Envelope Models in Gaussian Copula Regression and High-dimensional Hypothesis Testing

by

Wei Tu

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

Doctor of Philosophy

in

Statistics

Department of Mathematical and Statistical Sciences
University of Alberta

Abstract

Envelopes, introduced by Cook et al. (2007), encompass a class of methods for increasing efficiency in multivariate analyses without altering traditional objectives. Envelopes have been successfully incorporated to a variety of regression models from generalized linear models to quantile regression. Despite the potential of achieving substantial efficiency gains, inference based on an envelope is not invariant under rescaling and other transformation of the variables. Furthermore, envelopes have mostly been studied in the context of regression, but not hypothesis testing. This thesis mainly contains three parts. In the first part, we develop adaptive estimation and inference methods for envelope models that achieve the same performance without the knowledge of the marginal transformations of the responses and predictors in the context of response envelopes. In the second part, we study predictor envelopes and sparse envelopes. Using a Kendall's tau based covariance matrix estimator and the scaled envelope models, we propose an envelope-based Gaussian copula estimator. In the third part of the thesis, we propose an envelope-based high-dimensional multivariate test for mean vector that can efficiently exploit the dependency structures within the high-dimensional vector. In all three parts, theoretical properties of the procedures are studied, and exten-

sive simulation studies and data analysis have been conducted to illustrate the usefulness of the proposed procedures in practice.

Preface

Some of the research conducted for this thesis forms part of a collaboration with Dr. Linglong Kong and Dr. Rohana Karunamuni at the University of Alberta, Dr. Zhihua Su from University of Florida and Dr. Lan Wang from University of Miami. The model formulation and data analysis in Chapters 2, 3 and 4 are my original work.

Chapter 2 of the thesis is currently under preparation to be submitted as W. Tu, L. Kong, Z. Su and R. Karunamuni, "Gaussian copula response envelope model". I will be responsible for the manuscript composition. Dr. Linglong Kong and Dr. Zhihua Su will be the corresponding author, involved with data analysis and manuscript composition. Dr. Rohana Karunamuni will contribute to manuscript edits and assist with the data analysis.

Chapter 3 of the thesis is currently under preparation to be submitted as W. Tu, L. Kong, Z. Su and R. Karunamuni, "Predictor envelope model and partial least squares under Gaussian copula regression". I will be responsible for the manuscript composition. Dr. Linglong Kong and Dr. Zhihua Su will be the corresponding author, involved with data analysis and manuscript composition. Dr. Rohana Karunamuni will contribute to manuscript edits and assist with the data analysis.

Chapter 4 of the thesis is currently under preparation to be submitted as W. Tu, L. Kong, Z. Su, L. Wang and R. Karunamuni, "Envelope-based high-dimensional multivariate test for mean vector". I will be responsible for the manuscript composition. Dr. Linglong Kong and Dr. Zhihua Su will be the corresponding author, involved with data analysis and manuscript composition. Dr. Rohana Karunamuni and Dr. Lan Wang will contribute to manuscript edits and assist with the data analysis.

To my grandmother

For everything.

Acknowledgements

First and foremost I want to thank my supervisors Dr. Linglong Kong and Dr. Rohana Karunamuni for supporting me during the last 5 years. I want to thank Dr. Kong for his patience with me and for guiding me in the road of research, showing me what qualities a good researcher should have and all the encouragement and opportunities. I could not have been where I am today without the tremendous efforts he has dedicated to me. I want to thank Dr. Karunamuni for his help for the last 7 years. Seven years ago, I did not know much about statistics and thank you for guiding me through my Masters and PhD degrees.

I would also like to express my sincere thanks to Dr. Linglong Kong, Dr. Irina Dinu and Dr. Ivan Mizera and Dr. Niu Di for serving in my dissertation examination committee. My sincere thanks are extended to other members of my examining committee: Dr. Jiguo Cao for his thorough review, insightful comments. I want to thank Dr. Zhihua Su for introducing me to the world of Envelopes. I would also thank all the professors who taught me before for the mentorship. Also, I could't finish this degree without all the help, financially and many others, from Department of Mathematical and Statistical Sciences.

Last but not least, I would like to thank all my friends, and my grandmother

who passed away, my parents and my sister for supporting me unconditionally.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Overview of the Thesis

The evolution of data acquisition technologies and computing power has allowed researchers nowadays to collect and store data with high dimensionality and complex structure much more efficiently. Examples can be found in gene expression microarray data, single nucleotide polymorphism (SNP) data, magnetic resonance imaging (MRI) data, high-frequency financial data, and others. Estimation and testing are unarguably the most fundamental tasks in statistical inference. As a rule of the thumb, with the increase of the dimensionality of the data, the amount of data needed to achieve certain level of statistical accuracy increases greatly. However, in practice, it is often too expensive or even impossible to collect the amount of data needed; for example, in the setting of clinical trials and neuroimaging experiments. The blessings come from high dimensional data is that many of the variables collected might be correlated with each other, and some of them are not relevant to the tasks. Thus it is of great importance to learn the inner structure of the data and find an efficient representation of them. Broadly speaking, dimension reduction represents the

action of replacing data with a lower dimensional function of the data. In this thesis, we study a specific type of dimension reduction methods called envelopes.

Envelopes, which were introduced by Cook et al. (2007)[10] and developed for the multivariate linear model by Cook et al. (2010)[11], encompass a class of methods for increasing efficiency in multivariate analyses without altering traditional objectives. The original goal of the envelope model is to improve the estimation efficiency of standard multivariate analysis methods, which is equivalent to obtaining the same estimation efficiency with much fewer observations. The efficiency gains are usually obtained by removing immaterial information for the data. The monograph of Cook (2018)[7] gives a comprehensive review of the different types of envelope models and Lee and Su (2019)[30] presents a shorter review paper that provides a quick introduction to the envelope models.

Before we formally introduce the notation and structure of envelopes, we first explain the intuition behind the envelopes using an example. The first envelope model is proposed in the context of multivariate linear regression, where the goal is to estimate the effects of multiple predictors on multivariate responses. For example, the government initiated a new job training program, and several characteristics of the participants as well as many performance metrics of them in the job market before and after participating the program have been recorded. While some (or some linear combinations of) performance metrics are affected by the participation of the program, some of the linear combinations do not respond to the change in the predictors at all. These linear combinations can be called the immaterial part of the data. The in-

clusion of the immaterial part does not bring any extra information to the estimation process, rather it brings extraneous variation in the estimation, resulting in wider confidence intervals of the regression coefficients. The goal of envelope methods is to improve the estimation efficiency while still keeping the objectives of the regression analysis. For example, different tools such as multivariate linear regression, quantile regression or penalized linear regression can be used here and envelope models can be combined with any of these models to achieve more efficient estimation.

