = {} & \mathbf{L}\big(\mathbf{A}^{(s-1)}\big)\boldsymbol{\Omega}^{(s-1)}\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)^T \\
& + \mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big)\boldsymbol{\Omega}_0^{(s-1)}\mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big)^T,
\end{aligned}
$$

$$\mathbf{\Sigma}_{\boldsymbol{Y|X}}^{(s-1)} = \mathbf{R}\big(\mathbf{B}^{(s-1)}\big)\mathbf{\Phi}^{(s-1)}\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T$$
$$+ \mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big)\mathbf{\Phi}_0^{(s-1)}\mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big)^T.$$

2. Generate $\boldsymbol{\beta}_2^{(s)}$ from $\mathcal{MN}_{p_2,r}(\widetilde{\mathbf{M}}^{-1}\widetilde{\mathbf{Z}}^{(s-1)},\widetilde{\mathbf{M}}^{-1},\mathbf{\Sigma}_{\boldsymbol{Y|X}}^{(s-1)})$, where

$$\widetilde{\mathbf{M}} = \widetilde{\mathbb{X}}_2^T\widetilde{\mathbb{X}}_2 + \mathbf{M},$$
$$\widetilde{\mathbf{Z}}^{(s-1)} = \widetilde{\mathbb{X}}_2^T\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T$$
$$\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T\big)$$
$$+ \mathbf{Z}.$$

3. Generate $\boldsymbol{\gamma}^{(s)}$ from $\mathcal{MN}_{p_D,p_C}(\widetilde{\mathbf{\Lambda}}^{-1}\widetilde{\mathbf{F}}^{(s-1)},\widetilde{\mathbf{\Lambda}}^{-1},\mathbf{\Sigma}_{C|D}^{(s-1)})$, where $\widetilde{\mathbf{\Lambda}} = \widetilde{\mathbb{X}}_{1D}^T\widetilde{\mathbb{X}}_{1D} + \mathbf{\Lambda}$, and $\widetilde{\mathbf{F}}^{(s-1)} = \widetilde{\mathbb{X}}_{1D}^T\widetilde{\mathbb{X}}_{1C}^{(s-1)} + \mathbf{F}$.

4. Generate $(\boldsymbol{\mu}_{1C}^{(s)},\boldsymbol{\mu}_{\boldsymbol{Y}}^{(s)})^T$ from $\mathcal{N}_{p_C+r}((\frac{\overline{\boldsymbol{X}_{1C}}}{\overline{\boldsymbol{Y}}}),\mathbf{\Delta}^{(s-1)}/n)$ where $\mathbf{\Delta}^{(s-1)} =$

$$\begin{pmatrix} \mathbf{\Sigma}_{C|D}^{(s-1)} & \mathbf{L}(\mathbf{A}^{(s-1)})\mathbf{\Omega}^{(s-1)}(\boldsymbol{\eta}_C^{(s-1)})^T\mathbf{R}(\mathbf{B}^{(s-1)})^T \\ \mathbf{R}(\mathbf{B}^{(s-1)})\boldsymbol{\eta}_C^{(s-1)}\mathbf{\Omega}^{(s-1)}\mathbf{L}(\mathbf{A}^{(s-1)})^T & \mathbf{\Sigma}_{\boldsymbol{Y|X}}^{(s-1)}+\mathbf{R}(\mathbf{B}^{(s-1)})\boldsymbol{\eta}_C^{(s-1)}\mathbf{\Omega}^{(s-1)}(\boldsymbol{\eta}_C^{(s-1)})^T\mathbf{R}(\mathbf{B}^{(s-1)})^T \end{pmatrix}.$$

5. Generate $\boldsymbol{\eta}_C^{(s)}$ from $\mathcal{MN}_{d_{\boldsymbol{Y}},d_{\boldsymbol{X}}}(\widetilde{\mathbf{W}}^{(s-1)}(\widetilde{\mathbf{E}}^{(s-1)})^{-1},\mathbf{\Phi}^{(s-1)},(\widetilde{\mathbf{E}}^{(s-1)})^{-1})$, where

$$\widetilde{\mathbf{E}}^{(s-1)} = \mathbf{L}\big(\mathbf{A}^{(s-1)}\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)}\big)^T\widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}\big(\mathbf{A}^{(s-1)}\big) + \mathbf{E},$$
$$\widetilde{\mathbf{W}}^{(s-1)} = \big((\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)})\mathbf{R}(\mathbf{B}^{(s-1)}) - \widetilde{\mathbb{X}}_{1D}$$
$$(\boldsymbol{\eta}_D^{(s-1)})^T\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)}\big)\mathbf{L}\big(\mathbf{A}^{(s-1)}\big) + \mathbf{W}.$$

6. Generate $\boldsymbol{\eta}_D^{(s)}$ from $\mathcal{MN}_{d_{\boldsymbol{Y}},p_D}(\widetilde{\mathbf{T}}^{(s-1)}(\widetilde{\mathbf{Q}})^{-1},\mathbf{\Phi}^{(s-1)},(\widetilde{\mathbf{Q}})^{-1})$, where $\widetilde{\mathbf{Q}} = (\widetilde{\mathbb{X}}_{1D})^T\widetilde{\mathbb{X}}_{1D} + \mathbf{Q}$ and $\widetilde{\mathbf{T}}^{(s-1)} = ((\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)})\mathbf{R}(\mathbf{B}^{(s-1)}) - \widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}(\mathbf{A}^{(s-1)})(\boldsymbol{\eta}_C^{(s-1)})^T)^T\widetilde{\mathbb{X}}_{1D} + \mathbf{T}$.

7. Generate $\mathbf{\Omega}^{(s)}$ from $\mathcal{IW}_{d_{\boldsymbol{X}}}(\widetilde{\mathbf{\Psi}}_{\boldsymbol{X}}^{(s-1)},\widetilde{w}_{\boldsymbol{X}})$, where

$$\widetilde{\mathbf{\Psi}}_{\boldsymbol{X}}^{(s-1)} = \mathbf{L}\big(\mathbf{A}^{(s-1)}\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)}-$$
$$\widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)\mathbf{L}\big(\mathbf{A}^{(s-1)}\big) + \mathbf{L}\big(\mathbf{A}^{(s-1)}\big)^T\big(\boldsymbol{\gamma}^{(s-1)}-$$
$$\mathbf{\Lambda}^{-1}\mathbf{F}\big)^T\mathbf{\Lambda}\big(\boldsymbol{\gamma}^{(s-1)} - \mathbf{\Lambda}^{-1}\mathbf{F}\big)\mathbf{L}\big(\mathbf{A}^{(s-1)}\big) + \mathbf{\Psi}_{\boldsymbol{X}}$$

and $\widetilde{w}_{\boldsymbol{X}} = n + p_D + w_{\boldsymbol{X}}$.

8. Generate $\mathbf{\Omega}_0^{(s)}$ from $\mathcal{IW}_{p_C-d_{\boldsymbol{X}}}(\widetilde{\mathbf{\Psi}}_{\boldsymbol{X}_0}^{(s-1)},\widetilde{w}_{\boldsymbol{X}_0})$, where

$$\widetilde{\mathbf{\Psi}}_{\boldsymbol{X}_0}^{(s-1)} = \mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)^T\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)}-$$
$$\widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)\mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big) + \mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big)^T$$
$$\big(\boldsymbol{\gamma}^{(s-1)} - \mathbf{\Lambda}^{-1}\mathbf{F}\big)^T\mathbf{\Lambda}\big(\boldsymbol{\gamma}^{(s-1)} - \mathbf{\Lambda}^{-1}\mathbf{F}\big)$$
$$\mathbf{L}_0\big(\mathbf{A}^{(s-1)}\big) + \mathbf{\Psi}_{\boldsymbol{X}_0}$$

and $\widetilde{w}_{\boldsymbol{X}_0} = n + p_D + w_{\boldsymbol{X}_0}$.

9. Generate $\mathbf{\Phi}^{(s)}$ from $\mathcal{IW}_{d_{\boldsymbol{Y}}}(\widetilde{\mathbf{\Psi}}_{\boldsymbol{Y}}^{(s-1)},\widetilde{w}_{\boldsymbol{Y}})$, where

$$\widetilde{\mathbf{\Psi}}_{\boldsymbol{Y}}^{(s-1)} = \big((\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)})\mathbf{R}\big(\mathbf{B}^{(s-1)}\big) - \widetilde{\mathbb{X}}_{1C}^{(s-1)}$$

$$\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\big)^T$$
$$\big((\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)})\mathbf{R}\big(\mathbf{B}^{(s-1)}\big) - \widetilde{\mathbb{X}}_{1C}^{(s-1)}$$
$$\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\big)+$$
$$\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big)^T\mathbf{M}$$
$$\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big)\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)+$$
$$\big(\boldsymbol{\eta}_C^{(s-1)} - \mathbf{W}\mathbf{E}^{-1}\big)\mathbf{E}\big(\boldsymbol{\eta}_C^{(s-1)} - \mathbf{W}\mathbf{E}^{-1}\big)^T+$$
$$\big(\boldsymbol{\eta}_D^{(s-1)} - \mathbf{T}\mathbf{Q}^{-1}\big)\mathbf{Q}\big(\boldsymbol{\eta}_D^{(s-1)} - \mathbf{T}\mathbf{Q}^{-1}\big)^T + \mathbf{\Psi}_{\boldsymbol{Y}}$$

and $\widetilde{w}_{\boldsymbol{Y}} = n + d_{\boldsymbol{X}} + p_D + p_2 + w_{\boldsymbol{Y}}$.

10. Generate $\mathbf{\Phi}_0^{(s)}$ from $\mathcal{IW}_{r-d_{\boldsymbol{Y}}}(\widetilde{\mathbf{\Psi}}_{\boldsymbol{Y}_0}^{(s-1)},\widetilde{w}_{\boldsymbol{Y}0})$, where

$$\widetilde{\mathbf{\Psi}}_{\boldsymbol{Y}_0}^{(s-1)} = \mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big)^T\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)^T$$
$$\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)\mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big)+$$
$$\mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big)^T\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big)^T$$
$$\mathbf{M}\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big)\mathbf{R}_0\big(\mathbf{B}^{(s-1)}\big) + \mathbf{\Psi}_{\boldsymbol{Y}_0}$$

and $\widetilde{w}_{\boldsymbol{Y}_0} = n + p_2 + w_{\boldsymbol{Y}_0}$.

11. Let $h^{(s)}(\mathbf{A})$ be the log full conditional density of $\mathbf{A}$ at the $s$-th iteration, i.e.,

$$h^{(s)}(\mathbf{A})$$
$$= const - \frac{1}{2}\mathrm{tr}\big\{\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)\big(\mathbf{L}(\mathbf{A})$$
$$\mathbf{\Omega}^{(s-1)}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\mathbf{\Omega}_0^{(s-1)}\mathbf{L}_0(\mathbf{A})^T\big)^{-1}$$
$$\big(\widetilde{\mathbb{X}}_{1C}^{(s-1)} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}^{(s-1)}\big)^T + \big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_{1C}^{(s-1)}$$
$$\mathbf{L}(\mathbf{A})\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T$$
$$\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)\big(\mathbf{\Sigma}_{\boldsymbol{Y|X}}^{(s-1)}\big)^{-1}$$
$$\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}(\mathbf{A})\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T$$
$$- \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\mathbf{R}\big(\mathbf{B}^{(s-1)}\big)^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)^T$$
$$+ \big(\mathbf{L}(\mathbf{A})\mathbf{\Omega}^{(s-1)}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\mathbf{\Omega}_0^{(s-1)}$$
$$\mathbf{L}_0(\mathbf{A})^T\big)^{-1}\big(\boldsymbol{\gamma}^{(s-1)} - \mathbf{\Lambda}^{-1}\mathbf{F}\big)^T\mathbf{\Lambda}\big(\boldsymbol{\gamma}^{(s-1)}-$$
$$\mathbf{\Lambda}^{-1}\mathbf{F}\big) + \mathbf{\Sigma}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0)^T\mathbf{K}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0)\big\}.$$

We want to generate a Markov chain realization for $\mathbf{A}$ from the stationary density proportional to $\exp(h^{(s)}(\mathbf{A}))$. We update $\mathbf{A}^{(s-1)}$ to $\mathbf{A}^{(s)}$ columnwisely via Metropolis steps as follows. Let $\boldsymbol{A}_j^{(s-1)} \in \mathbb{R}^{p_C-d_{\boldsymbol{X}}}$ denote the $j$th column of $\mathbf{A}^{(s-1)}$, $j = 1,\ldots,d_{\boldsymbol{X}}$. Let $\tau_A > 0$ be a given tuning parameter and let $\{i_1,\ldots,i_{d_{\boldsymbol{X}}}\}$ be a random permutation of $\{1,\ldots,d_{\boldsymbol{X}}\}$. For each $k = i_j$, $j = 1,\ldots,d_{\boldsymbol{X}}$,

(a) Generate $\boldsymbol{A}_k^*$ from $\mathcal{N}_{p_C-d_{\boldsymbol{X}}}(\boldsymbol{A}_k^{(s-1)},\tau_A^2\mathbf{I}_{p_C-d_{\boldsymbol{X}}})$.

(b) Calculate the ratio

$$r = \exp\{h(\mathbf{A}^*)\}/\exp\{h(\mathbf{A}^{(s-1)})\},$$

where $\mathbf{A}^*$ is the resulting matrix after $A_k^{(s-1)}$ is replaced with $A_k^*$ in $\mathbf{A}^{(s-1)}$.