With the goal of envelope models in mind, here we use the Berkeley guidance data to see how it achieves the goal. The Berkeley Guidance Study [42] was a longitudinal study that monitored the growth of children born in Berkeley, California between 1928 and 1929. The data recorded the weight, height, leg circumference and strength of these children from age 2 to age 18. Here we use the heights at ages 13 and 14 as the bivariate responses $(Y_1, Y_2)^T$ and sex as the predictor X, with $X = 1$ being boys and 0 being girls. The regression coefficient $\beta = (\beta_1, \beta_2)^T$ estimates the height difference between boys and girls at ages 13 and 14. Both standard multivariate linear regression and envelope response model were applied to the data. Both models result in the same estimation of $\widehat{\beta}$ with very different standard deviations. The bootstrap standard deviation of β based on the residual bootstrap with 200 replications are 1.90 and 1.81 for standard multivariate linear regression and are 0.19 and 0.19 for envelope response model. This implies that the sample size needed for standard model is about $(1.80/0.19)^2 = 90$ times of the envelope model to achieve the same accuracy. Figure 1 visualizes the differences in inference using the standard method and the envelope model. In this scatterplot, the
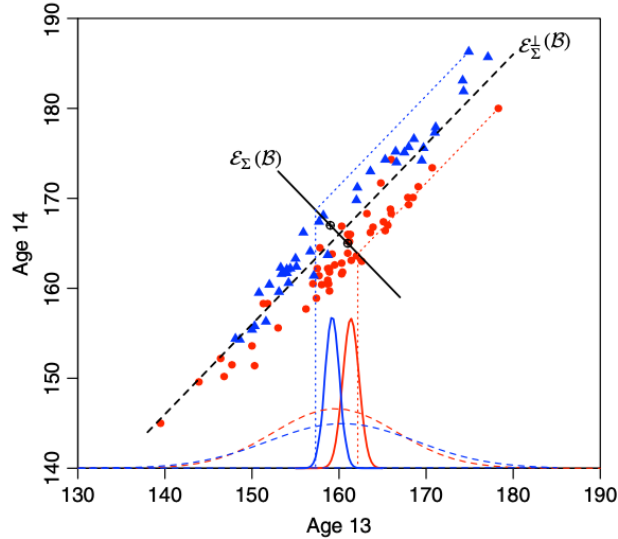
Figure 1.1: Berkeley guidance study: height at age 13 versus 14 with blue triangles represent males and red circles represent females. The lines marked by $\varepsilon_\Sigma(\mathcal{B})$ and $\varepsilon_\Sigma^\perp(\mathcal{B})$ denote the estimated envelopes and its orthogonal complement.

two histograms in dash represent the projection of data onto horizontal axis, and there are a lot of overlap between these two histograms, which means that the height of boys and girls are not separable at age 13. The two histograms in solid represent the projection of the material part of the data found using the envelope. The material part $G_0^T \mathbf{Y}$ here can be interpreted as the height difference $Y_1 - Y_2$, and the two solid distribution curves are well separated. This example shows that the efficiency gains in the envelope model allows us to detect weak signals that standard models fail to detect.

The first envelope model was introduced by [10] under the classical multivariate linear regression model. Consider a multivariate response vector $\mathbf{Y} \in \mathbb{R}^r$ and multiple covariates $\mathbf{X} \in \mathbb{R}^p$. A standard multivariate linear

4

regression assumes that $\mathbf{Y}$ and $\mathbf{X}$ are linearly related as follows:

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}, \tag{1.1}$$

where the predictors are centered in the sample such that $\sum_{i=1}^{n} \mathbf{X}_i = 0$, the errors $\boldsymbol{\epsilon} \in \mathbb{R}^r$ are independently and identically distributed normal vectors with mean 0 and covariance matrix $\boldsymbol{\Sigma} > 0$, $\boldsymbol{\alpha} \in \mathbb{R}^r$ is an unknown vector of intercepts, and $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ is an unknown matrix of regression coefficients.

Under the classical multivariate linear regression model (1.1), the estimation of $\boldsymbol{\beta}$ does not use the relationship among the response variables. The response envelope model exploits the dependencies among $\mathbf{Y}$ to achieve efficient estimation of $\boldsymbol{\beta}$. The intuition behind the envelope model is that there exist linear combinations of the response vector whose distribution is invariant to changes in the predictor vector. If such linear combinations exist, the variation of estimation for $\boldsymbol{\beta}$ in (1.1) can be drastically reduced.

Denote $\mathcal{S}$ as a d-dimensional subspace of $\mathbb{R}^r$ with $u < r$. Let $\mathbf{G} \in \mathbb{R}^{r \times u}$ be an orthonormal basis of $\mathcal{S}$ and $\mathbf{G}_0 \in \mathbb{R}^{r \times (r-u)}$ be an orthonormal basis of $\mathcal{S}^\perp$, the orthogonal complement of $\mathcal{S}$. The envelope model arises by imposing the following two conditions on $\mathbf{G}^T\mathbf{Y}$ and $\mathbf{G}_0^T\mathbf{Y}$:

(i) $\mathbf{G}_0^T\mathbf{Y}|\mathbf{X} \sim \mathbf{G}_0^T\mathbf{Y}$ and (ii) $\mathrm{cov}(\mathbf{G}^T\mathbf{Y}, \mathbf{G}_0^T\mathbf{Y}|\mathbf{X}) = 0$.

Condition (i) implies that the marginal distribution of $\mathbf{G}_0^T\mathbf{Y}$ must be unaffected by changes in $\mathbf{X}$, and condition (ii) requires that $\mathbf{G}_0^T\mathbf{Y}$ must be unaffected by changes in $\mathbf{X}$ through an association with $\mathbf{G}^T\mathbf{Y}$. The above two conditions combined imply that any dependence of $\mathbf{Y}$ on $\mathbf{X}$ must be concentrated on $\mathbf{G}^T\mathbf{Y}$, and $\mathbf{G}_0^T\mathbf{Y}$ carries no material information for the estimation of

$\boldsymbol{\beta}$ in (1.1). The two conditions above hold if and only if (a) $\mathcal{B} := \text{span}(\boldsymbol{\beta}) \in \mathcal{S}$ and (b) $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}$. There are many subspaces that reduce $\boldsymbol{\Sigma}$ and contain $\mathcal{B}$. We define the $\boldsymbol{\Sigma}$-envelope of $\mathcal{B}$ as the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathcal{B}$. The smallest here means the intersection of all such subspaces. The $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ is often denoted as $\varepsilon_{\boldsymbol{\Sigma}}(\mathcal{B})$. The multivariate linear regression model (1.1) now can be reparametrized in terms of $\varepsilon_{\boldsymbol{\Sigma}}(\mathcal{B})$ by using a basis. Let $u = \dim(\mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B}))$, and let $\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times r}$ be an orthogonal matrix with $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$, $\text{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Sigma}}(\mathcal{B})$, and $\text{span}(\boldsymbol{\Gamma}_0) = \mathcal{E}_{\boldsymbol{\Sigma}}^{\perp}(\mathcal{B})$. Then the envelope copula model can be written as

$$\mathbf{Y} = \alpha + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T,$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, and $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ is the coordinate of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\eta}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$, respectively. The variable of interest is still $\boldsymbol{\beta}$, and we normally do not infer about the constitute parameters $\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$. The values of the $\boldsymbol{\eta}, \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ depend on the choice of $\boldsymbol{\Gamma}$. It should also be noted that the basis matrix $\boldsymbol{\Gamma}$ is not identifiable because $\boldsymbol{\Gamma}\mathbf{O}$ for any orthogonal matrix $\mathbf{O}$ will lead to an equivalent model. However, the envelope $\varepsilon_{\Sigma}(\mathcal{B}) = \text{span}(\boldsymbol{\Gamma})$ is identifiable, which is the key for the estimation of $\boldsymbol{\beta}$. Under this formulation, it is also straightforward to see the covariance matrix $\boldsymbol{\Sigma}$ can be decomposed as the sum of the covariance matrix of the material part of the response $\text{var}(\mathbf{P}_\varepsilon \mathbf{Y}) = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T$ and immaterial part of the response $\text{var}(\mathbf{Q}_\varepsilon \mathbf{Y}) = \boldsymbol{\Gamma}_0\boldsymbol{\Omega}\boldsymbol{\Gamma}_0^T$.

Unlike most sufficient dimension reduction methods, which usually avoid specifying the parametric form of a model, envelopes work mostly in model-

based contexts without altering traditional objectives. As we can see from the last example, envelope methods can result in massive efficiency gains relative to standard methods, gains that are equivalent to increasing the sample size many times over. Since its introduction, many advances have taken place, including but not limited to response envelope model, predictor envelope model and partial least squares, generalized linear models, Bayesian analysis, variable selection and quantile regression. At the same time of the potential in bringing in huge efficiency gains, envelopes are very "sensitive" to many aspects of the data generating process. For example, the original envelope models were not scale-invariant; the current algorithms used in computing envelopes strongly depends on the starting values and in general do not have global minimizer; almost all envelope models essentially relies on the normality assumption of the error; the envelope models can only handle linear models and are quite sensitive to any transformation of the data. Some of the previous issues have been addressed. For example, scaled response envelope was proposed by Cook and Su[12] to address the scale-invariant issue. Quantile envelope model proposed in [17] does not rely on the normality assumption and is based on generalized estimating equations.

In this thesis, we focus on a few issues that have not been properly addressed in the current literature. First, most of current envelope models are based on the multivariate linear regression model. The linearity assumption there can be very strict and essentially uncheckable in practice. For example, it is very common to log transform a highly skewed variable such as white cell counts or viral loads to improve the linearity between responses and co-variates. Other transformations such as Box-Cox transformation, Fisher's z

transformation and variance stabilization transformation have been frequently used to improve the linear fit. The envelope models available in the current literature are not invariant or equivariant to these transformations, and even the dimension of the envelope will change. We have observed in practice that such transformations often resulting in envelope based estimators reduce back to ordinary least square estimators, which means no efficiency gain can be achieved. In order to address this issue, we propose to combine Gaussian copula regression model and envelopes. The essential tool we are using is the rank-based correlation estimators, which are invariant to continuous monotone transformations. Specifically, we focus on both response envelope model and predictor envelope model. In Chapter 2, we study the response envelope model and propose a novel Gaussian copula response envelope model, which is equivariant to a large class of popular transformation of the responses. In Chapter 3, we focus on the predictor envelope model, which aims to achieve dimension reduction in the predictor space. Even though the goal is the same, the techniques used in Chapter 3 are different from those in Chapter 2. The proposed Gaussian copula envelope models have the potential for bringing huge efficiency gains when the original envelope model fails to do so.

Most of the current literature on envelopes focus on regression models. In the area of hypothesis testing in high-dimensional data, envelope models actually can bring in huge efficiency gains. In Chapter 4, we focus on the use of envelope models in the domain of hypothesis testing. Specifically, we focus on an envelope-based Hotelling's T^2 test and a likelihood ratio test. The proposed tests are more efficient when the high-dimensional mean vector can be represented using a lower dimensional envelope structure in the covariance

matrix.

The rest of the thesis is organized as follows. Chapter 2 introduces the Gaussian copula response envelope model and describes the framework of model formulation, estimation, selection of envelope dimension and inference under the model. Chapter 3 introduces the Gaussian copula predictor envelope model and partial least squares. The theoretical results highlight the scenarios when the proposed model can result in substantial efficiency gains, which were verified by the simulation results and data analysis. Chapter 4 introduces several envelope-based test statistics for high-dimensional data. The reference distribution of the test statistics have been established and simulation studies suggest substantial improvement over the methods in the current literature.

Chapter 2

Gaussian Copula Response Envelope Model

2.1 Introduction

Regression analysis is one of most common tasks in data analysis, where we try to establish the relationship between a set of responses and a set of covariates. Linear regression analysis is arguably the most commonly used method in statistics. However, the linearity assumption is often too restrictive in practice. To deal with the nonlinearity, one can either transform the data so that there is relationship between the transformed variables or use a non-linear function. The use of linear models is greatly enhanced through the use of various transformations of the data. There are a few reasons why a transformation of the data makes sense in practice. For example, sometimes there is some prior information that the data generating process corresponds to a theoretical model, which has a specific parametric form; sometimes the data itself appears a nonlinear pattern between the responses and covariates. What kind of data can be transformed into linear relationship is often hard to determine. Data transformations, such as the Box-Cox transformation, Fisher's z transfor-

mation and variance stabilization transformation, have been frequently used to improve the linear fit. However, typically these transformations need to be applied before the regression, and various transformations have to be applied in order to find the right one.

Consider multivariate response vector $\mathbf{Y} \in \mathbb{R}^r$ and multiple covariates $\mathbf{X} \in \mathbb{R}^p$. A standard multivariate linear regression assumes that $\mathbf{Y}$ and $\mathbf{X}$ are linearly related as follows:

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\epsilon}, \tag{2.1}$$

where the predictors are centered in the sample such that $\sum_{i=1}^{n} \mathbf{X}_i = 0$, the errors $\boldsymbol{\epsilon} \in \mathbb{R}^r$ are independently and identically distributed normal vectors with mean 0 and covariance matrix $\boldsymbol{\Sigma} > 0$, $\boldsymbol{\alpha} \in \mathbb{R}^r$ is an unknown vector of intercepts, and $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ is an unknown matrix of regression coefficients.

Model (2.1) is not always realistic as we have discussed in the Introduction. The question is whether it is possible to estimate $\boldsymbol{\beta}$ without knowing specific transformations that can make the linearity assumption more realistic. Let us consider the following model, which has been widely used in a range of applications,

$$g(\mathbf{Y}) = \boldsymbol{\alpha} + \boldsymbol{\beta}f(\mathbf{X}) + \boldsymbol{\epsilon}, \tag{2.2}$$

where $f(\mathbf{X}) = (f_1(X_1), \ldots, f_p(X_p))^\top$ and $g(\mathbf{Y}) = (g_1(Y_1), \ldots, g_r(Y_r))^\top$ are unknown functions. Examples of model (2.2) include the additive regression model, single index model, copula regression model, and semiparametric proportional hazards models. Typically when we apply a transformation to data, the transformation functions are continuous and one-to-one; for example, log transformation, Box-cox transformation and Fisher's z transformation. The

11

functions that satisfy these two assumptions must be strictly monotone.

The model can be further formulated as follows. Suppose we have an independent and identically distributed random sample $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{X}_1), \ldots, \mathbf{Z}_n = (\mathbf{Y}_n, \mathbf{X}_n) \in \mathbb{R}^{p+r}$. The observations $(\mathbf{Y}_i, \mathbf{X}_i)$ satisfy a Gaussian copula regression model, if there exists a set of strictly increasing functions $g = \{g_1, \ldots, g_r\}$ and $f = \{f_1, \ldots, f_p\}$ such that the marginally transformed random vectors $\widetilde{\mathbf{Z}}_i = (\widetilde{\mathbf{Y}}_i, \widetilde{\mathbf{X}}_i) = (g_1(Y_1), \ldots, g_r(Y_r), f_1(X_1), \ldots, f_p(X_p))$ satisfy $\widetilde{\mathbf{Z}}_i \sim N_{p+r}(0, \mathbf{\Sigma}_Z)$ for some positive-definite covariance matrix $\mathbf{\Sigma}_Z$. Under Gaussian copula regression model, one has the following linear relationship for the transformed data:

$$\widetilde{\mathbf{Y}}_i = \boldsymbol{\beta}\widetilde{\mathbf{X}}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{2.3}$$

where $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ and ϵ_i are i.i.d. zero-mean Gaussian variables. The fundamental difference between the Gaussian copula regression model (2.3) and the conventional linear regression model is that one observes $\{(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n)\}$, instead of $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$ as the transformations f and g are unknown.

Cai and Zhang (2018)[4] focused on a high-dimensional setting where p is comparable to or much larger than n and $\boldsymbol{\beta}$ is sparse and they only considered the situation of univariate response. Compared with other methods such as those for the additive regression model and single index model, a significant advantage for their proposed estimation and inference procedures is that they do not require estimation of the marginal transformations. For example, one can select the important variables x_i without any knowledge of the transformations f_i. This makes the methods more flexible and adaptive, and

12

achieves the same optimal rate as that for high-dimensional linear regression. Zhao and Genest (2019)[47] generalized their work to multivariate responses scenarios, and also covered cases in which the variables exhibit greater tail dependence than if they were jointly normal; for example, the multivariate Student t distribution.