(c) Generate a binary indicator with success probability $r$. If a success is achieved, update $A_k^{(s-1)}$ to $A_k^*$ as the $k$-th column of $\mathbf{A}^{(s)}$. Otherwise, retain $A_k^{(s-1)}$ as the $k$-th column of $\mathbf{A}^{(s)}$.

Update $\mathbf{A}^{(s-1)}$ to $\mathbf{A}^{(s)}$ after all columns are investigated.

12. Let $h^{(s)}(\mathbf{B})$ be the log full conditional density of $\mathbf{B}$ at the $s$-th iteration, i.e.,

$$h^{(s)}(\mathbf{B})$$
$$= const - \frac{1}{2}\mathrm{tr}\Big\{\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)$$
$$\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\mathbf{R}(\mathbf{B})^T -$$
$$\widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}^{(s-1)}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0^{(s-1)}$$
$$\mathbf{R}_0(\mathbf{B})^T\big)^{-1}\big(\widetilde{\mathbb{Y}}^{(s-1)} - \widetilde{\mathbb{X}}_{1C}^{(s-1)}\mathbf{L}\big(\mathbf{A}^{(s-1)}\big)\big(\boldsymbol{\eta}_C^{(s-1)}\big)^T$$
$$\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\big(\boldsymbol{\eta}_D^{(s-1)}\big)^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2^{(s-1)}\big)^T +$$
$$\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}^{(s-1)}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0^{(s-1)}\mathbf{R}_0(\mathbf{B})^T\big)^{-1}$$
$$\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big)^T\mathbf{M}\big(\boldsymbol{\beta}_2^{(s-1)} - \mathbf{M}^{-1}\mathbf{Z}\big) + \boldsymbol{\Sigma}_{\mathbf{B}}^{-1}$$
$$(\mathbf{B} - \mathbf{B}_0)^T\mathbf{K}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)\Big\}.$$

The procedure to generate a Markov chain realization for $\mathbf{B}$ from the stationary density proportional to $\exp(h^{(s)}(\mathbf{B}))$ is similar as that for $\mathbf{A}$ in Step 11.

**Remark 1.** *Note that when $d_{\boldsymbol{X}} = 0$ (or, $d_{\boldsymbol{X}} = p_C$), steps 5, 7 and 11 (or, steps 8 and 11) should be skipped, $\mathbf{L}^{(s)}$ does not exist, $\mathbf{L}_0^{(s)} = \mathbf{I}_{p_C}$ (or, $\mathbf{L}^{(s)} = \mathbf{I}_{p_C}$ and $\mathbf{L}_0^{(s)}$ does not exist) always and any terms involving either $\mathbf{L}^{(s)}$ or $\boldsymbol{\Omega}^{(s)}$ or $\boldsymbol{\eta}_C^{(s)}$ (or, either $\mathbf{L}_0^{(s)}$ or $\boldsymbol{\Omega}_0^{(s)}$) or $\mathbf{A}^{(s)}$ are omitted. Similarly, when $d_{\boldsymbol{Y}} = 0$ (or, $d_{\boldsymbol{Y}} = r$), steps 5, 6, 9 and 12 (or, steps 10 and 12) should be skipped, $\mathbf{R}^{(s)}$ does not exist, $\mathbf{R}_0^{(s)} = \mathbf{I}_r$ (or, $\mathbf{R}^{(s)} = \mathbf{I}_r$ and $\mathbf{R}_0^{(s)}$ does not exist) always and any terms involving either $\mathbf{R}^{(s)}$ or $\boldsymbol{\Phi}^{(s)}$ or $\boldsymbol{\eta}_C^{(s)}$ or $\boldsymbol{\eta}_D^{(s)}$ (or, either $\mathbf{R}_0^{(s)}$ or $\boldsymbol{\Phi}_0^{(s)}$) or $\mathbf{B}^{(s)}$ are omitted.*

**Remark 2.** *In our actual implementation of this Metropolis-within-Gibbs algorithm, once any parameter is updated, it will be immediately used for the updating of other parameters in the same iteration. Even when we are updating $\mathbf{A}$ and $\mathbf{B}$, the previously updated columns will be used immediately to update the rest of columns.*

**Remark 3.** *The updating of each column of $\mathbf{A}$ and $\mathbf{B}$ in steps 11 and 12 is implemented in a random order as illustrated. Instead, it could be in a deterministic order as well. Meanwhile, the order of updating different parameter blocks in an iteration could also be random without affecting the convergence.*

## APPENDIX B

### B.1 Initial Estimator for MCMC Algorithm of SIMP

In this section, we propose a warm start initial estimator for all parameters in SIMP, including $\boldsymbol{\mu}_{1C}$, $\boldsymbol{\mu}_{\boldsymbol{Y}}$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}$, $\boldsymbol{\eta}_C$, $\boldsymbol{\eta}_D$, $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Omega}$, $\boldsymbol{\Omega}_0$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}_0$ if all existed (i.e. $0 < d_{\boldsymbol{X}} < p_C$, $0 < d_{\boldsymbol{Y}} < r$ and $r$, $p_C$, $p_D$, $p_2 > 0$). It will be used as the initial value $\boldsymbol{\Theta}^{(0)}$ for our implementation of the MCMC algorithm for SIMP, for faster convergence. Given the data matrices $\mathbb{X}_{1C}$, $\mathbb{X}_{1D}$, $\mathbb{X}_2$ and $\mathbb{Y}$, and any valid envelope dimensions $d_{\boldsymbol{X}}$, $d_{\boldsymbol{Y}}$ as input, we implement the following steps.

1. Estimate $\boldsymbol{\mu}_{\boldsymbol{Y}}^{(0)} = \mathbb{Y}^T\mathbf{1}_n/n$ and $\boldsymbol{\mu}_{1C}^{(0)} = \mathbb{X}_{1C}^T\mathbf{1}_n/n$.
2. Center $\mathbb{X}_{1C}$, $\mathbb{X}_{1D}$, $\mathbb{X}_2$ and $\mathbb{Y}$ to be $\widetilde{\mathbb{X}}_{1C} = \mathbb{X}_{1C} - \mathbf{1}_n(\boldsymbol{\mu}_{1C}^{(0)})^T$, $\widetilde{\mathbb{X}}_{1D} = \mathbb{X}_{1D} - \mathbf{1}_n\overline{\boldsymbol{X}}_{1D}^T$, $\widetilde{\mathbb{X}}_2 = \mathbb{X}_2 - \mathbf{1}_n\overline{\boldsymbol{X}}_2^T$ and $\widetilde{\mathbb{Y}} = \mathbb{Y} - \mathbf{1}_n(\boldsymbol{\mu}_{\boldsymbol{Y}}^{(0)})^T$.
3. Perform the frequentist ordinary least squares for $\widetilde{\mathbb{X}}_{1C}$ on $\widetilde{\mathbb{X}}_{1D}$, and get $\boldsymbol{\gamma}^{(0)}$ and $\boldsymbol{\Sigma}_{C|D}^{(0)}$ as the estimate for the regression coefficients matrix and the residual covariance matrix, respectively.
4. Perform the frequentist partial response envelope model with dimension $d_{\boldsymbol{Y}}$ (by the function *penv()* in the R library *Renvlp*) on

$$\text{regressing } \widetilde{\mathbb{Y}} \text{ on } \widetilde{\mathbb{X}}_{1C}, \widetilde{\mathbb{X}}_{1D} \text{ and } \widetilde{\mathbb{X}}_2,$$

where we treat $\widetilde{\mathbb{X}}_{1C}$ and $\widetilde{\mathbb{X}}_{1D}$ together as our predictors of interest in this partial response envelope model. From this fitting, $\boldsymbol{\beta}_{1C}^{(0)}$, $\boldsymbol{\beta}_{1D}^{(0)}$, $\boldsymbol{\beta}_2^{(0)}$, $\mathbf{R}(\mathbf{B}^{(0)})$, $\mathbf{R}_0(\mathbf{B}^{(0)})$, $\boldsymbol{\Phi}^{(0)}$ and $\boldsymbol{\Phi}_0^{(0)}$ could be directly obtained. Compute $\mathbf{B}^{(0)}$ from $\mathbf{R}(\mathbf{B}^{(0)})$ by the trick of re-parameterization, as illustrated in Section 3.2. Estimate $\boldsymbol{\eta}_D$ by $\boldsymbol{\eta}_D^{(0)} = (\boldsymbol{\beta}_{1D}^{(0)}\mathbf{R}(\mathbf{B}^{(0)}))^T$.
5. Perform the frequentist predictor envelope model with dimension $d_{\boldsymbol{X}}$ (by the function *xenv()* in the R library *Renvlp*) on the regression from $\widetilde{\mathbb{Y}}^*$ on $\widetilde{\mathbb{X}}_{1C}^*$, where $\widetilde{\mathbb{Y}}^*$ and $\widetilde{\mathbb{X}}_{1C}^*$ are the residual matrices from the following two frequentist ordinary least squares regressions respectively:

$$\text{Regress } \widetilde{\mathbb{Y}} \text{ on } \widetilde{\mathbb{X}}_{1D} \text{ and } \widetilde{\mathbb{X}}_2,$$
$$\text{Regress } \widetilde{\mathbb{X}}_{1C} \text{ on } \widetilde{\mathbb{X}}_{1D} \text{ and } \widetilde{\mathbb{X}}_2.$$

Estimates $\mathbf{L}(\mathbf{A}^{(0)})$, $\mathbf{L}_0(\mathbf{A}^{(0)})$ could be immediately obtained from this regression. Then, $\boldsymbol{\Omega}^{(0)}$ and $\boldsymbol{\Omega}_0^{(0)}$ could be calculated as $\boldsymbol{\Omega}^{(0)} = \mathbf{L}(\mathbf{A}^{(0)})^T\boldsymbol{\Sigma}_{C|D}^{(0)}\mathbf{L}(\mathbf{A}^{(0)})$ and $\boldsymbol{\Omega}_0^{(0)} = \mathbf{L}_0(\mathbf{A}^{(0)})^T\boldsymbol{\Sigma}_{C|D}^{(0)}\mathbf{L}_0(\mathbf{A}^{(0)})$, and $\mathbf{A}^{(0)}$ could be computed from $\mathbf{L}(\mathbf{A}^{(0)})$ by the idea in Section 3.2 similarly. Estimate $\boldsymbol{\eta}_C^{(0)} = \mathbf{R}(\mathbf{B}^{(0)})^T(\boldsymbol{\beta}_{1C}^{(0)})^T\mathbf{L}(\mathbf{A}^{(0)})$.

Note when $d_{\boldsymbol{X}} = 0$ (or, $d_{\boldsymbol{X}} = p_C$), $\boldsymbol{\Omega}$, $\mathbf{A}$ and $\boldsymbol{\eta}_C$ (or, $\boldsymbol{\Omega}_0$ and $\mathbf{A}$) don't exist. When $d_{\boldsymbol{Y}} = 0$ (or, $d_{\boldsymbol{Y}} = r$), $\boldsymbol{\Phi}$, $\mathbf{B}$,

$\boldsymbol{\eta}_C$ and $\boldsymbol{\eta}_D$ (or, $\boldsymbol{\Phi}_0$ and $\mathbf{B}$) don't exist. Therefore, it is not necessary to initialize them in the corresponding cases.

## APPENDIX C

## C.1 Additional Simulation Results for Section 7.2

Tables C.1 and C.2 show the parameter estimation performance for $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ under exactly the same set-

ting for (M1) in Section 7.2, except now $p_2 = 4$ with $\boldsymbol{\mu}_2 = (2, -1, -3, 5)$ or $p_D = 4$ respectively. We call these two data generating mechanisms as (M5) and (M6). As Table C.1 reveals, MSE for both $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ from our method and our competitors remain roughly the same or become little inflated as $p_2$ increases from 2 to 4, except few cases. Table C.2 suggests the MSE for $\boldsymbol{\beta}_{1D}$ grow linearly with respect to $p_D$, and the MSE for $\boldsymbol{\beta}_{1C}$ remain almost the same for our method and most of our competitors, except few cases. Therefore, SIMP is observed to show more advanta-