Borrowing the idea from Cai and Zhang (2018)[4], in this chapter, we apply the use of Gaussian copula regression on response envelope models. Originally proposed envelope models are not invariant or equivariant under the rescaling of the variables, let along other strictly increasing transformations. This brings some inconvenience for the practical use of envelope models. For example, it is difficult to ensure all the responses are measured by the same or similar scales, and sometimes the responses might measure different things, which make them not comparable. We have observed in practice that such transformations often resulting in envelope based estimators reduce back to ordinary least square estimators, which means no efficiency gain can be achieved. The use of envelope models will be affected greatly as certain variable with large variance might dominate the results, just as what we see in methods like principal component regression, ridge regression and partial least squares. To address this issue, Cook and Su (2013)[12] proposed scaled response model, which is invariant to a scaling of the response variables. In both Cai and Zhang (2018)[4] and Zhao and Genest (2019)[47], the response variables were standardized to have the same variance. However, since the original response envelope model is not variant towards scaling, the direct application of Gaussian copula model will not work. Instead, we combine the scaled response envelope model and Gaussian copula model.

The rest of this chapter is organized as follows. In Sections 2.2 and 2.3, we first introduce the background of response envelope model and scaled response envelope models. In Section 2.4, we introduce the proposed Gaussian copula response envelope model. The model formulation, estimation, selection of envelope dimension have been established. In section 2.5, the theoretical properties have been established, including the uniqueness, consistency and asymptotic normality. Simulation studies and data analysis can be found in Section 2.6.

2.2   Response Envelope Model

Under the classical multivariate linear regression model (2.1), estimation of $\boldsymbol{\beta}$ does not use the relationship among the response variables. The response envelope model exploits the dependencies among $\mathbf{Y}$ to achieve efficient estimation of $\boldsymbol{\beta}$. The intuition behind the model is that there exist linear combinations of the response vector whose distribution is invariant to changes in the predictor vector. If such linear combinations exist, the variation of estimation for $\boldsymbol{\beta}$ in (2.1) can be drastically reduced.

Denote $\mathcal{S}$ as a d-dimensional subspace of $\mathbb{R}^r$ with $u < r$. Let $\mathbf{G} \in \mathbb{R}^{r \times u}$ be an orthonormal basis of $\mathcal{S}$ and $\mathbf{G}_0 \in \mathbb{R}^{r \times (r-u)}$ be an orthonormal basis of $\mathcal{S}^\perp$, the orthogonal complement of $\mathcal{S}$. The envelope model imposes the following two conditions on $\mathbf{G}^T\mathbf{Y}$ and $\mathbf{G}_0^T\mathbf{Y}$:

(i) $\mathbf{G}_0^T\mathbf{Y}|\mathbf{X} \sim \mathbf{G}_0^T\mathbf{Y}$ and   (ii) $\mathrm{cov}(\mathbf{G}^T\mathbf{Y}, \mathbf{G}_0^T\mathbf{Y}|\mathbf{X}) = 0$.

Condition (i) implies that the marginal distribution of $\mathbf{G}_0^T\mathbf{Y}$ must be unaffected by changes in $\mathbf{X}$, and condition (ii) requires that $\mathbf{G}_0^T\mathbf{Y}$ must be un-

affected by changes in $\mathbf{X}$ through an association with $\mathbf{G}^T\mathbf{Y}$. The above two conditions combined imply that any dependence of $\mathbf{Y}$ on $\mathbf{X}$ must be concentrated on $\mathbf{G}^T\mathbf{Y}$, and $\mathbf{G}_0^T\mathbf{Y}$ carries no material information for the estimation of $\boldsymbol{\beta}$ in (2.1). The two conditions above hold if and only if (a) $\mathcal{B} := \mathrm{span}(\boldsymbol{\beta}) \in \mathcal{S}$ and (b) $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}$. The definition of reducing subspace is given below:

Definition 1 *A subspace $\mathcal{R} \in \mathbb{R}^r$ is said to be a reducing subspace of $\mathbf{M} \in \mathbb{S}^{r \times r}$ if $\mathcal{R}$ decomposes $\mathbf{M}$ as $\mathbf{M} = \mathbf{P}_{\mathcal{R}} \mathbf{M} \mathbf{P}_{\mathcal{R}} + \mathbf{Q}_{\mathcal{R}} \mathbf{M} \mathbf{Q}_{\mathcal{R}}$, where $\mathbf{P}$ represents a projection operator and $\mathbf{Q} = \mathbf{I} - \mathbf{P}$. If $\mathcal{R}$ is a reducing subspace of $\mathbf{M}$, we say that $\mathcal{R}$ reduces $\mathbf{M}$.*

There are many subspaces that reduce $\boldsymbol{\Sigma}$ and contain $\mathcal{B}$. We define the $\boldsymbol{\Sigma}$-envelope of $\mathcal{B}$ as the smallest reducing subspace of $\boldsymbol{\Sigma}$ that contains $\mathcal{B}$. The smallest here means the intersection of all such subspaces. The $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ is often denoted as $\varepsilon_{\boldsymbol{\Sigma}}(\mathcal{B})$. The multivariate linear regression model (2.1) now can be reparametrized in terms of $\varepsilon_{\boldsymbol{\Sigma}}(\mathcal{B})$ by using a basis. Let $u = \dim(\boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}}(\mathcal{B}))$, and let $\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0 \in \mathbb{R}^{r \times r}$ be an orthogonal matrix with $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$, $\mathrm{span}(\boldsymbol{\Gamma}) = \boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}}(\mathcal{B})$, and $\mathrm{span}(\boldsymbol{\Gamma}_0) = \boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}}^{\perp}(\mathcal{B})$. Then the envelope copula model can be written as

$$\mathbf{Y} = \alpha + \boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \tag{2.4}$$

where $\boldsymbol{\beta} = \boldsymbol{\Gamma}\boldsymbol{\eta}$, and $\boldsymbol{\eta} \in \mathbb{R}^{u \times p}$ is the coordinate of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\eta}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(r-u) \times (r-u)}$ are the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$, respectively. The variable of interest is still $\boldsymbol{\beta}$, and we normally do not infer about the constitute parameters $\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}$ and

15

$\boldsymbol{\Omega}_0$. The values of the $\boldsymbol{\eta}, \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ depend on the choice of $\boldsymbol{\Gamma}$. It should also be noted that the basis matrix $\boldsymbol{\Gamma}$ is not identifiable in (2.4) because $\boldsymbol{\Gamma O}$ for any orthogonal matrix $\mathbf{O}$ will lead to an equivalent model. However, the envelope $\varepsilon_\Sigma(\mathcal{B}) = \text{span}(\boldsymbol{\Gamma})$ is identifiable, which is the key for the estimation of $\boldsymbol{\beta}$. Under this formulation, it is also straightforward to see the covariance matrix $\boldsymbol{\Sigma}$ can be decomposed as a sum of the covariance matrix of the material part of the response $\text{var}(\mathbf{P}_\varepsilon \mathbf{Y}) = \boldsymbol{\Gamma \Omega \Gamma}^T$ and immaterial part of the response $\text{var}(\mathbf{Q}_\varepsilon \mathbf{Y}) = \boldsymbol{\Gamma}_0 \boldsymbol{\Omega \Gamma}_0^T$.