*Table C.1.* MSE *comparison for* SIMP *and other 11 estimators of* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M5). The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face.*

| $r$ | $n$ | | OLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.16 | 0.07 | 0.18 | 0.17 | 3.82 | 0.10 | 0.06 | 0.05 | 0.05 | **0.02** | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.60 | 0.26 | 1.57 | 1.33 | 14.70 | 1.16 | 0.20 | 0.58 | 0.17 | 0.25 | 0.18 | **0.10** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.09 | 0.04 | 0.10 | 0.10 | 3.78 | 0.05 | 0.03 | 0.03 | 0.03 | **0.01** | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.35 | 0.15 | 0.67 | 0.64 | 14.47 | 0.58 | 0.11 | 0.22 | 0.10 | 0.14 | 0.10 | **0.06** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.05 | 0.02 | 0.05 | 0.05 | 3.85 | 0.02 | **0.01** | **0.01** | **0.01** | **0.01** | **0.01** | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.17 | 0.07 | 0.17 | 0.21 | 14.98 | 0.19 | 0.05 | 0.07 | 0.05 | 0.06 | 0.05 | **0.03** |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.69 | 0.11 | 0.59 | 0.52 | 22.98 | 0.27 | 0.07 | 0.08 | 0.06 | 0.04 | 0.06 | **0.03** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.46 | 0.34 | 4.29 | 4.08 | 84.50 | 3.31 | 0.24 | 0.54 | 0.18 | 0.90 | 0.18 | **0.12** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.41 | 0.06 | 0.42 | 0.41 | 22.82 | 0.19 | 0.04 | 0.04 | 0.03 | **0.02** | 0.03 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.46 | 0.19 | 1.88 | 1.90 | 83.58 | 1.75 | 0.13 | 0.21 | 0.10 | 0.55 | 0.10 | **0.07** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.20 | 0.03 | 0.21 | 0.20 | 22.76 | 0.09 | 0.02 | 0.02 | 0.02 | **0.01** | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.73 | 0.08 | 0.84 | 1.19 | 83.02 | 1.02 | 0.05 | 0.08 | 0.05 | 0.27 | 0.05 | **0.03** |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | 1.44 | 0.17 | 0.20 | 0.16 | 30.25 | 0.13 | 0.09 | 0.13 | 0.08 | **0.06** | 0.08 | **0.06** |
| | | $\boldsymbol{\beta}_{1D}$ | 4.66 | 0.46 | 2.25 | 2.20 | 100.73 | 2.21 | 0.22 | 0.81 | **0.17** | 1.94 | **0.17** | 0.23 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.85 | 0.09 | 0.17 | 0.14 | 30.13 | 0.11 | 0.05 | 0.06 | 0.04 | 0.04 | 0.05 | **0.03** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.81 | 0.22 | 2.23 | 2.17 | 101.26 | 2.22 | 0.11 | 0.26 | **0.09** | 1.16 | **0.09** | 0.13 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.42 | 0.04 | 0.17 | 0.13 | 30.18 | 0.10 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.38 | 0.10 | 2.18 | 2.10 | 101.38 | 2.11 | 0.05 | 0.10 | 0.05 | 0.57 | 0.05 | **0.04** |

*Table C.2.* MSE *comparison for* SIMP *and other 11 estimators of* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M6). The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face.*

| $r$ | $n$ | | OLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.15 | 0.06 | 0.17 | 0.17 | 3.13 | 0.10 | 0.05 | 0.07 | 0.04 | 0.02 | 0.05 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.18 | 0.50 | 1.50 | 1.43 | 24.57 | 1.59 | 0.36 | 0.97 | 0.35 | 0.44 | 0.35 | **0.15** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.09 | 0.04 | 0.10 | 0.10 | 3.16 | 0.07 | 0.03 | 0.04 | 0.03 | **0.01** | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.71 | 0.29 | 0.84 | 0.82 | 24.43 | 1.06 | 0.21 | 0.51 | 0.20 | 0.25 | 0.20 | **0.09** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.04 | 0.02 | 0.05 | 0.05 | 3.14 | 0.04 | 0.01 | 0.02 | 0.01 | **0.00** | 0.01 | **0.00** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.34 | 0.14 | 0.36 | 0.40 | 23.84 | 0.43 | 0.10 | 0.16 | 0.10 | 0.13 | 0.10 | **0.04** |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.69 | 0.09 | 0.64 | 0.62 | 18.75 | 0.31 | 0.06 | 0.08 | 0.05 | 0.03 | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 4.61 | 0.76 | 3.53 | 3.37 | 122.38 | 2.96 | 0.36 | 0.80 | 0.33 | 1.87 | 0.33 | **0.20** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.41 | 0.05 | 0.40 | 0.36 | 18.62 | 0.20 | 0.03 | 0.04 | 0.03 | 0.02 | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.75 | 0.40 | 2.47 | 1.70 | 122.52 | 1.85 | 0.20 | 0.43 | 0.20 | 1.12 | 0.20 | **0.11** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.20 | 0.03 | 0.21 | 0.18 | 18.43 | 0.11 | 0.02 | 0.02 | 0.01 | 0.01 | 0.01 | **0.00** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.36 | 0.21 | 1.47 | 0.87 | 121.33 | 1.22 | 0.10 | 0.18 | 0.10 | 0.55 | 0.10 | **0.06** |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | 1.45 | 0.16 | 0.38 | 0.34 | 30.92 | 0.31 | 0.07 | 0.12 | 0.06 | 0.06 | 0.06 | **0.04** |
| | | $\boldsymbol{\beta}_{1D}$ | 9.46 | 1.17 | 7.10 | 6.68 | 200.78 | 6.85 | 0.42 | 0.88 | 0.36 | 3.88 | 0.37 | **0.32** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.86 | 0.08 | 0.48 | 0.37 | 30.68 | 0.32 | 0.04 | 0.06 | 0.04 | 0.03 | 0.04 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 5.52 | 0.57 | 6.74 | 6.39 | 198.22 | 4.80 | 0.23 | 0.43 | 0.21 | 2.30 | 0.21 | **0.17** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.42 | 0.03 | 0.45 | 0.38 | 30.49 | 0.21 | 0.02 | 0.03 | 0.02 | 0.02 | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.79 | 0.26 | 3.26 | 2.20 | 198.11 | 2.35 | 0.11 | 0.20 | 0.10 | 1.14 | 0.10 | **0.07** |

*Table C.3. MSE comparison between FPY-env, PX-env, PY-env and SIMP for estimating $\beta_{1C}$ and $\beta_{1D}$ over 500 repetitions, with the true envelope dimensions known under data generating mechanisms (M1)–(M3), (M5) and (M6), and the MSE of the RRR estimator with the true rank known under data generating mechanism (M4). "\*" in the names of methods means that the true envelope dimension/rank is used. The lowest MSE for each combination of $r$ and $n$ under each true data generating mechanism is highlighted in bold face.*

| | $r$ | 3 | | | | | | 8 | | | | | | 15 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | 300 | | 500 | | 1000 | | 300 | | 500 | | 1000 | | 300 | | 500 | | 1000 | |
| | | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ | $\beta_{1C}$ | $\beta_{1D}$ |
| M1 | FPY-env* | 0.05 | 0.16 | 0.03 | 0.09 | 0.01 | 0.05 | 0.05 | 0.18 | 0.03 | 0.11 | 0.01 | 0.05 | 0.08 | 0.15 | 0.04 | 0.09 | 0.02 | 0.04 |
| | PX-env* | **0.01** | 0.22 | **0.01** | 0.14 | **0.00** | 0.07 | 0.03 | 0.92 | 0.02 | 0.57 | 0.01 | 0.27 | 0.06 | 1.87 | 0.04 | 1.12 | 0.02 | 0.55 |
| | PY-env* | 0.05 | 0.16 | 0.03 | 0.09 | 0.01 | 0.05 | 0.05 | 0.18 | 0.03 | 0.11 | 0.01 | 0.05 | 0.08 | 0.15 | 0.04 | 0.09 | 0.02 | 0.04 |
| | SIMP* | **0.01** | **0.07** | **0.01** | **0.04** | **0.00** | **0.02** | **0.02** | **0.09** | **0.01** | **0.05** | **0.00** | **0.03** | **0.04** | **0.09** | **0.03** | **0.05** | **0.01** | **0.03** |
| M2 | FPY-env* | **0.02** | 0.10 | **0.01** | **0.06** | **0.01** | **0.03** | **0.04** | **0.11** | **0.02** | 0.07 | **0.01** | **0.03** | **0.10** | **0.09** | **0.06** | **0.05** | **0.03** | **0.03** |
| | PX-env* | 0.05 | 0.30 | 0.03 | 0.19 | **0.01** | 0.09 | 0.23 | 1.25 | 0.13 | 0.76 | 0.06 | 0.37 | 0.46 | 2.04 | 0.28 | 1.24 | 0.14 | 0.59 |
| | PY-env* | **0.02** | 0.10 | **0.01** | **0.06** | **0.01** | **0.03** | **0.04** | **0.11** | **0.02** | 0.07 | **0.01** | **0.03** | 0.11 | **0.09** | **0.06** | **0.05** | **0.03** | **0.03** |
| | SIMP* | **0.02** | **0.09** | **0.01** | **0.06** | **0.01** | **0.03** | **0.04** | **0.11** | **0.02** | **0.06** | **0.01** | **0.03** | 0.11 | **0.09** | **0.06** | **0.05** | **0.03** | **0.03** |
| M3 | FPY-env* | 0.23 | 0.81 | 0.13 | 0.46 | 0.06 | 0.23 | 0.22 | 0.80 | 0.13 | 0.49 | 0.06 | 0.23 | 0.23 | 0.70 | 0.13 | 0.41 | 0.07 | 0.21 |
| | PX-env* | 0.02 | 0.33 | **0.01** | 0.20 | **0.00** | 0.10 | **0.02** | 0.47 | **0.01** | 0.28 | **0.01** | 0.14 | **0.03** | 0.66 | **0.02** | 0.40 | **0.01** | 0.20 |
| | **PY-env*** | 0.23 | 0.81 | 0.13 | 0.46 | 0.06 | 0.23 | 0.22 | 0.80 | 0.13 | 0.49 | 0.06 | 0.23 | 0.23 | 0.69 | 0.13 | 0.41 | 0.07 | 0.21 |
| | SIMP* | **0.01** | **0.30** | **0.01** | **0.19** | **0.00** | **0.09** | **0.02** | **0.30** | **0.01** | **0.18** | **0.01** | **0.09** | **0.03** | **0.30** | **0.02** | **0.19** | **0.01** | **0.09** |
| M4 | RRR* | 0.39 | 0.05 | 0.23 | 0.03 | 0.11 | 0.01 | 0.32 | 0.04 | 0.17 | 0.02 | 0.08 | 0.01 | 9.04 | 1.00 | 3.75 | 0.40 | 1.56 | 0.16 |
| M5 | FPY-env* | 0.05 | 0.17 | 0.03 | 0.10 | 0.01 | 0.05 | 0.06 | 0.18 | 0.03 | 0.10 | 0.02 | 0.05 | 0.07 | 0.16 | 0.04 | 0.09 | 0.02 | 0.05 |
| | PX-env* | **0.01** | 0.22 | **0.01** | 0.13 | **0.00** | 0.06 | 0.03 | 0.90 | 0.02 | 0.55 | **0.01** | 0.27 | 0.06 | 1.94 | 0.03 | 1.16 | 0.02 | 0.57 |
| | PY-env* | 0.05 | 0.17 | 0.03 | 0.10 | 0.01 | 0.05 | 0.06 | 0.17 | 0.03 | 0.10 | 0.02 | 0.05 | 0.07 | 0.16 | 0.04 | 0.09 | 0.02 | 0.05 |
| | SIMP* | **0.01** | **0.08** | **0.01** | **0.05** | **0.00** | **0.02** | **0.02** | **0.09** | **0.01** | **0.05** | **0.01** | **0.03** | **0.04** | **0.09** | **0.02** | **0.06** | **0.01** | **0.03** |
| M6 | FPY-env* | 0.04 | 0.35 | 0.03 | 0.20 | 0.01 | 0.10 | 0.05 | 0.33 | 0.03 | 0.20 | 0.01 | 0.10 | 0.06 | 0.36 | 0.04 | 0.21 | 0.02 | 0.10 |
| | PX-env* | 0.01 | 0.40 | **0.00** | 0.24 | **0.00** | 0.12 | 0.03 | 1.86 | 0.02 | 1.13 | 0.01 | 0.55 | 0.06 | 3.87 | 0.03 | 2.31 | 0.02 | 1.14 |
| | PY-env* | 0.04 | 0.35 | 0.03 | 0.20 | 0.01 | 0.10 | 0.05 | 0.33 | 0.03 | 0.19 | 0.01 | 0.10 | 0.06 | 0.37 | 0.04 | 0.21 | 0.02 | 0.10 |
| | SIMP* | **0.00** | **0.13** | **0.00** | **0.08** | **0.00** | **0.04** | **0.01** | **0.18** | **0.01** | **0.11** | **0.00** | **0.05** | **0.03** | **0.23** | **0.02** | **0.13** | **0.01** | **0.06** |

geous numerical performance with increased $p_D$ in our limited simulation studies.

Table C.3 summarizes the performance of FPY-env, PX-env, PY-env and SIMP with the true envelope dimensions known under (M1)–(M3), (M5) and (M6), or the performance of RRR with the true rank known under (M4). With the true envelope dimensions known, SIMP almost always obtains the best estimation performance under all scenarios except the model mis-specification case (M4).

Table C.4 shows the performance of all three Bayesian envelope estimators (PX-env, PY-env and SIMP) of $\boldsymbol{\beta}_{1C}$ and $\boldsymbol{\beta}_{1D}$ under the data generating mechanism (M1), except now the posterior median rather than the posterior mean is chosen as the point estimator. Comparing the results in this table and Table 3, we empirically find no observable difference in performance between the posterior mean and the posterior median as the point estimator under (M1).

## APPENDIX D

### D.1 Theoretical Proofs

*Proof of Theorem 1.* The integrability of the posterior density could be immediately obtained following the proof of Theorem 2. □

*Proof of Theorem 2.* To prove the Harris ergodicity of the Markov chain generated by our Metropolis-within-Gibbs algorithm, we need to prove 3 properties, namely, (a) $\phi$-irreducibility with respect to some measure $\phi$, (b) Aperiodicity, and (c) Harris recurrence. Firstly, we prove $\phi$-irreducibility which requires that the Markov chain could reach any set with positive measure from any state in the state space. The rigorous definition is given in Definition 3.