Estimation of the parameters in (2.4) depends on the dimension of the envelope u. Assuming that u is known and a normal likelihood function, Cook et al. (2010)[11] developed some estimation procedures. The maximum likelihood estimators $\widehat{\boldsymbol{\mathcal{E}}}_\Sigma(\mathcal{B})$ of $\boldsymbol{\mathcal{E}}_\Sigma(\mathcal{B})$ and of the remaining parameters are determined as

$$
\begin{aligned}
\widehat{\boldsymbol{\mathcal{E}}}_{\boldsymbol{\Sigma}}(\mathcal{B}) &= \text{span}\{\arg\min_{\mathbf{G}}(\log|\mathbf{G}^T S_{\mathbf{Y|X}}\mathbf{G}| + \log|\mathbf{G}^T S_{\mathbf{Y}}^{-1}\mathbf{G}|)\}, \\
\widehat{\boldsymbol{\eta}} &= \widehat{\boldsymbol{\Gamma}}^T \widehat{\boldsymbol{\beta}}_{\text{ols}}, \\
\widehat{\boldsymbol{\beta}} &= \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\eta}}, \\
\widehat{\boldsymbol{\Omega}} &= \widehat{\boldsymbol{\Gamma}}^T S_{\mathbf{Y|X}}\widehat{\boldsymbol{\Gamma}}, \\
\widehat{\boldsymbol{\Omega}}_0^T &= \widehat{\boldsymbol{\Gamma}}_0^T S_{\mathbf{Y}}\widehat{\boldsymbol{\Gamma}}_0, \\
\widehat{\boldsymbol{\Sigma}} &= \widehat{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Gamma}}^T + \widehat{\boldsymbol{\Gamma}}_0\widehat{\boldsymbol{\Omega}}_0\widehat{\boldsymbol{\Gamma}}_0^T,
\end{aligned}
\tag{2.5}
$$

where $S_{\mathbf{Y}}$ and $S_{\mathbf{X}}$ are the sample covariance matrices of $\mathbf{Y}$ and $\mathbf{X}$, respectively, $S_{\mathbf{YX}}$ is the sample covariance matrix of $\mathbf{Y}$ and $\mathbf{X}$, $\widehat{\boldsymbol{\beta}}_{\text{ols}} = S_{\mathbf{YX}}S_{\mathbf{X}}^{-1}$ is the ordinary least square (OLS) estimator of $\boldsymbol{\beta}$, $S_{\mathbf{Y|X}}$ is the sample covariance matrix of the conditional distribution of $\mathbf{Y}$ given $\mathbf{X}$, $\min_{\mathbf{G}}$ is over all semi-orthogonal matrices $\mathbf{G} \in \mathbb{R}^{r \times u}$, $\hat{\boldsymbol{\Gamma}}$ is any semi-orthogonal basis matrix for $\widehat{\boldsymbol{\mathcal{E}}}_\Sigma(\mathcal{B})$, and $\hat{\boldsymbol{\Gamma}}_0$

is any semi-orthogonal basis matrix for the orthogonal complement of $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$. The fully maximized log-likelihood for fixed u is then given by

$$
\begin{aligned}
L_u = & -(nr/2)\log(2\pi) - nr/2 - (n/2)\log|S_{\mathbf{Y}}| \\
& - (n/2)\log|\mathbf{\Gamma}^T S_{\mathbf{Y}|\mathbf{X}}\mathbf{\Gamma}| - (n/2)\log|\mathbf{\Gamma}^T S_{\mathbf{Y}}^{-1}\mathbf{\Gamma}|.
\end{aligned}
\tag{2.6}
$$

The above log-likelihood can be further simplified as the following objective function for the estimation of $\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B})$

$$
\widehat{\mathcal{E}}_{\mathbf{\Sigma}}(\mathcal{B}) = \arg\min_{\mathbf{\Gamma}} \log|\mathbf{\Gamma}^T S_{\mathbf{Y}|\mathbf{X}}\mathbf{\Gamma}| + \log|\mathbf{\Gamma}^T S_{\mathbf{Y}}^{-1}\mathbf{\Gamma}|,
$$

where the minimum is taken over all semi-orthogonal matrices $\mathbf{\Gamma} \in \mathbb{R}^{r \times u}$. It is worth pointing out that the envelope model (2.4) does not rely on the normality assumption, and only the estimation procedure used the normal likelihood. The performance of envelope models under the departure from normality has been studied by Cook (2018)[7], Su and Cook (2011)[39] and Park et al. (2017)[36].

The optimization problem here is not trivial due to two reasons: (1) this is a constrained optimization problem since $\mathbf{\Gamma}$ is semi-orthogonal and the optimization is done on a Grassmannian manifold instead of the Euclidean space; (2) the objective function in (2.6) is not convex with respect to $\mathbf{\Gamma}$, which means a global optimizer does not always exist and good starting values are essential. Chapter 6 in Cook (2018)[7] provides a good summary of state-of-the-art developments of envelope algorithms. Popular algorithms include the one-direction-at-a-time (1D) algorithm (Cook and Zhang (2016)[14]), envelope coordinate descent algorithm (Cook and Zhang (2018)[15]) and non-Grassmann estimation algorithm (Cook et al. (2016)[8]).

The maximum likelihood estimation method described above are based on

17

a known u. The dimension of the envelope subspace, u, is a model selection parameter. In practice, the popular ways of selecting u include sequential likelihood ratio test, information criterion such as AIC or BIC, and cross-validation.

2.3   Scaled Response Envelope Model

The response envelope model presented in Section 2.2 has the potential to yield an estimator of $\boldsymbol{\beta}$ with much less variation than the ordinary least squares estimator, which is of great importance since this is equivalent to the increase of sample size in practice. However, suppose we rescale the response $\mathbf{Y}$ by the form of $\mathbf{Y} \to \Lambda \mathbf{Y}$, where Λ is a non-singular diagonal matrix. Similar to principal component analysis, partial least squares and other methods, envelope is not invariant or equivariant under scale transformations. Figure (2.1) shows an example of how rescaling the response can affect an envelope analysis. After the rescaling, all linear combinations of $\mathbf{Y}$ are material to the regression and the envelope model is the same as the standard model and no efficiency gains have been achieved. This could potentially greatly limit the use of envelope models in practice since the measuring units might vary greatly for different response variables. In order to address this issue, Cook and Su (2013)[12] proposed scaled response envelope model. The essential idea in the scaled response envelope model is that the process of estimating $\Lambda \mathbf{P_\Gamma} \Lambda^{-1}$ is the same as treating Λ^{-1} as a diagonal similarity transformation to represent $\mathbf{P_\Gamma}$ in original coordinate system as $\Lambda \mathbf{P_\Gamma} \Lambda^{-1}$.

We introduce a diagonal scaling matrix $\Lambda = \operatorname{diag}\{1, \lambda_2, \ldots, \lambda_r\}$, and the
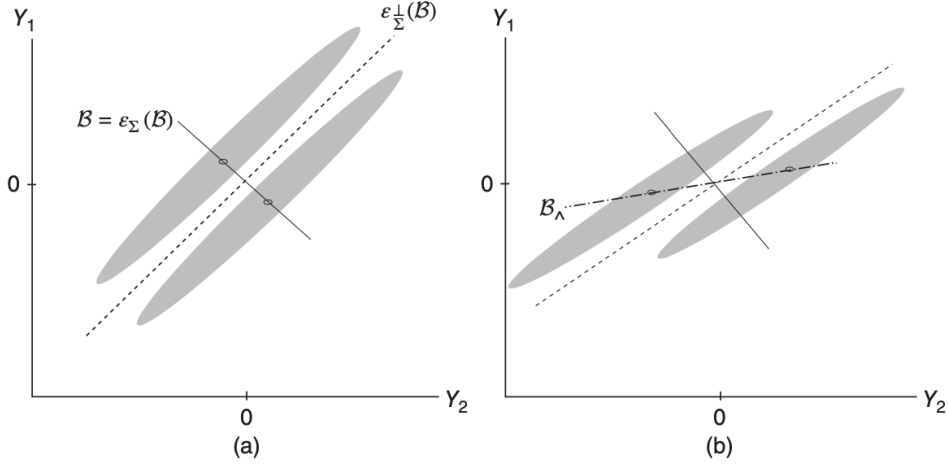
Figure 2.1: Schematic illustration of how rescaling the response can affect an envelope analysis. (a) original distributions, (b) rescaled distributions.

first element of Λ is set to 1 for identifiability. Assume that the scaled response vector $\Lambda^{-1}\mathbf{Y}$ follows the response envelope model (2.4), so the scaled envelope model for the original responses $\mathbf{Y}$ becomes

$$\mathbf{Y} = \alpha + \Lambda\boldsymbol{\Gamma}\boldsymbol{\eta}\mathbf{X} + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\Sigma} = \Lambda\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\Lambda + \Lambda\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T\Lambda. \qquad (2.7)$$

The scaled response $\Lambda^{-1}\mathbf{Y}$ follows an envelope model with u-dimensional envelope $\varepsilon_{\Lambda^{-1}\boldsymbol{\Sigma}\Lambda^{-1}}(\Lambda^{-1}\mathcal{B})$ and semi-orthogonal basis matrix $\boldsymbol{\Gamma}$. The material and immaterial pairs of $\mathbf{Y}$ are $\Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$, respectively. To show that model (2.7) follows the same envelope structure as (2.8), we denote $\mathbf{U} = \Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\mathbf{V} = \Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$. It is easy to verify that

$$\text{(i) } \mathbf{V}\mathbf{Y}|\mathbf{X} \sim \mathbf{V}\mathbf{Y} \quad \text{and} \quad \text{(ii) } \text{cov}(\mathbf{U}\mathbf{Y}, \mathbf{V}\mathbf{Y}|\mathbf{X}) = 0. \qquad (2.8)$$

The maximum likelihood estimators $\widehat{\Lambda}$ and $\widehat{\boldsymbol{\Gamma}}$ of Λ and $\boldsymbol{\Gamma}$, respectively, can be obtained by minimizing the objective function

$$L_u(\Lambda, \boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^T\Lambda^{-1}S_{\mathbf{Y}|\mathbf{X}}\Lambda^{-1}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^T\Lambda S_{\mathbf{Y}}^{-1}\Lambda\boldsymbol{\Gamma}|. \qquad (2.9)$$

19

Compared with the standard response envelope model, the scaled response envelope introduces a new scaling parameter Λ. The maximum likelihood estimators of the remaining parameters are as follows:

$$\widehat{\boldsymbol{\alpha}} = \bar{\mathbf{Y}};$$
$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\beta}}_{\mathrm{ols}},$$
$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\eta}},$$
$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} S_{\mathbf{Y}|\mathbf{X}} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}},$$
$$\widehat{\boldsymbol{\Gamma}}_0^T = \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^{-1} S_{\mathbf{Y}} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}}_0,$$
$$\widehat{\boldsymbol{\Sigma}} = \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Omega}} \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^T + \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Omega}}_0 \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^T.$$

(2.10)

The algorithms used in minimizing (2.9) are similar to those used in the standard response envelope model. The strategies used in choosing the envelope dimension u here are also similar, including likelihood-based methods such as AIC and BIC, and nonparametric methods such as cross validation and permutation test.

By introducing a scaling parameter for each response, the scaled envelope model further broadens the scope of envelope model, and can potentially bring efficiency gains that can not be offered by the ordinary envelope model. Cook and Su (2013)[12] also found that the scaled response envelope gives good results when the error distribution does not deviate substantially from the multivariate normal. The idea of furthering applying scaled envelope model to partial envelope model and inner envelope model has also been mentioned by above authors. The transformation matrix used in scaled envelope model is a diagonal matrix Λ. However, in practice, the linear relationship assumed in the multivariate linear model cannot be fixed by simply rescaling of the variables.

For example, in medical data, variables such as white cell counts and viral loads are often log transformed. The original response envelope model is not variant to such transformation, in fact, the original envelope model almost always reduce to the standard linear model, which brings no efficiency gain. To address this issues, we further generalize the original response envelope model using Gaussian copula regression in the next section.

2.4 Gaussian Copula Response Envelope Model

2.4.1 Motivation

The two sections above illustrates the potential efficiency gains of using envelope response models. However, model (2.4) and (2.7) are not always realistic in practice. The linearity assumption here between $\mathbf{X}$ and $\mathbf{Y}$ is often too restrictive and unrealistic. In introductory statistics classes, we teach the use of data transformation such as log transformation and Fisher's z transformation to improve a linear fit and sometimes an the equal variance assumption. Carroll and Rupert (1987)[5] presented some detailed discussions on the use of transformations in linear regression.

The linearity and constant covariance conditions posed in the envelope models are essentially uncheckable, which can be dangerous in practice. Lee et al. (2013)[28] extended the foundations of sufficient dimension reduction to allow for nonlinear reduction. Li and Song (2007)[31] developed foundations and methodology for nonlinear sufficient dimension reduction for functional data. Unlike most sufficient dimension reduction methods, envelope methods do require a parametric form of the regression model for estimation. Inspired

by the use of Gaussian copula regression in variable settings such as variable selection [4], variable screening [25], Gaussian mixture model [3] ,etc., we relax the linearity assumption on response envelop model to allow certain parametric form of the responses.

Let us consider the following model, which has been widely used in a range of applications,

$$g(\mathbf{Y}) = \boldsymbol{\alpha} + \boldsymbol{\beta} f(\mathbf{X}) + \boldsymbol{\epsilon}, \tag{2.11}$$

where $f(\mathbf{X}) = (f_1(X_1), \ldots, f_p(X_p)^\top$ and $g(\mathbf{Y}) = (g_1(Y_1), \ldots, g_r(Y_r)^\top$ are unknown functions. Examples of model (2.11) include the additive regression model, single index model, copula regression model, and semiparametric proportional hazards models. Typically when we apply a transformation to data, the transformation functions are continuous and one-to-one, for example, log transformation, Box-cox transformation and Fisher's z transformation. The functions that satisfy these two assumptions must be strictly monotone.

Under model (2.11), the linearity is assumed between $g(\mathbf{Y})$ and $f(X)$, and if the transformation functions f and g are known, we can use envelope models to further more efficiently estimate $\boldsymbol{\beta}$. In practice, often one has to try different transformation functions and decide whether they improve the linearity fit between the response and covariates, and this process can be arbitrary and exhausting. To address this issue, we propose the Gaussian copula response envelope model.

2.4.2 Model Formulation

The model (2.11) can be further formulated as follows. Suppose we have an independent and identically distributed random sample $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{X}_1), \ldots, \mathbf{Z}_n =$

$(\mathbf{Y}_n, \mathbf{X}_n) \in \mathbb{R}^{p+r}$. The observations $(\mathbf{Y}_i, \mathbf{X}_i)$ satisfies a Gaussian copula regression model, if there exists a set of strictly increasing functions $g = \{g_1, \ldots, g_r\}$ and $f = \{f_1, \ldots, f_p\}$ such that the marginally transformed random vectors $\widetilde{\mathbf{Z}}_i = (\widetilde{\mathbf{Y}}_i, \widetilde{\mathbf{X}}_i) = (g_1(Y_1), \ldots, g_r(Y_r), f_1(X_1), \ldots, f_p(X_p))$ satisfy $\widetilde{\mathbf{Z}}_i \sim N_{p+r}(0, \boldsymbol{\Sigma}_Z)$ for some positive-definite covariance matrix $\boldsymbol{\Sigma}_Z$. Under Gaussian copula regression model, one has the following linear relationship for the transformed data:

$$\widetilde{\mathbf{Y}}_i = \boldsymbol{\beta}\widetilde{\mathbf{X}}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n,$$

where $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ and ϵ_i are i.i.d. zero-mean Gaussian variables. The fundamental difference between the Gaussian copula regression model and conventional linear regression model (2.11) is that one observes $\{(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n)\}$, instead of $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$ as the transformations f and g are unknown.

Under Gaussian copula regression model, the linear relationship is assumed between the unobserved $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$. Similarly, we assume an envelope structure exists for the unobserved $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$, and we can use the observed $\{(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n)\}$ to estimate the envelope structure, which means the proposed Gaussian copula response envelope is invariant towards any strictly increasing transformation of the responses and predictors.