**Definition 3** ($\phi$-irreducibility). *For a Markov chain $(\boldsymbol{X}_n)_{n=1}^{\infty}$, suppose that the state space is $\mathcal{X}$ with $\sigma$-algebra $\mathcal{F}$, and the $n$-step transition kernel is $\kappa_n : \mathcal{X} \times \mathcal{F} \to [0,1]$. Given a nonzero measure $\phi$, the Markov chain is called $\phi$-irreducible, if for every state $\boldsymbol{x} \in \mathcal{X}$ and every subset*

*Table C.4.* MSE *comparison for three Bayesian envelope estimators (PX-env, PY-env and* SIMP*) on estimating* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M1), except the posterior median rather than the posterior mean is chosen as the point estimator. The lowest* MSE *under each combination of* $r$ *and* $n$ *is in bold face. The tiny differences between the results in this table and those in Table 3 are colored in red.*

| $r$ | $n$ | | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | **0.02** | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.23 | 0.17 | **0.09** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | **0.01** | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.14 | 0.10 | **0.05** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | **0.00** | 0.01 | **0.00** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.07 | 0.05 | **0.02** |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.03 | 0.05 | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.93 | 0.18 | **0.10** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.02 | 0.03 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.57 | 0.11 | **0.06** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | **0.01** | 0.02 | **0.01** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.27 | 0.05 | **0.03** |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | **0.06** | 0.08 | **0.06** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.87 | **0.15** | 0.17 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.04 | 0.04 | **0.03** |
| | | $\boldsymbol{\beta}_{1D}$ | 1.13 | **0.09** | 0.10 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | **0.02** | **0.02** | **0.02** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.54 | 0.04 | **0.03** |

$A \in \mathcal{F}$ with $\phi(A) > 0$, there exists an integer $n$ such that $\kappa_n(\boldsymbol{x}, A) > 0$.

To prove the $\phi$-irreducibility of the Markov chain generated by the Metropolis-within-Gibbs algorithm that we developed, we just need to check that the proposal densities and the acceptance probabilities for updating all parameter blocks are positive, then the $\phi$-irreducibility is proved under 1-step transition (take $n = 1$). For 1-dimensional MCMC, it is obvious that,

$$\kappa_1(x, A) = \int_A q(x, y)a(x, y)\phi(\mathrm{d}y) > 0$$
$$\text{for } \forall \, x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})$$

as long as the proposal density $q(x, y) > 0$ and the acceptance probability $a(x, y) > 0$ always. To extend this conclusion to the multi-dimensional case, the technique of induction could be applied. See details in Section 4.1.8 of [22]. Among 12 parameter blocks $\{\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_Y, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D\}$, if all existed, the updating of $\mathbf{A}$ and $\mathbf{B}$ requires the Metropolis steps, and that for the rest of parameter blocks only needs normal Gibbs steps by the conjugacy. Since the exponential function and the vector Normal density function are always positive, we can conclude that the acceptance probabilities and the

proposal densities of the Metropolis steps, if present, are always positive. For the Gibbs steps, the proposal distributions are Matrix normal and Inverse-Wishart distributions, both with densities always positive. The acceptance probabilities are always 1, which is positive as well. This completes the proof for the $\phi$-irreducibility.

Next, we intend to show the Aperiodicity, whose definition for a $\phi$-irreducible Markov chain is given in Definition 4.

**Definition 4** (Aperiodicity)**.** *For a $\phi$-irreducible Markov chain $(\boldsymbol{X}_n)_{n=1}^{\infty}$, suppose that the state space is $\mathcal{X}$ with $\sigma$-algebra $\mathcal{F}$, the (1-step) transition probability is $\kappa_1 : \mathcal{X} \times \mathcal{F} \rightarrow [0, 1]$, and the stationary probability distribution is $\pi(\cdot)$. The period of $(\boldsymbol{X}_n)_{n=1}^{\infty}$ is defined as the largest positive integer $T$ for which there exist disjoint subsets $A_1, A_2, \ldots, A_T \in \mathcal{F}$ with $\pi(A_i) > 0$, such that $\kappa_1(\boldsymbol{x}, A_{i+1}) = 1$ for all $\boldsymbol{x} \in A_i$ ($1 \leq i \leq T - 1$), and $\kappa_1(\boldsymbol{x}, A_1) = 1$ for all $\boldsymbol{x} \in A_T$. The Markov chain is called Aperiodic if $T = 1$.*

Aperiodicity could be immediately obtained, since every set with nonzero probability (under the stationary probability distribution $\pi$) in the $\sigma$-algebra of the state space could be reached from any point in the state space through one step transition in the Markov chain.

Lastly, we intend to prove the Harris recurrence. The definition of the Harris recurrence for a $\phi$-irreducible Markov chain is given in Definition 5.

**Definition 5** (Harris recurrence)**.** *For a $\phi$-irreducible Markov chain $(\boldsymbol{X}_n)_{n=1}^{\infty}$, suppose that the state space is $\mathcal{X}$ with $\sigma$-algebra $\mathcal{F}$, and the stationary probability distribution is $\pi(\cdot)$. The Markov chain is called Harris recurrent, if for all $A \subseteq \mathcal{X}$ with $\pi(A) > 0$ and any starting point $\boldsymbol{x} \in \mathcal{X}$, we have the conditional probability $P(\boldsymbol{X}_n \in A$ infinitely often in $n \mid \boldsymbol{X}_1 = \boldsymbol{x}) = 1$.*

When $d_{\boldsymbol{X}} \in \{0, p_C\}$ and $d_{\boldsymbol{Y}} \in \{0, r\}$, $\mathbf{A}$ and $\mathbf{B}$ don't exist, then the Metropolis-within-Gibbs algorithm is actually a normal Gibbs sampler. Then Harris recurrence is directly implied by $\phi$-irreducibility ([48], Corollary 13). When either $0 < d_{\boldsymbol{X}} < p_C$ or $0 < d_{\boldsymbol{Y}} < r$, there exists Metropolis step for either $\mathbf{A}$ or $\mathbf{B}$. The Harris recurrence could be obtained by proving the Lebesgue integrability of the posterior density over every combination of parameter blocks ([48], Corollary 19). For simplicity, we assume $0 < d_{\boldsymbol{X}} < p_C$ and $0 < d_{\boldsymbol{Y}} < r$. Hence we only need to prove the posterior density $f(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_Y, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$ is Lebesgue integrable over every $k$ ($1 \leq k \leq 12$) elements from $\{\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_Y, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D\}$. Theorem 2 under the other two cases (1) $d_{\boldsymbol{X}} \in \{0, p_C\}$ and $0 < d_{\boldsymbol{Y}} < r$; (2) $0 < d_{\boldsymbol{X}} < p_C$ and $d_{\boldsymbol{Y}} \in \{0, r\}$ could be similarly proved with simple modifications of the following proof. Theorem 1 should be immediately proved by letting $k$ to be the maximal value 12, i.e., the Lebesgue integrability

*Table C.5.* MSE *comparison between* SIMP *and other 11 competitors for estimating* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *over* 500 *repetitions under the data generating mechanism (M4). Different from Table 6, this table displays the estimation performance of 12 methods, if we pretend that the true values of the parameters* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *are known, and the "optimal" tuning parameters or the envelope dimensions are selected by minimizing the* $\ell_2$ *distances between the estimates and the true values of* $\boldsymbol{\beta}_{1C}$ *and* $\boldsymbol{\beta}_{1D}$ *separately, instead of applying the model selection methods that are introduced in Section 7.2.1. The lowest* MSE *under each combination of* $r$ *and* $n$ *among all methods and within envelope methods are in bold face and blue respectively.*

| $r$ | $n$ | | FOLS | RRR | PCR | PLSR | CCA | FX-env | FY-env | FS-env | FPY-env | PX-env | PY-env | SIMP |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.92 | 0.39 | 0.51 | 0.39 | 27.61 | 0.37 | 0.65 | **0.35** | 0.64 | 0.37 | 0.64 | 0.37 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.12 | 0.05 | 0.06 | 0.04 | 3.58 | **0.02** | 0.05 | **0.02** | 0.05 | 0.10 | 0.09 | 0.04 |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.53 | **0.23** | 0.36 | 0.26 | 27.60 | 0.27 | 0.47 | 0.26 | 0.47 | 0.27 | 0.47 | 0.27 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.07 | 0.03 | 0.04 | 0.03 | 3.72 | **0.01** | 0.04 | **0.01** | 0.04 | 0.07 | 0.07 | 0.04 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.27 | **0.11** | 0.23 | 0.15 | 28.57 | 0.19 | 0.26 | 0.18 | 0.26 | 0.18 | 0.26 | 0.18 |
| | | $\boldsymbol{\beta}_{1D}$ | 0.03 | 0.01 | 0.02 | 0.01 | 3.37 | **0.00** | 0.03 | **0.00** | 0.03 | 0.03 | 0.03 | 0.02 |
| 8 | 300 | $\boldsymbol{\beta}_{1C}$ | 0.75 | 0.32 | 0.65 | 0.49 | 20.53 | 0.58 | 0.32 | 0.26 | 0.28 | 0.48 | 0.28 | **0.24** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.09 | 0.04 | 0.08 | 0.05 | 2.49 | 0.05 | 0.04 | **0.02** | 0.03 | 0.08 | 0.03 | **0.02** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 0.43 | 0.17 | 0.39 | 0.28 | 20.32 | 0.35 | 0.18 | 0.15 | 0.15 | 0.28 | 0.15 | **0.13** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.06 | 0.02 | 0.05 | 0.03 | 2.62 | 0.03 | 0.02 | **0.01** | 0.02 | 0.05 | 0.02 | 0.02 |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 0.20 | 0.08 | 0.20 | 0.13 | 20.14 | 0.16 | 0.08 | **0.07** | **0.07** | 0.14 | **0.07** | **0.07** |
| | | $\boldsymbol{\beta}_{1D}$ | 0.03 | 0.01 | 0.02 | 0.01 | 2.55 | 0.01 | 0.01 | **0.00** | 0.01 | 0.02 | 0.01 | 0.01 |
| 15 | 300 | $\boldsymbol{\beta}_{1C}$ | 37.13 | 9.04 | 9.76 | 8.61 | 20.23 | 1.33 | 2.84 | 1.06 | 2.82 | 0.82 | 2.82 | **0.71** |
| | | $\boldsymbol{\beta}_{1D}$ | 4.74 | 1.00 | 0.01 | 0.01 | 2.10 | 0.01 | 0.01 | **0.00** | **0.00** | 4.32 | 0.12 | **0.00** |
| | 500 | $\boldsymbol{\beta}_{1C}$ | 22.73 | 3.75 | 8.37 | 6.43 | 20.95 | 0.62 | 2.39 | 0.57 | 2.37 | 0.41 | 2.37 | **0.40** |
| | | $\boldsymbol{\beta}_{1D}$ | 2.85 | 0.40 | 0.01 | 0.01 | 2.05 | **0.00** | **0.00** | **0.00** | **0.00** | 2.62 | 0.08 | **0.00** |
| | 1000 | $\boldsymbol{\beta}_{1C}$ | 11.54 | 1.56 | 6.38 | 4.47 | 21.22 | 0.31 | 2.06 | 0.30 | 2.05 | **0.23** | 2.05 | 0.24 |
| | | $\boldsymbol{\beta}_{1D}$ | 1.46 | 0.16 | 0.01 | 0.01 | 2.08 | **0.00** | **0.00** | **0.00** | **0.00** | 1.39 | 0.07 | **0.00** |

over all parameter blocks is considered. Below we show the proof. The unnormalized posterior density is

$$
\begin{aligned}
& f(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_Y, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}) \\
& \propto f_0(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_Y, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}) \\
& = \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ (\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \right.\right. \\
& \quad \left.\left. \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T \right)^{-1}(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T \right\}\right\} \\
& \quad \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ (\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T \right.\right. \\
& \quad \mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2)(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0 \\
& \quad \mathbf{R}_0(\mathbf{B})^T)^{-1}(\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T \\
& \quad \left.\left. \mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2)^T \right\}\right\} |\boldsymbol{\Omega}|^{-(n+p_D+w_X+d_X+1)/2} \\
& \quad |\boldsymbol{\Omega}_0|^{-(n+p_D+w_{X_0}+p_C-d_X+1)/2} \\
& \quad |\boldsymbol{\Phi}|^{-(n+p_2+w_Y+d_Y+d_X+p_D+1)/2} \\
& \quad |\boldsymbol{\Phi}_0|^{-(n+p_2+w_{Y_0}+r-d_Y+1)/2} \\
& \quad \exp\left\{ -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}_X\boldsymbol{\Omega}^{-1}) \right\} \exp\left\{ -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}_{X_0}\boldsymbol{\Omega}_0^{-1}) \right\} \\
& \quad \exp\left\{ -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}_Y\boldsymbol{\Phi}^{-1}) \right\} \exp\left\{ -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Psi}_{Y_0}\boldsymbol{\Phi}_0^{-1}) \right\}
\end{aligned}
$$

$$
\begin{aligned}
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \boldsymbol{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F})^T\boldsymbol{\Lambda}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F}) \right\}\right\} \\
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \boldsymbol{\Sigma}_{Y|X}^{-1}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z})^T\mathbf{M}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z}) \right\}\right\} \\
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \boldsymbol{\Sigma}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)^T\mathbf{K}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0) \right\}\right\} \\
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \boldsymbol{\Sigma}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)^T\mathbf{K}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0) \right\}\right\} \\
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \mathbf{E}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})^T\boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1}) \right\}\right\} \\
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ \mathbf{Q}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})^T\boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1}) \right\}\right\}.
\end{aligned}
$$