The key element in Gaussian copula regression is the use of rank-based estimator of correlation matrix. If we take a closer look at the likelihood function in (2.6), the essential information we need is the covariance matrix of $\widetilde{\mathbf{Y}}$ and the residuals, so the question now becomes how do we estimate $S_{\widetilde{\mathbf{Y}}}$ and

$S_{\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}}$ using only the observed $\mathbf{X}$ and $\mathbf{Y}$. In Cai and Zhang (2015)[4], the co-variance matrix $\mathbf{\Sigma}_Z$ is assumed to satisfy $\mathrm{diag}(\mathbf{\Sigma}_Z) = \mathbf{1}$. This condition is for identifiability because the scaling and shifting are absorbed in the marginal transformations. However, the original response envelope model is not invariant to the rescaling. In other words, if we only know the correlation matrix of $\widetilde{\mathbf{Y}}$ and the residuals, it will be impossible to get the same envelope estimates. Thus we need to use the scale invariant response envelope model. Before we introduce the estimation procedures, we first introduce the rank-based estimator of correlation matrix.

2.4.3 Rank-based Estimator of Correlation Matrix

The goal here is to estimate the $\mathbf{\Sigma}_Z$, the covariance matrix of $\widetilde{\mathbf{Z}}$, using only the observed $(\mathbf{Y}, \mathbf{X})$. Since the marginal transformations f and g are unknown, we use rank-based correlation of the observed data $\mathbf{Z}$ to estimate the covariance/correlation matrix $\mathbf{\Sigma}_Z$. Given that f and g are strictly monotone, $\mathbf{Z}$ and $\widetilde{\mathbf{Z}}$ have the same elliptical copula. Set $d = p + r$. If $\widetilde{\mathbf{Z}}_i \sim N_d(0, \mathbf{\Sigma}_Z)$ with $\mathbf{\Sigma}_Z = (\sigma_{jk})_{1 \leq j,k \leq d}$, then

$$\sigma_{jk} = \sin(\frac{\pi}{2}\tau_{jk}),$$

where τ_{jk} is Kendall's tau and is defined as

$$\tau_{jk} = E[\mathrm{sgn}(\tilde{Z}_{1j} - \tilde{Z}_{2j})\mathrm{sgn}(\tilde{Z}_{1k} - \tilde{Z}_{2k})] \tag{2.12}$$

with $\widetilde{\mathbf{Z}}_i = (\tilde{Z}_{i1}, \ldots, \tilde{Z}_{id})^T, i = 1, 2$ being two independent copies of $N_d(0, \mathbf{\Sigma}_Z)$.

The Kendall's tau τ_{jk} in (2.12) is invariant under strictly increasing marginal transformations. This promises that using the observed data $\mathbf{Z}$ will return the

same correlation matrix as using the observed $\widetilde{\mathbf{Z}}$. Specifically, we have

$$\widehat{\tau}_{jk} = \frac{2}{n(n-1)} \sum_{1 \le i_1 \le i_2 \le n} \mathrm{sgn}(\tilde{Z}_{i_1,j} - \tilde{Z}_{i_2,j})\mathrm{sgn}(\tilde{Z}_{i_1,k} - \tilde{Z}_{i_2,k})$$

$$= \frac{2}{n(n-1)} \sum_{1 \le i_1 \le i_2 \le n} \mathrm{sgn}(Z_{i_1,j} - Z_{i_2,j})\mathrm{sgn}(Z_{i_1,k} - Z_{i_2,k}), \quad 1 \le j, k \le d.$$

$$(2.13)$$

Using the Kendall's tau correlation estimator, we obtain the following estimator for the correlation matrix $\mathbf{\Sigma_Z}$,

$$\widehat{\mathbf{\Sigma}}_{\mathbf{Z}} = (\widehat{\sigma}_{jk}) \text{ with } \widehat{\sigma}_{jk} = \sin(\frac{\pi}{2}\widehat{\tau}_{jk}). \tag{2.14}$$

We divide $\widehat{\mathbf{\Sigma}}_{\mathbf{Z}}$ into four sub-matrices, denoted by $\Sigma_{XX}, \Sigma_{XY}, \Sigma_{YX}, \Sigma_{YY}$, and their corresponding Kendall's tau based estimators are $\widehat{\Sigma}_{XX}, \widehat{\Sigma}_{XY}, \widehat{\Sigma}_{YX}, \widehat{\Sigma}_{YY}$. These covariance estimators will be essential in the estimation of $\boldsymbol{\beta}$. However, keep in mind that here we assume $\mathrm{diag}(\mathbf{\Sigma}_Z) = \mathbf{1}$, so the correlation matrix is the same as covariance matrix. Since the original response envelope model is not invariant to rescaling of variables, we use the scaled envelope response model.

2.4.4 Estimation of $\boldsymbol{\beta}$

The goal of envelope methods is to increase the efficiency in multivariate regression estimation and prediction by exploiting variation in the data that is effectively immaterial. In the setting of Gaussian copula model, if $\widetilde{\mathbf{Z}}$ were available, the standard envelope response model can be used. However, as shown in previous examples, some simple changes such as rescaling and log transformation to the variables can result in serious changes in the envelope structure, even the dimension of envelope can change. We assume the following sparse

response envelope exists between the unobserved $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$.:

$$\widetilde{\mathbf{Y}} = \alpha + \Lambda\boldsymbol{\Gamma}\boldsymbol{\eta}\widetilde{\mathbf{X}} + \boldsymbol{\epsilon}, \ \ \text{with} \ \ \widetilde{\boldsymbol{\Sigma}} = \Lambda\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\Lambda + \Lambda\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T\Lambda. \qquad (2.15)$$

The scaled response $\Lambda^{-1}\widetilde{\mathbf{Y}}$ follows an envelope model with u-dimensional envelope $\varepsilon_{\Lambda^{-1}\boldsymbol{\Sigma}\Lambda^{-1}}(\Lambda^{-1}\mathcal{B})$ and semi-orthogonal basis matrix $\boldsymbol{\Gamma}$. The material and immaterial pairs of $\mathbf{Y}$ are $\Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$, respectively. The maximum likelihood estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ of Λ and Γ, respectively, can be obtained by minimizing the objective function

$$L_u(\Lambda, \boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^T\Lambda^{-1}\Sigma_{\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}}\Lambda^{-1}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^T\Lambda\Sigma_{\widetilde{\mathbf{Y}}}^{-1}\Lambda\boldsymbol{\Gamma}|.$$

The quantities $\Sigma_{\widetilde{\mathbf{Y}}}$ and $\Sigma_{\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}}$ can be estimated using the Kendall's τ estimator as discussed in Section 2.4.3. Specifically, we use $\widehat{\Sigma}_{YY}$ and $\widehat{\Sigma}_{YY} - \widehat{\Sigma}_{YX}\widehat{\Sigma}_{XX}^{-1}\widehat{\Sigma}_{XY}$ as estimators for $\Sigma_{\widetilde{\mathbf{Y}}}$ and $\Sigma_{\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}}$ respectively. By substituting them into the objective function, we have

$$\begin{aligned}
\Lambda = \arg\min_{\Lambda, \boldsymbol{\Gamma}} & \log|\boldsymbol{\Gamma}^T\Lambda^{-1}(\widehat{\Sigma}_{YY} - \widehat{\Sigma}_{YX}\widehat{\Sigma}_{XX}^{-1}\widehat{\Sigma}_{XY})\Lambda^{-1}\boldsymbol{\Gamma}| \\
& + \log|\boldsymbol{\Gamma}^T\Lambda\widehat{\Sigma}_{YY}^{-1}\Lambda\boldsymbol{\Gamma}|.
\end{aligned} \qquad (2.16)$$

Note that Kendall's tau based correlation matrix may not be positive semidefinite and we do need to invert some of these matrices in (2.16). Take $\widehat{\Sigma}_{XX}$ for example, we can project $\widehat{\Sigma}_{XX}$ onto the cone of the positive semidefinite matrices. This can be done using the following convex optimization problem

$$\widehat{\Sigma}_{XX}^+ = \arg\min_{\Sigma \geq 0} ||\widehat{\Sigma}_{XX} - \Sigma||_{2,s}.$$