Specifically, in the following term, what inside trace function is a product of two positive semi-definite matrices, so

$$
\begin{aligned}
& \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ (\mathbb{Y} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \right.\right. \\
& \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2)(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0\mathbf{R}_0(\mathbf{B})^T)^{-1}(\mathbb{Y} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A}) \\
& \left.\left. \boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2)^T\left( \mathbf{I}_n - \frac{1}{n}\mathbf{J}_n \right) \right\}\right\} \leq 1,
\end{aligned}
$$

where $\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n^T$. Therefore,

$$
\int \exp\left\{ -\frac{1}{2}\operatorname{tr}\left\{ (\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_Y^T - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \right.\right.
$$

$$\widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0$$

$$\mathbf{R}_0(\mathbf{B})^T\big)^{-1}\big(\mathbb{Y} - \mathbf{1}_n\boldsymbol{\mu}_{\boldsymbol{Y}}^T - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}$$

$$\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)^T\big\}\big\}\,\mathrm{d}\boldsymbol{\mu}_{\boldsymbol{Y}}$$

$$= \left(\frac{2\pi}{n}\right)^{r/2}|\boldsymbol{\Phi}|^{1/2}|\boldsymbol{\Phi}_0|^{1/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big\{\big(\mathbb{Y} - \widetilde{\mathbb{X}}_{1C}$$

$$\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}$$

$$\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0\mathbf{R}_0(\mathbf{B})^T\big)^{-1}\big(\mathbb{Y} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T$$

$$\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)^T\Big(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\Big)\Big\}\Big\}$$

$$\leq \left(\frac{2\pi}{n}\right)^{r/2}|\boldsymbol{\Phi}|^{1/2}|\boldsymbol{\Phi}_0|^{1/2}.$$

Meanwhile,

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big\{\big(\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T -$$

$$\widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)\big(\mathbf{R}(\mathbf{B})\boldsymbol{\Phi}\mathbf{R}(\mathbf{B})^T + \mathbf{R}_0(\mathbf{B})\boldsymbol{\Phi}_0\mathbf{R}_0(\mathbf{B})^T\big)^{-1}\big(\widetilde{\mathbb{Y}} - \widetilde{\mathbb{X}}_{1C}$$

$$\mathbf{L}(\mathbf{A})\boldsymbol{\eta}_C^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\eta}_D^T\mathbf{R}(\mathbf{B})^T - \widetilde{\mathbb{X}}_2\boldsymbol{\beta}_2\big)^T\Big\}\Big\} \leq 1.$$

We can conclude

$$f_0(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$$
$$\leq c_1 f_1^{(0)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

and

$$\int f_0(\boldsymbol{\mu}_{1C}, \boldsymbol{\mu}_{\boldsymbol{Y}}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})\,\mathrm{d}\boldsymbol{\mu}_{\boldsymbol{Y}}$$
$$\leq c_1 f_1^{(1)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

where $c_1 = 1 + (\frac{2\pi}{n})^{r/2}$ and for $\Xi = 0, 1$,

$$f_1^{(\Xi)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$$
$$= \exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})$$

$$\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T\big)^{-1}(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T\big\}\Big\}$$

$$|\boldsymbol{\Omega}|^{-(n+p_D+w_{\boldsymbol{X}}+d_{\boldsymbol{X}}+1)/2}$$
$$|\boldsymbol{\Omega}_0|^{-(n+p_D+w_{\boldsymbol{X}_0}+p_C-d_{\boldsymbol{X}}+1)/2}$$
$$|\boldsymbol{\Phi}|^{-(n+p_2+w_{\boldsymbol{Y}}+d_{\boldsymbol{Y}}+d_{\boldsymbol{X}}+p_D+1-\Xi)/2}$$
$$|\boldsymbol{\Phi}_0|^{-(n+p_2+w_{\boldsymbol{Y}_0}+r-d_{\boldsymbol{Y}}+1-\Xi)/2}$$
$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Psi}_{\boldsymbol{X}}\boldsymbol{\Omega}^{-1}\big)\Big\}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Psi}_{\boldsymbol{X}_0}\boldsymbol{\Omega}_0^{-1}\big)\Big\}$$
$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Psi}_{\boldsymbol{Y}}\boldsymbol{\Phi}^{-1}\big)\Big\}\exp\Big\{-\frac{1}{2}\mathrm{tr}\big(\boldsymbol{\Psi}_{\boldsymbol{Y}_0}\boldsymbol{\Phi}_0^{-1}\big)\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\boldsymbol{\Sigma}_{C|D}^{-1}\big(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F}\big)^T\boldsymbol{\Lambda}\big(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F}\big)\big\}\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}\big(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z}\big)^T\mathbf{M}\big(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z}\big)\big\}\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\boldsymbol{\Sigma}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)^T\mathbf{K}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)\big\}\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\boldsymbol{\Sigma}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)^T\mathbf{K}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)\big\}\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\mathbf{E}\big(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1}\big)^T\boldsymbol{\Phi}^{-1}\big(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1}\big)\big\}\Big\}$$

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\mathbf{Q}\big(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1}\big)^T\boldsymbol{\Phi}^{-1}\big(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1}\big)\big\}\Big\}.$$

Next, we just need to show $f_1^{(\Xi)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$ is integrable with respect to any $1 \leq k \leq 11$ elements in $\{\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D\}$, for any value of $\Xi$. Similarly, note that

$$\int \exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{\big(\mathbb{X}_{1C} - \mathbf{1}_n\boldsymbol{\mu}_{1C}^T - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}\big)\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}$$

$$\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T\big)^{-1}\big(\mathbb{X}_{1C} - \mathbf{1}_n\boldsymbol{\mu}_{1C}^T -$$

$$\widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}\big)^T\big\}\Big\}\,\mathrm{d}\boldsymbol{\mu}_{1C}$$

$$= \left(\frac{2\pi}{n}\right)^{p_C/2}|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Omega}_0|^{1/2}\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big\{\big(\mathbb{X}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}\big)$$

$$\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0\mathbf{L}_0(\mathbf{A})^T\big)^{-1}(\mathbb{X}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T$$

$$\Big(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\Big)\Big\}\Big\}$$

$$\leq \left(\frac{2\pi}{n}\right)^{p_C/2}|\boldsymbol{\Omega}|^{1/2}|\boldsymbol{\Omega}_0|^{1/2},$$

since

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\Big\{\big(\mathbb{X}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma}\big)\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0$$

$$\mathbf{L}_0(\mathbf{A})^T\big)^{-1}(\mathbb{X}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T\Big(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n\Big)\Big\}\Big\} \leq 1.$$

Together with the fact that

$$\exp\Big\{-\frac{1}{2}\mathrm{tr}\big\{(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})\big(\mathbf{L}(\mathbf{A})\boldsymbol{\Omega}\mathbf{L}(\mathbf{A})^T + \mathbf{L}_0(\mathbf{A})\boldsymbol{\Omega}_0$$

$$\mathbf{L}_0(\mathbf{A})^T\big)^{-1}(\widetilde{\mathbb{X}}_{1C} - \widetilde{\mathbb{X}}_{1D}\boldsymbol{\gamma})^T\big\}\Big\} \leq 1,$$

we get for any $\Xi = 0, 1$,

$$f_1^{(\Xi)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$$
$$\leq c_2 f_2^{(\Xi,0)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

and

$$\int f_1^{(\Xi)}(\boldsymbol{\mu}_{1C}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}) \, \mathrm{d}\boldsymbol{\mu}_{1C}$$

$$\leq c_2 f_2^{(\Xi,1)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

where $c_2 = 1 + (\frac{2\pi}{n})^{p_C/2}$ and for $\Xi = 0, 1$, $\zeta = 0, 1$,

$$f_2^{(\Xi,\zeta)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$$

$$= |\boldsymbol{\Omega}|^{-(n+p_D+w_{\boldsymbol{X}}+d_{\boldsymbol{X}}+1-\zeta)/2}$$

$$|\boldsymbol{\Omega}_0|^{-(n+p_D+w_{\boldsymbol{X}_0}+p_C-d_{\boldsymbol{X}}+1-\zeta)/2}$$

$$|\boldsymbol{\Phi}|^{-(n+p_2+w_{\boldsymbol{Y}}+d_{\boldsymbol{Y}}+d_{\boldsymbol{X}}+p_D+1-\Xi)/2}$$

$$|\boldsymbol{\Phi}_0|^{-(n+p_2+w_{\boldsymbol{Y}_0}+r-d_{\boldsymbol{Y}}+1-\Xi)/2}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{X}}\boldsymbol{\Omega}^{-1})\right\} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{X}_0}\boldsymbol{\Omega}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{Y}}\boldsymbol{\Phi}^{-1})\right\} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{Y}_0}\boldsymbol{\Phi}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F})^T \boldsymbol{\Lambda}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F})\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z})^T \mathbf{M}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z})\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)^T \mathbf{K}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)^T \mathbf{K}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{E}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{Q}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})\}\right\}.$$

Then it is enough to show $f_2^{(\Xi,\zeta)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$ is integrable over every $1 \leq k \leq 10$ members in $\{\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D\}$, for any value of $\Xi$ and $\zeta$. Observe

$$\int \exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{E}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})\}\right\} \mathrm{d}\boldsymbol{\eta}_C$$

$$= (2\pi)^{d_{\boldsymbol{X}} d_{\boldsymbol{Y}}/2} |\mathbf{E}|^{-d_{\boldsymbol{Y}}/2} |\boldsymbol{\Phi}|^{d_{\boldsymbol{X}}/2}$$

and

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{E}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_C - \mathbf{W}\mathbf{E}^{-1})\}\right\} \leq 1.$$

Hence for any $\Xi = 0, 1$, $\zeta = 0, 1$,

$$f_2^{(\Xi,\zeta)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D})$$

$$\leq c_3 f_3^{(\Xi,\zeta,0)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

and

$$\int f_2^{(\Xi,\zeta)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_C, \boldsymbol{\eta}_D \mid \mathcal{D}) \, \mathrm{d}\boldsymbol{\eta}_C$$

$$\leq c_3 f_3^{(\Xi,\zeta,1)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D}),$$

with $c_3 = 1 + (2\pi)^{d_{\boldsymbol{X}} d_{\boldsymbol{Y}}/2} |\mathbf{E}|^{-d_{\boldsymbol{Y}}/2}$. For $\Xi$, $\zeta$, $\lambda = 0, 1$,

$$f_3^{(\Xi,\zeta,\lambda)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D})$$

$$= |\boldsymbol{\Omega}|^{-(n+p_D+w_{\boldsymbol{X}}+d_{\boldsymbol{X}}+1-\zeta)/2}$$

$$|\boldsymbol{\Omega}_0|^{-(n+p_D+w_{\boldsymbol{X}_0}+p_C-d_{\boldsymbol{X}}+1-\zeta)/2}$$

$$|\boldsymbol{\Phi}|^{-(n+p_2+w_{\boldsymbol{Y}}+d_{\boldsymbol{Y}}+p_D+(1-\lambda)d_{\boldsymbol{X}}+1-\Xi)/2}$$

$$|\boldsymbol{\Phi}_0|^{-(n+p_2+w_{\boldsymbol{Y}_0}+r-d_{\boldsymbol{Y}}+1-\Xi)/2}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{X}}\boldsymbol{\Omega}^{-1})\right\} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{X}_0}\boldsymbol{\Omega}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{Y}}\boldsymbol{\Phi}^{-1})\right\} \exp\left\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{\Psi}_{\boldsymbol{Y}_0}\boldsymbol{\Phi}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F})^T \boldsymbol{\Lambda}(\boldsymbol{\gamma} - \boldsymbol{\Lambda}^{-1}\mathbf{F})\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\boldsymbol{Y}|\boldsymbol{X}}^{-1}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z})^T \mathbf{M}(\boldsymbol{\beta}_2 - \mathbf{M}^{-1}\mathbf{Z})\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)^T \mathbf{K}_{\mathbf{A}}^{-1}(\mathbf{A} - \mathbf{A}_0)\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\boldsymbol{\Sigma}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)^T \mathbf{K}_{\mathbf{B}}^{-1}(\mathbf{B} - \mathbf{B}_0)\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{Q}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})\}\right\}.$$

Therefore, the task right now is to prove $f_3^{(\Xi,\zeta,\lambda)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D})$ is integrable over any $1 \leq k \leq 9$ elements in $\{\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D\}$, for any value of $\Xi$, $\zeta$ and $\lambda$. Since

$$\int \exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{Q}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})\}\right\} \mathrm{d}\boldsymbol{\eta}_D$$

$$= (2\pi)^{d_{\boldsymbol{Y}} p_D/2} |\mathbf{Q}|^{-d_{\boldsymbol{Y}}/2} |\boldsymbol{\Phi}|^{p_D/2}$$

and

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\{\mathbf{Q}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})^T \boldsymbol{\Phi}^{-1}(\boldsymbol{\eta}_D - \mathbf{T}\mathbf{Q}^{-1})\}\right\} \leq 1,$$

we get for any $\Xi$, $\zeta$, $\lambda = 0, 1$,

$$f_3^{(\Xi,\zeta,\lambda)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D})$$

$$\leq c_4 f_4^{(\Xi,\zeta,\lambda,0)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B} \mid \mathcal{D}),$$

and

$$\int f_3^{(\Xi,\zeta,\lambda)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B}, \boldsymbol{\eta}_D \mid \mathcal{D}) \, \mathrm{d}\boldsymbol{\eta}_D$$

$$\leq c_4 f_4^{(\Xi,\zeta,\lambda,1)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B} \mid \mathcal{D}),$$

with $c_4 = 1 + (2\pi)^{d_{\boldsymbol{Y}} p_D/2} |\mathbf{Q}|^{-d_{\boldsymbol{Y}}/2}$ and for $\Xi$, $\zeta$, $\lambda$, $\tau = 0, 1$,

$$f_4^{(\Xi,\zeta,\lambda,\tau)}(\boldsymbol{\beta}_2, \boldsymbol{\gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega}_0, \boldsymbol{\Phi}, \boldsymbol{\Phi}_0, \mathbf{A}, \mathbf{B} \mid \mathcal{D})$$

$$= |\boldsymbol{\Omega}|^{-(n+p_D+w_{\boldsymbol{X}}+d_{\boldsymbol{X}}+1-\zeta)/2}$$

$|\mathbf{\Omega}_0|^{-(n+p_D+w_{\mathbf{X}_0}+p_C-d_{\mathbf{X}}+1-\zeta)/2}$

$|\mathbf{\Phi}|^{-(n+p_2+w_{\mathbf{Y}}+d_{\mathbf{Y}}+(1-\tau)p_D+(1-\lambda)d_{\mathbf{X}}+1-\Xi)/2}$

$|\mathbf{\Phi}_0|^{-(n+p_2+w_{\mathbf{Y}_0}+r-d_{\mathbf{Y}}+1-\Xi)/2}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_X}\mathbf{\Omega}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{X_0}}\mathbf{\Omega}_0^{-1})\right\}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_Y}\mathbf{\Phi}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{Y_0}}\mathbf{\Phi}_0^{-1})\right\}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})^T\mathbf{\Lambda}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})\right\}\right\}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})^T\mathbf{M}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})\right\}\right\}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_A}^{-1}(\mathbf{A}-\mathbf{A}_0)^T\mathbf{K_A}^{-1}(\mathbf{A}-\mathbf{A}_0)\right\}\right\}$

$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_B}^{-1}(\mathbf{B}-\mathbf{B}_0)^T\mathbf{K_B}^{-1}(\mathbf{B}-\mathbf{B}_0)\right\}\right\}.$

Then we intend to show $f_4^{(\Xi,\zeta,\lambda,\tau)}(\boldsymbol{\beta}_2,\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},$ $\mathbf{B}\mid\mathcal{D})$ is integrable over any $1\leq k\leq 8$ elements in $\{\boldsymbol{\beta}_2,\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\}$, for any value of $\Xi$, $\zeta$, $\lambda$ and $\tau$. Since

$$\int\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})^T\mathbf{M}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})\right\}\right\}\mathrm{d}\boldsymbol{\beta}_2$$

$$=(2\pi)^{p_2r/2}|\mathbf{M}|^{-r/2}|\mathbf{\Phi}|^{p_2/2}|\mathbf{\Phi}_0|^{p_2/2},$$

and

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})^T\mathbf{M}(\boldsymbol{\beta}_2-\mathbf{M}^{-1}\mathbf{Z})\right\}\right\}\leq 1,$$

we get for any $\Xi$, $\zeta$, $\lambda$, $\tau=0,1$,

$$f_4^{(\Xi,\zeta,\lambda,\tau)}(\boldsymbol{\beta}_2,\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})$$
$$\leq c_5 f_5^{(\Xi,\zeta,\lambda,\tau,0)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D}),$$

and

$$\int f_4^{(\Xi,\zeta,\lambda,\tau)}(\boldsymbol{\beta}_2,\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})\mathrm{d}\boldsymbol{\beta}_2$$

$$\leq c_5 f_5^{(\Xi,\zeta,\lambda,\tau,1)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D}),$$

where $c_5=1+(2\pi)^{p_2r/2}|\mathbf{M}|^{-r/2}$ and for $\Xi$, $\zeta$, $\lambda$, $\tau$, $\chi=0,1$,

$$f_5^{(\Xi,\zeta,\lambda,\tau,\chi)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})$$
$$=|\mathbf{\Omega}|^{-(n+p_D+w_{\mathbf{X}}+d_{\mathbf{X}}+1-\zeta)/2}$$
$$|\mathbf{\Omega}_0|^{-(n+p_D+w_{\mathbf{X}_0}+p_C-d_{\mathbf{X}}+1-\zeta)/2}$$
$$|\mathbf{\Phi}|^{-(n+w_{\mathbf{Y}}+d_{\mathbf{Y}}+(1-\chi)p_2+(1-\tau)p_D+(1-\lambda)d_{\mathbf{X}}+1-\Xi)/2}$$
$$|\mathbf{\Phi}_0|^{-(n+w_{\mathbf{Y}_0}+r-d_{\mathbf{Y}}+(1-\chi)p_2+1-\Xi)/2}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_X}\mathbf{\Omega}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{X_0}}\mathbf{\Omega}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_Y}\mathbf{\Phi}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{Y_0}}\mathbf{\Phi}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})^T\mathbf{\Lambda}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})\right\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_A}^{-1}(\mathbf{A}-\mathbf{A}_0)^T\mathbf{K_A}^{-1}(\mathbf{A}-\mathbf{A}_0)\right\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_B}^{-1}(\mathbf{B}-\mathbf{B}_0)^T\mathbf{K_B}^{-1}(\mathbf{B}-\mathbf{B}_0)\right\}\right\}.$$

Now it is suffice to show $f_5^{(\Xi,\zeta,\lambda,\tau,\chi)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})$ is integrable with respect to any $1\leq k\leq 7$ members in $\{\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\}$, for any value of $\Xi$, $\zeta$, $\lambda$, $\tau$ and $\chi$. Applying the similar trick as before, since

$$\int\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})^T\mathbf{\Lambda}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})\right\}\right\}\mathrm{d}\boldsymbol{\gamma}$$

$$=(2\pi)^{p_Dp_C/2}|\mathbf{\Lambda}|^{-p_C/2}|\mathbf{\Omega}|^{p_D/2}|\mathbf{\Omega}_0|^{p_D/2},$$

and

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma}_{C|D}^{-1}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})^T\mathbf{\Lambda}(\boldsymbol{\gamma}-\mathbf{\Lambda}^{-1}\mathbf{F})\right\}\right\}\leq 1,$$

we conclude for any $\Xi$, $\zeta$, $\lambda$, $\tau$, $\chi=0,1$,

$$f_5^{(\Xi,\zeta,\lambda,\tau,\chi)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})$$
$$\leq c_6 f_6^{(\Xi,\zeta,\lambda,\tau,\chi,0)}(\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D}),$$

and

$$\int f_5^{(\Xi,\zeta,\lambda,\tau,\chi)}(\boldsymbol{\gamma},\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})\mathrm{d}\boldsymbol{\gamma}$$

$$\leq c_6 f_6^{(\Xi,\zeta,\lambda,\tau,\chi,1)}(\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D}),$$

with $c_6=1+(2\pi)^{p_Dp_C/2}|\mathbf{\Lambda}|^{-p_C/2}$ and for $\Xi$, $\zeta$, $\lambda$, $\tau$, $\chi$, $\nu=0,1$,

$$f_6^{(\Xi,\zeta,\lambda,\tau,\chi,\nu)}(\mathbf{\Omega},\mathbf{\Omega}_0,\mathbf{\Phi},\mathbf{\Phi}_0,\mathbf{A},\mathbf{B}\mid\mathcal{D})$$
$$=|\mathbf{\Omega}|^{-(n+w_{\mathbf{X}}+d_{\mathbf{X}}+p_D(1-\nu)+1-\zeta)/2}$$
$$|\mathbf{\Omega}_0|^{-(n+w_{\mathbf{X}_0}+p_C-d_{\mathbf{X}}+p_D(1-\nu)+1-\zeta)/2}$$
$$|\mathbf{\Phi}|^{-(n+w_{\mathbf{Y}}+d_{\mathbf{Y}}+(1-\chi)p_2+(1-\tau)p_D+(1-\lambda)d_{\mathbf{X}}+1-\Xi)/2}$$
$$|\mathbf{\Phi}_0|^{-(n+w_{\mathbf{Y}_0}+r-d_{\mathbf{Y}}+(1-\chi)p_2+1-\Xi)/2}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_X}\mathbf{\Omega}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{X_0}}\mathbf{\Omega}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_Y}\mathbf{\Phi}^{-1})\right\}\exp\left\{-\frac{1}{2}\mathrm{tr}(\mathbf{\Psi_{Y_0}}\mathbf{\Phi}_0^{-1})\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_A}^{-1}(\mathbf{A}-\mathbf{A}_0)^T\mathbf{K_A}^{-1}(\mathbf{A}-\mathbf{A}_0)\right\}\right\}$$

$$\exp\left\{-\frac{1}{2}\mathrm{tr}\left\{\mathbf{\Sigma_B}^{-1}(\mathbf{B}-\mathbf{B}_0)^T\mathbf{K_B}^{-1}(\mathbf{B}-\mathbf{B}_0)\right\}\right\}.$$

Notice this is a constant multiple of the product of densities for $\mathcal{IW}_{d_{\mathbf{X}}}(\mathbf{\Psi_X},n+w_{\mathbf{X}}+p_D(1-\nu)-\zeta)$, $\mathcal{IW}_{p_C-d_{\mathbf{X}}}(\mathbf{\Psi_{X_0}},n+$

$w_{\boldsymbol{X}_0} + p_D(1 - \nu) - \zeta)$, $\mathcal{IW}_{d_{\boldsymbol{Y}}}(\boldsymbol{\Psi}_{\boldsymbol{Y}}, n + w_{\boldsymbol{Y}} + p_2(1 - \chi) + p_D(1 - \tau) + d_{\boldsymbol{X}}(1 - \lambda) - \Xi)$, $\mathcal{IW}_{r-d_{\boldsymbol{Y}}}(\boldsymbol{\Psi}_{\boldsymbol{Y}_0}, n + w_{\boldsymbol{Y}_0} + p_2(1 - \chi) - \Xi)$, $\mathcal{MN}_{p_C-d_{\boldsymbol{X}}.d_{\boldsymbol{X}}}(\mathbf{A}_0, \mathbf{K_A}, \boldsymbol{\Sigma_A})$ and $\mathcal{MN}_{r-d_{\boldsymbol{Y}}.d_{\boldsymbol{Y}}}(\mathbf{B}_0, \mathbf{K_B}, \boldsymbol{\Sigma_B})$. These Inverse-Wishart distributions are well-defined by checking

$$n + w_{\boldsymbol{X}} + p_D(1 - \nu) - \zeta \;>\; d_{\boldsymbol{X}} - 1,$$

$$n + w_{\boldsymbol{X}_0} + p_D(1 - \nu) - \zeta \;>\; p_C - d_{\boldsymbol{X}} - 1,$$

$$n + w_{\boldsymbol{Y}_0} + p_2(1 - \chi) - \Xi \;>\; r - d_{\boldsymbol{Y}} - 1,$$

and

$$n + w_{\boldsymbol{Y}} + p_2(1 - \chi) + p_D(1 - \tau) + d_{\boldsymbol{X}}(1 - \lambda) - \Xi > d_{\boldsymbol{Y}} - 1,$$

for any value of $\Xi, \zeta, \lambda, \tau, \chi, \nu = 0, 1$, since $w_{\boldsymbol{X}} > d_{\boldsymbol{X}} - 1$, $w_{\boldsymbol{X}_0} > p_C - d_{\boldsymbol{X}} - 1$, $w_{\boldsymbol{Y}} > d_{\boldsymbol{Y}} - 1$, $w_{\boldsymbol{Y}_0} > r - d_{\boldsymbol{Y}} - 1$, $p_D$, $p_2$, $d_{\boldsymbol{X}} \geq 0$ and $n \geq 1$. This completes the proof. $\square$

# APPENDIX E

## E.1 Review of the Matrix Normal and the Inverse-Wishart Distributions

We introduce the Matrix normal and the Inverse-Wishart distributions here, for the description of the prior and the posterior distributions:

Let $\mathbf{X} \sim \mathcal{MN}_{a,b}(\mathbf{M}, \mathbf{U}, \mathbf{V})$ denote the Matrix normal distribution with location parameter $\mathbf{M} \in \mathbb{R}^{a \times b}$ and scale parameters $\mathbf{U} \in \mathbb{R}^{a \times a}$ and $\mathbf{V} \in \mathbb{R}^{b \times b}$ such that $\mathbf{U}$ and $\mathbf{V}$ are positive definite, and it has the density function

$$\frac{\exp\{-\mathrm{tr}(\mathbf{V}^{-1}(\mathbf{X} - \mathbf{M})^T \mathbf{U}^{-1}(\mathbf{X} - \mathbf{M}))/2\}}{(2\pi)^{ab/2} |\mathbf{V}|^{a/2} |\mathbf{U}|^{b/2}}.$$

Let $\mathbf{X} \sim \mathcal{IW}_a(\mathbf{X}_0, \nu)$ denote the Inverse-Wishart distribution with scale matrix parameter $\mathbf{X}_0 \in \mathbb{S}_+^{a \times a}$ and degrees of freedom parameter $\nu > a - 1$. The density function is given by

$$\frac{|\mathbf{X}_0|^{\nu/2}}{2^{\nu a/2} \Gamma_a(\nu/2)} |\mathbf{X}|^{-(\nu+a+1)/2} \exp\{-\mathrm{tr}(\mathbf{X}_0 \mathbf{X}^{-1})/2\}.$$

Here $\Gamma_a(\cdot)$ denotes the Multivariate gamma function, which is defined by

$$\Gamma_a(u) = \pi^{a(a-1)/4} \prod_{j=1}^{a} \Gamma\{u + (1-j)/2\}, \quad \text{for } u > 0,$$

where $a$ is a positive integer, and $\Gamma\{\cdot\}$ is the usual Gamma function.

# APPENDIX F

## F.1 Review of DIC, WAIC and Bayesian CV

We introduce the definitions of two information criteria, DIC and WAIC, and Bayesian CV here. Let $\widehat{\boldsymbol{\Theta}}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{mean}$ denote the estimated posterior mean of all parameters $\boldsymbol{\Theta}$ among retained posterior samples, when fixing the envelope dimensions at $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$. Recall that in (4.1), we denote $l(\boldsymbol{\Theta})$ as the log-likelihood function of $\boldsymbol{\Theta}$ conditional on all observations $\mathcal{D}$. For convenience, for $i = 1, 2, \ldots, n$, we further denote $f(\mathcal{D}_i \mid \boldsymbol{\Theta}) = \exp\{l_i(\boldsymbol{\Theta})\}$ as the density function value of the observation $i$ when using parameter $\boldsymbol{\Theta}$, and its log-transformation is denoted as $l_i(\boldsymbol{\Theta})$, where $\mathcal{D}_i = \{\boldsymbol{Y}_i, \boldsymbol{X}_{1C,i}, \boldsymbol{X}_{1D,i}, \boldsymbol{X}_{2,i}\}$. Then DIC and WAIC are defined as

$$\mathrm{DIC} = -2l(\widehat{\boldsymbol{\Theta}}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{mean})$$
$$+ 2\left\{ l(\widehat{\boldsymbol{\Theta}}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{mean}) - \frac{1}{S} \sum_{s=1}^{S} l(\boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{(s)}) \right\},$$

$$\mathrm{WAIC} = -2 \sum_{i=1}^{n} \log\left\{ \frac{1}{S} \sum_{s=1}^{S} f(\mathcal{D}_i \mid \boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{(s)}) \right\}$$
$$+ 2 \sum_{i=1}^{n} \left\{ \frac{1}{S-1} \sum_{s=1}^{S} \left( l_i(\boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{(s)}) \right. \right.$$
$$\left. \left. - \frac{1}{S} \sum_{s=1}^{S} l_i(\boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{(s)}) \right)^2 \right\},$$

where $\{\boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{(s)}\}_{s=1}^{S}$ are the retained posterior samples, and we use the subscript for $\boldsymbol{\Theta}$ to highlight that the envelope dimensions are fixed at $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$. Compared with AIC$-$MCMC as defined in (6.1), DIC replaces $\widehat{\boldsymbol{\Theta}}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{max}$ and $K(d_{\boldsymbol{X}}, d_{\boldsymbol{Y}})$ with $\widehat{\boldsymbol{\Theta}}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{mean}$ and a bias correction term based on data respectively, while WAIC replaces the maximized log-likelihood function by a term called the computed log pointwise posterior predictive density, and estimates $K(d_{\boldsymbol{X}}, d_{\boldsymbol{Y}})$ by another bias correction term.

To introduce Bayesian CV, suppose $n$ samples are partitioned into $K$ subsets, with corresponding index sets as $\{\mathcal{C}_k\}_{k=1}^{K}$. For fixed $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$, and each $k = 1, 2, \ldots, K$, we run the MCMC algorithm on all observations $\mathcal{D}$ except those in $\mathcal{C}_k$, and denote the retained posterior samples as $\{\boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{-k,(s)}\}_{s=1}^{S}$. Instead of computing the MSPE that is usually used in the frequentist setting, Bayesian CV aims at finding $d_{\boldsymbol{X}}$ and $d_{\boldsymbol{Y}}$ that minimize the minus of the average out-of-sample estimate of the log predictive density, with expression

$$-\frac{1}{K} \sum_{k=1}^{K} \log\left\{ \frac{1}{S} \sum_{s=1}^{S} \prod_{i \in \mathcal{C}_k} f(\mathcal{D}_i \mid \boldsymbol{\Theta}_{d_{\boldsymbol{X}}, d_{\boldsymbol{Y}}}^{-k,(s)}) \right\}.$$

# APPENDIX G

## G.1 MCMC Diagnostic and Posterior Density Plots for Section 7.2

Figure G.1 shows the convergence of the proposed MCMC algorithm by showing the traceplots, the autocorrelation plots and the evolution of the Gelman-Rubin's shrink factor for a randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$, $\widehat{\boldsymbol{\beta}}_{1D}$, $\widehat{\boldsymbol{\beta}}_2$, $\widehat{\boldsymbol{\Sigma}}_{C|D}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}|\boldsymbol{X}}$ from SIMP, under the data generating mechanism (M1) of Section 7.2, with $r = 15$ and $n = 300$ (the most challenging case). All of those plots indicate the convergence and weak autocorrelation between MCMC samples.

Figure G.2 displays the empirical posterior density plots of the randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$ and $\widehat{\boldsymbol{\beta}}_{1D}$ under the same setting. These (estimated) posterior distributions are bell-shaped and the shape is very common among the (estimated) posterior distributions of all elements in $\widehat{\boldsymbol{\beta}}_{1C}$ and $\widehat{\boldsymbol{\beta}}_{1D}$.



Figure G.2: The (marginal) empirical posterior density plots for a randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$ and $\widehat{\boldsymbol{\beta}}_{1D}$ from SIMP, under the data generating mechanism (M1) of Section 7.2, with $r = 15$ and $n = 300$. These density plots under other simulated settings are similar.

# APPENDIX H

## H.1 Plots of Exploring the Population Stratification for Section 8



Figure G.1: The trace plots (left), the autocorrelation plots (middle; based only on Chain 1) and the evolution of the (median and 97.5% upper confidence limit of) Gelman-Rubin's shrink factor as the number of iterations increases (right; PSRF represents the point estimate, i.e. the potential scale reduction factor) for a randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$, $\widehat{\boldsymbol{\beta}}_{1D}$, $\widehat{\boldsymbol{\beta}}_2$, $\widehat{\boldsymbol{\Sigma}}_{C|D}$ and $\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{Y}|\boldsymbol{X}}$ from SIMP (correspondingly from row 1 to row 5), under the data generating mechanism (M1) of Section 7.2, with $r = 15$ and $n = 300$ (the most challenging case).



Figure H.1: Plots of all selected participants with their $i$-th SNP PC and $(i+1)$-th SNP PC as $X$ and $Y$ coordinates respectively, ($i = 1, 3, 5, 7, 9, 11$ for the upper left, upper middle, upper right, lower left, lower middle and lower right panels. The percentage of the total variation that is explained by that PC is included in the parenthesis) for the real data application. No population stratification is observed in all six plots.

## H.2 MCMC Diagnostic and Posterior Density Plots for Section 8.4
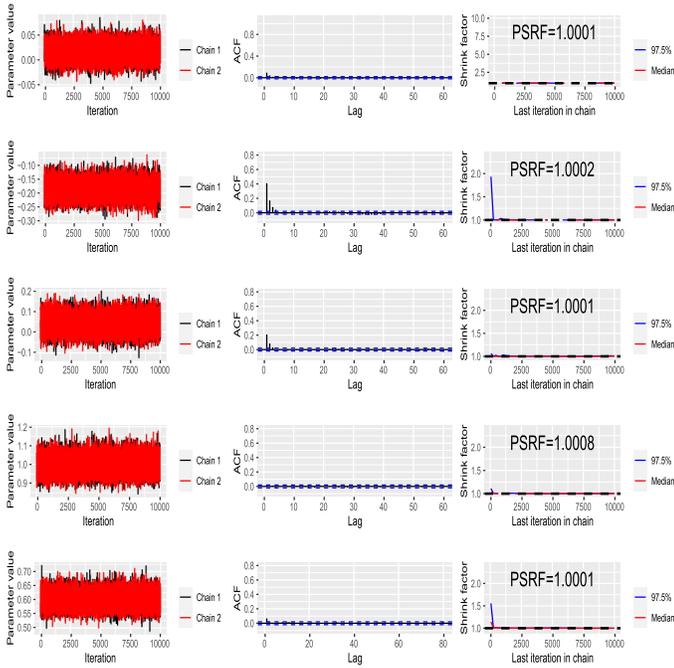


Figure H.2: The trace plots (left), the autocorrelation plots (middle; based only on Chain 1) and the evolution of the (median and 97.5% upper confidence limit of) Gelman-Rubin's shrink factor as the number of iterations increases (right; PSRF represents the point estimate, i.e. the potential scale reduction factor) for a randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$, $\widehat{\boldsymbol{\beta}}_{1D}$, $\widehat{\boldsymbol{\beta}}_{2}$, $\widehat{\boldsymbol{\Sigma}}_{C|D}$ and $\widehat{\boldsymbol{\Sigma}}_{Y|X}$ from SIMP (correspondingly from row 1 to row 5) for the real data application.
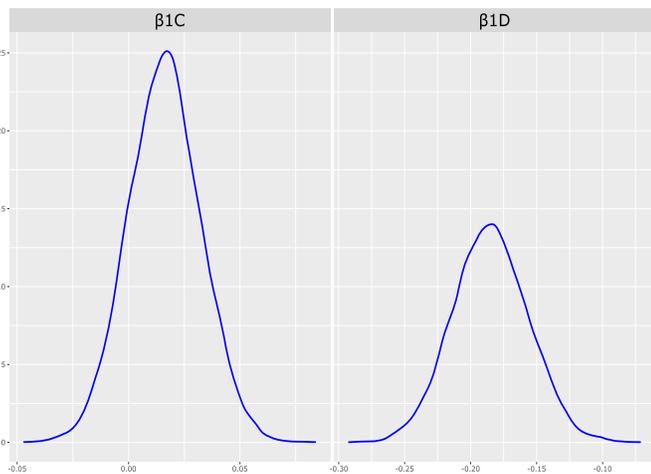


Figure H.3: The (marginal) empirical posterior density plots for a randomly selected and representative element of each of $\widehat{\boldsymbol{\beta}}_{1C}$ and $\widehat{\boldsymbol{\beta}}_{1D}$ for the real data application.

Figures H.2 and H.3 are the MCMC diagnostic plots and the (marginal) empirical posterior density plots for the real data application. The observations are similar to those in Appendix G.

## H.3 37 Significant SNPs Besides APOE $\epsilon 4$ in Section 8.4

Table H.1. Besides APOE $\epsilon 4$, the following 37 SNPs are found to be significant with all 12 IPs from SIMP. The RefSeq gene is the associated (or the nearest if known) gene for the SNP in the NCBI Reference Sequence Database, where "-" means that no such associated gene is found by us.

| SNP | RefSeq gene | SNP | RefSeq gene |
|---|---|---|---|
| rs10493854 | - | rs7120548 | MTCH2 |
| rs12059556 | - | rs10501927 | CNTN5 |
| rs1229119 | - | rs489243 | - |
| rs6428182 | - | rs10894473 | NTM |
| rs10921394 | - | rs11064498 | C1S |
| rs12138394 | - | rs10877700 | - |
| rs12756397 | - | rs757402 | OAS2 |
| rs11685593 | LOC105373605/BIN1 | rs845757 | SPATA7 |
| rs7561528 | LOC105373605/BIN1 | rs2274736 | PTPN21 |
| rs11706690 | CHL1 | rs10498633 | SLC24A4/RIN3 |
| rs10513391 | P2RY14 | rs2554389 | ADAMTSL3 |
| rs1623177 | - | rs437649 | LOC105371501 |
| rs10515893 | - | rs470268 | - |
| rs2439538 | TBC1D7 | rs17809911 | CCDC102B |
| rs9381563 | CD2AP | rs9959820 | - |
| rs1516895 | - | rs4809760 | SLC9A8 |
| rs11038268 | LOC105376650 | rs2281223 | SLC9A8 |
| rs2280231 | NDUFS3 | | |

## H.4 Boxplot of Bilateral Correlations for Section 8.4
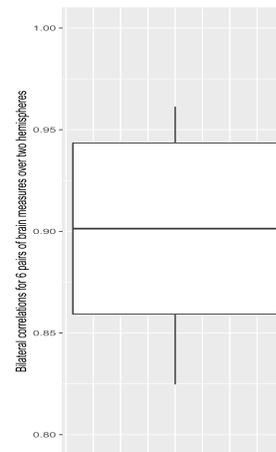


Figure H.4: Boxplot of the Pearson correlation coefficients within all 6 pairs of brain measures over two hemispheres. Strong bilateral correlations are observed in this figure.

# ACKNOWLEDGEMENTS

# REFERENCES

[1] BARNARD, J., MCCULLOCH, R. and MENG, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* 1281–1311.

[2] BOADA, M., ANTUNEZ, C., RAMÍREZ-LORCA, R., DESTEFANO, A. L., GONZALEZ-PEREZ, A., GAYÁN, J., LÓPEZ-ARRIETA, J., IKRAM, M. A., HERNÁNDEZ, I., MARIN, J. et al. (2014). ATP5H/KCTD2 locus is associated with Alzheimer's disease risk. *Molecular Psychiatry* **19**(6) 682–687.

[3] BROCE, I. J., TAN, C. H., FAN, C. C., JANSEN, I., SAVAGE, J. E., WITOELAR, A., WEN, N., HESS, C. P., DILLON, W. P., GLASTONBURY, C. M. et al. (2019). Dissecting the genetic relationship between cardiovascular risk factors and Alzheimer's disease. *Acta Neuropathologica* **137**(2) 209–226.

[4] BURA, E. and COOK, R. D. (2003). Rank estimation in reduced-rank regression. *Journal of Multivariate Analysis* **87**(1) 159–176.

[5] CHAKRABORTY, S. and SU, Z. A comprehensive Bayesian framework for envelope models. *Technical Report*.

[6] CHAMBERLAIN, G. (1982). Multivariate regression models for panel data. *Journal of Econometrics* **18**(1) 5–46.

[7] CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. and LEE, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**(1) 13742–015.

[8] CHEN, T., SU, Z., YANG, Y., DING, S. et al. (2020). Efficient estimation in expectile regression using envelope models. *Electronic Journal of Statistics* **14**(1) 143–173.

[9] CHIB, S. and GREENBERG, E. (1998). Analysis of multivariate probit models. *Biometrika* **85**(2) 347–361.

[10] CONWAY, J. (1990) *A course in functional analysis.* New York: Springer.

[11] COOK, R. D. and ZHANG, X. (2015). Foundations for envelope models and methods. *Journal of the American Statistical Association* **110**(510) 599–611.

[12] COOK, R. D., HELLAND, I. S. and SU, Z. (2013). Envelopes and partial least squares regression. *Journal of the Royal Statistical Society: Series B* **75**(5) 851–877.

[13] COOK, R. D. (2018). An introduction to envelopes: Dimension reduction for efficient estimation in multivariate statistics.

[14] COOK, R. D. and ZHANG, X. (2015). Simultaneous envelopes for multivariate linear regression. *Technometrics* **57**(1) 11–25.

[15] COOK, R. D., LI, B. and CHIAROMONTE, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression (with discussion). *Statistica Sinica* **20** 927–1010.

[16] DE JONG, S. (1993). SIMPLS: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems* **18**(3) 251–263.

[17] DIBBLE, C. C., ELIS, W., MENON, S., QIN, W., KLEKOTA, J., ASARA, J. M., FINAN, P. M., KWIATKOWSKI, D. J., MURPHY, L. O. and MANNING, B. D. (2012). TBC1D7 is a third subunit of the TSC1-TSC2 complex upstream of mTORC1. *Molecular Cell* **47**(4) 535–546.

[18] DING, S. and DENNIS COOK, R. (2018). Matrix variate regressions and envelope models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**(2) 387–408.

[19] DING, S., SU, Z., ZHU, G. and WANG, L. (2020). Envelope quantile regression. *Statistica Sinica* **31**(1) 79–105.

[20] DUBOIS, B., HAMPEL, H., FELDMAN, H. H., SCHELTENS, P., AISEN, P., ANDRIEU, S., BAKARDJIAN, H., BENALI, H., BERTRAM, L., BLENNOW, K. et al. (2016). Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia* **12**(3) 292–323.

[21] FISCHL, B. (2012). FreeSurfer. *Neuroimage* **62**(2) 774–781.

[22] GEYER, C. J. (1998). Markov chain Monte Carlo lecture notes. *Course notes, Spring Quarter* **80**.

[23] GREENLAW, K., SZEFER, E., GRAHAM, J., LESPERANCE, M., NATHOO, F. S. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2017). A Bayesian group sparse multi-task regression model

for imaging genetics. *Bioinformatics* **33**(16) 2513–2522.

[24] HAROLD, D., ABRAHAM, R., HOLLINGWORTH, P., SIMS, R., GER-RISH, A., HAMSHERE, M. L., PAHWA, J. S., MOSKVINA, V., DOWZELL, K., WILLIAMS, A. et al. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics* **41**(10) 1088–1093.

[25] HIBAR, D. P., STEIN, J. L., KOHANNIM, O., JAHANSHAD, N., SAYKIN, A. J., SHEN, L., KIM, S., PANKRATZ, N., FOROUD, T., HUENTELMAN, M. J. et al. (2011). Voxelwise gene-wide association study (vGeneWAS): Multivariate gene-based association testing in 731 elderly subjects. *NeuroImage* **56**(4) 1875–1891.

[26] HOTELLING, H. (1936). Relations between two sets of variates. *Biometrika* **28**(3-4) 321–377.

[27] HOWIE, B. N., DONNELLY, P. and MARCHINI, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLOS Genetics* **5**(6) 1000529.

[28] HOWIE, B., FUCHSBERGER, C., STEPHENS, M., MARCHINI, J. and ABECASIS, G. R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nature Genetics* **44**(8) 955–959.

[29] IZENMAN, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis* **5**(2) 248–264.

[30] JANSEN, I. E., SAVAGE, J. E., WATANABE, K., BRYOIS, J., WILLIAMS, D. M., STEINBERG, S., SEALOCK, J., KARLSSON, I. K., HÄGG, S., ATHANASIU, L. et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nature Genetics* **51**(3) 404–413.

[31] KARCH, C. M., EZERSKIY, L. A., BERTELSEN, S., (ADGC), A. D. G. C. and GOATE, A. M. (2016). Alzheimer's disease risk polymorphisms regulate gene expression in the ZCWPW1 and the CELF1 loci. *PLOS One* **11**(2) 0148717.

[32] KENDALL, M. G. (1957). A course in multivariate analysis. Technical Report.

[33] KHARE, K., PAL, S. and SU, Z. (2017). A Bayesian approach for envelope models. *The Annals of Statistics* **45**(1) 196–222.

[34] KIM, S., SOHN, K.-A. and XING, E. P. (2009). A multivariate regression approach to association analysis of a quantitative trait network. *Bioinformatics* **25**(12) 204–212.

[35] KUMAR, N., BANSAL, A., SARMA, G. and RAWAL, R. K. (2014). Chemometrics tools used in analytical chemistry: An overview. *Talanta* **123** 186–199.

[36] LAMBERT, J.-C., IBRAHIM-VERBAAS, C. A., HAROLD, D., NAJ, A. C., SIMS, R., BELLENGUEZ, C., JUN, G., DESTEFANO, A. L., BIS, J. C., BEECHAM, G. W. et al. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics* **45**(12) 1452–1458.

[37] LEE, M. and SU, Z. (2020). A review of envelope models. *International Statistical Review* **88**(3) 658–676.

[38] LEE, M., CHAKRABORTY, S. and SU, Z. (2022). A Bayesian approach to envelope quantile regression. *Statistica Sinica* **32** 1–19.

[39] LI, L. and ZHANG, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association* **112**(519) 1131–1146.

[40] LIU, J. S. and WU, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**(448) 1264–1274.

[41] NAJ, A. C., LEONENKO, G., JIAN, X., GRENIER-BOLEY, B., DAL-MASSO, M. C., BELLENGUEZ, C., SHA, J., ZHAO, Y., VAN DER LEE, S. J., SIMS, R. et al. (2021). Genome-wide meta-analysis of late-onset Alzheimer's disease using rare variant imputation in 65,602 subjects identifies novel rare variant locus NCK2: The International Genomics of Alzheimer's Project (IGAP). *medRxiv*.

[42] NATHOO, F. S., KONG, L., ZHU, H. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2019). A review of statistical methods in imaging genetics. *Canadian Journal of Statistics* **47**(1) 108–131.

[43] NESTOR, S. M., RUPSINGH, R., BORRIE, M., SMITH, M., AC-COMAZZI, V., WELLS, J. L., FOGARTY, J., BARTHA, R. and

ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2008). Ventricular enlargement as a possible measure of Alzheimer's disease progression validated using the Alzheimer's disease neuroimaging initiative database. *Brain* **131**(9) 2443–2454.

[44] NOOROSSANA, R., EYVAZIAN, M., AMIRI, A. and MAHMOUD, M. A. (2010). Statistical monitoring of multivariate multiple linear regression profiles in phase I with calibration application. *Quality and Reliability Engineering International* **26**(3) 291–303.

[45] PARK, Y., SU, Z. and CHUNG, D. (2022). Envelope-based partial partial least squares with application to cytokine-based biomarker analysis for COVID-19. *Statistics in Medicine* 1–15.

[46] PARK, Y., SU, Z. and ZHU, H. (2017). Groupwise envelope models for imaging genetic analysis. *Biometrics* **73**(4) 1243–1253.

[47] POULIN, S. P., DAUTOFF, R., MORRIS, J. C., BARRETT, L. F., DICKERSON, B. C. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2011). Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Research: Neuroimaging* **194**(1) 7–13.

[48] ROBERTS, G. O. and ROSENTHAL, J. S. (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *The Annals of Applied Probability* **16**(4) 2123–2139.

[49] ROSENTHAL, S. L., BARMADA, M. M., WANG, X., DEMIRCI, F. Y. and KAMBOH, M. I. (2014). Connecting the dots: Potential of data integration to identify regulatory SNPs in late-onset Alzheimer's disease GWAS findings. *PLOS One* **9**(4) 95152.

[50] ROWE, B. Bayesian variable selection methods for genome wide association studies with categorical phenotypes (2020). PhD thesis, University of Nevada, Las Vegas.

[51] SELIGE, T., BÖHNER, J. and SCHMIDHALTER, U. (2006). High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures. *Geoderma* **136**(1-2) 235–244.

[52] SONG, Y., GE, S., CAO, J., WANG, L. and NATHOO, F. S. (2022). A Bayesian spatial model for imaging genetics. *Biometrics* **78**(2) 742–753.

[53] SQUILLARIO, M., TOMASI, F., TOZZO, V., BARLA, A., UBERTI, D. and ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE (2018). A 3-fold kernel approach for characterizing late-onset Alzheimer's disease. *bioRxiv* 397760.

[54] STEIN, J. L., HUA, X., LEE, S., HO, A. J., LEOW, A. D., TOGA, A. W., SAYKIN, A. J., SHEN, L., FOROUD, T., PANKRATZ, N. et al. (2010). Voxelwise genome-wide association study (vGWAS). *NeuroImage* **53**(3) 1160–1174.

[55] STRICKLAND, S. L., REDDY, J. S., ALLEN, M., N'SONGO, A., BURGESS, J. D., CORDA, M. M., BALLARD, T., WANG, X., CAR-RASQUILLO, M. M., BIERNACKA, J. M. et al. (2020). MAPT haplotype–stratified GWAS reveals differential association for AD risk variants. *Alzheimer's & Dementia* **16**(7) 983–1002.

[56] SU, Z., ZHU, G., CHEN, X. and YANG, Y. (2016). Sparse envelope model: Efficient estimation and response variable selection in multivariate linear regression. *Biometrika* **103**(3) 579–593.

[57] SU, Z. and COOK, R. D. (2011). Partial envelopes for efficient estimation in multivariate linear regression. *Biometrika* **98** 133–146.

[58] SU, Z. and COOK, R. D. (2013). Estimation of multivariate means with heteroscedastic errors using envelope models. *Statistica Sinica* **23**(1) 213–230.

[59] TALHOUK, A., DOUCET, A. and MURPHY, K. (2012). Efficient Bayesian inference for multivariate probit models with sparse inverse correlation matrices. *Journal of Computational and Graphical Statistics* **21**(3) 739–757.

[60] TAN, M.-S., YANG, Y.-X., XU, W., WANG, H.-F., TAN, L., ZUO, C.-T., DONG, Q., TAN, L., SUCKLING, J. and YU, J.-T. (2021). Associations of Alzheimer's disease risk variants with gene expression, amyloidosis, tauopathy, and neurodegeneration. *Alzheimer's Research & Therapy* **13**(1) 1–11.

[61] TORVELL, M., CARPANINI, S. M., DASKOULIDOU, N., BYRNE, R. A., SIMS, R. and MORGAN, B. P. (2021). Genetic Insights into the Impact of Complement in Alzheimer's Disease. *Genes* **12**(12) 1990.

[62] Veera Manikandan, R. and Anand, R. (2015). P2-012: A genome wide scan for genetic variations with inverse association between Alzheimer's disease and breast cancer. *Alzheimer's & Dementia* **11**(7S_Part_10) 485–485.

[63] Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A. and Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *The American Journal of Human Genetics* **101**(1) 5–22.

[64] Vounou, M., Nichols, T. E., Montana, G. and Alzheimer's Disease Neuroimaging Initiative (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: A sparse reduced-rank regression approach. *NeuroImage* **53**(3) 1147–1159.

[65] Wang, H., Nie, F., Huang, H., Risacher, S. L., Saykin, A. J., Shen, L. and Alzheimer's Disease Neuroimaging Initiative (2012). Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* **28**(12) 127–136.

[66] White, C. C., Yang, H.-S., Schneider, J. A., Bennett, D. A., De Jager, P. L. and group, C. A. F. (2021). A genome-wide investigation of clinicopathologic endophenotypes uncovers a new susceptibility locus for tau pathology at Neurotrimin (NTM). *Alzheimer's & Dementia* **17** 051682.

[67] Zhang, C. and Yu, T. (2008). Semiparametric detection of significant activation for brain fMRI. *The Annals of Statistics* **36**(4) 1693–1725.

[68] Zhu, H., Khondker, Z., Lu, Z. and Ibrahim, J. G. (2014). Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of the American Statistical Association* **109**(507) 977–990.

Yanbo Shen. Department of Statistics,
University of Wisconsin-Madison, Madison, WI, 53706, USA.
E-mail address: yshen84@wisc.edu

Yeonhee Park. Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, 53726, USA. E-mail address: ypark56@wisc.edu

Saptarshi Chakraborty. Department of Biostatistics,
State University of New York at Buffalo, Buffalo, NY, 14214, USA. E-mail address: chakrab2@buffalo.edu

Chunming Zhang. Department of Statistics,
University of Wisconsin-Madison, Madison, WI, 53706, USA.
E-mail address: cmzhang@stat.wisc.edu