The $||\cdot||_{2,s}$ norm is used here is mostly for theoretical considerations, and other norms such as the spectral norm and $l_{\max}$ norm can also be used, and

they differ very little in practice. Define

$$\widehat{\boldsymbol{\alpha}} = \bar{\mathbf{Y}};$$

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} \widetilde{\boldsymbol{\beta}},$$

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\eta}},$$

$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} (\widehat{\Sigma}_{YY} - \widehat{\Sigma}_{YX} \widehat{\Sigma}_{XX}^{-1} \widehat{\Sigma}_{XY}) \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}},$$

$$\widehat{\boldsymbol{\Gamma}}_0^T = \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^{-1} \widehat{\Sigma}_{YY}^{-1} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}}_0,$$

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Omega}} \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^T + \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Omega}}_0 \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^T.$$

where $\widetilde{\boldsymbol{\beta}} = (\widehat{\Sigma}_{XX})^{-1} \widehat{\Sigma}_{XY}$ is the Kendall's tau based coefficient estimator. When solving (2.16), the objective function almost always has a unique pair as the global minimizer. However, occasionally $\boldsymbol{\Gamma}$ and $\mathrm{span}(\Gamma)$ are not identifiable. When this happens, objective function will be flat along some directions. However, this potential non-uniqueness is not an issue since the parameters we are interested in $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are always uniquely defined. This implies that we obtain the same estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Sigma}$ whether the global minimizer $\{\widehat{\Lambda}, \mathrm{span}(\widehat{\Gamma})\}$ is unique or not.

2.4.5   Selection of u

The dimension of the envelope u is an important parameter in determining the right envelope structure. Likelihood-based methods such as AIC, BIC or other information criteria can be used to select the dimension u.

Before we introduce the AIC estimator, let us first count the number of parameters in the estimation problem. For a copula response envelope model with dimension u, we need r parameters for $\boldsymbol{\alpha}$, $(r-1)$ parameters for Λ, pu parameters for η, $u(u+1)/2$ parameters for $\boldsymbol{\Omega}$, and $(r-u)(r-u+1)/2$

parameters for $\boldsymbol{\Omega}_0$. For the envelope, $\boldsymbol{\Gamma}$ itself is not identifiable, but span($\boldsymbol{\Gamma}$) needs $u(r-u)$ parameters. The total number of parameters is $N(u) = 2r - 1 + pu + r(r+1/2)$.

The AIC estimator of u is $\arg\min -2\widehat{L}(u) + 2N(u)$, where the minimum is taken over the set of integers. The maximized log likelihood under the copula response envelope model with dimension u takes the following form

$$\widehat{L}(u) = -\frac{nr}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Gamma}^T\widehat{\Lambda}^{-1}(\widehat{\Sigma}_{YY} - \widehat{\Sigma}_{YX}\widehat{\Sigma}_{XX}^{-1}\widehat{\Sigma}_{XY})\widehat{\Lambda}^{-1}\boldsymbol{\Gamma}|$$
$$-\frac{n}{2}\log|\widehat{\Sigma}_{YY}^{-1}| - \frac{n}{2}\log|\boldsymbol{\Gamma}^T\widehat{\Lambda}\widehat{\Sigma}_{YY}^{-1}\widehat{\Lambda}\boldsymbol{\Gamma}|,$$

where $\widehat{\Lambda}$ and span($\widehat{\Gamma}$) are the maximum likelihood estimators. The BIC estimator of u is $\arg\min -2\widehat{L}(u) + \log(n)N(u)$.

2.5 Theoretical Results

The proposed copula response envelope model enjoys most of the theoretical qualities similar to the scaled response envelope model since the objective functions are essentially the same. Here we establish the consistency and asymptotic normality of the proposed estimator.

We introduce the operator vec : $\mathbb{R}^{a\times b} \to \mathbb{R}^{ab}$ stacks the columns of a matrix, and the operator vech : $\mathbb{R}^{a\times a} \to \mathbb{R}^{a(a+1)/2}$ stacks the lower triangular part of a symmetric matrix. We combine the parameters in model (2.15) into the vector $\phi = \{\lambda^T, \text{vec}(\eta)^T, \text{vec}(\boldsymbol{\Gamma})^T, \text{vech}(\boldsymbol{\Omega})^T, \text{vech}(\boldsymbol{\Omega}_0)^T\}^T = (\lambda^T, \phi_0^T)^T$, where $\phi_0 = \{\text{vec}(\eta)^T, \text{vec}(\boldsymbol{\Gamma})^T, \text{vech}(\boldsymbol{\Omega})\}^T$ contains the constituent parameters and $\lambda = (\lambda_2, \ldots, \lambda_r)^T$ is the vector of the 2nd to the rth diagonal elements of Λ. Both $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are functions of ϕ. First we establish the identifiability of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

Theorem 2.5.1 *Assume that (2.15) has independent but not necessarily normal errors with finite second moments, and that $n^{-1}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i^T > 0$. Then $\boldsymbol{\beta}(\phi)$ and $\boldsymbol{\Sigma}(\phi)$ are identifiable and $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are uniquely defined.*

As we mentioned earlier, sometimes $\hat{\phi}$ might be trapped in local minimum during the optimization, but $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ will still be identifiable.

The copula response envelope model in (2.15) is similar to a regular envelope model with response $\Lambda^{-1}\mathbf{Y}$, which we use the subscript o to denote. Based on Cook et al. (2010)[11] , the gradient matrix $G_o = \partial\{\text{vec}(\boldsymbol{\beta}_o)^T, \text{vech}(\boldsymbol{\Sigma}_o)^T\}^T/\partial\phi_0^T$ for model (2.4) has dimension $\{pr + r(r+1)/2\} \times \{pu + r(r+1)/2\}$ and has the following form

$$\begin{pmatrix} I_p \otimes \boldsymbol{\Gamma} & \boldsymbol{\eta}^T \otimes I_r & 0 & 0 \\ 0 & 2C_r(\boldsymbol{\Gamma\Omega} \otimes I_r - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T & C - r(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})E_u) & C_r(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)E_{r-u} \end{pmatrix}.$$

The Fisher information for $\{\text{vec}(\boldsymbol{\beta}_o)^T, \text{vech}(\boldsymbol{\Sigma}_o)^T\}^T$ is the $\{rp + r(r+1)/2\} \times \{rp + r(r+1)/2\}$ block diagonal matrix $J_o = \text{bdiag}\{\boldsymbol{\Sigma}_\mathbf{X} \otimes \boldsymbol{\Sigma}_o^{-1}, 0.5E_r^T(\boldsymbol{\Sigma}_o^{-1} \otimes \boldsymbol{\Sigma}_o^{-1})E_r\}$, where bdiag(.) indicates a block diagonal matrix with the diagonal blocks as arguments. Let $h_o = \{(\boldsymbol{\beta}_o \otimes I_r), 2(\boldsymbol{\Sigma}_o \otimes I_r)C_r^T\}^T$, which is the gradient component $h_o = \partial\{\text{vec}(\boldsymbol{\beta})^T, \text{vech}(\boldsymbol{\Sigma})^T\}^T/\partial\Lambda$ for the copula model evaluated at $\Lambda = I_r$. Let $A_o = Q_{G_o(J_o)}h_o L$ and let $D_\Lambda = \text{bdiag}\{I_p \otimes \Delta, C_r(\Lambda \otimes \Lambda)E_r\}$, which is a block diagonal matrix with the same dimensions as J_o.

The gradient matrix $H = \partial\{\text{vec}(\boldsymbol{\beta})^T, \text{vech}(\boldsymbol{\Sigma})^T\}^T/\partial\phi_0^T$ for the copula response model (2.15) has dimension $\{pr + r(r+1)/2\} \times \{r - 1 + pu + r(r+1)/2\}$ and can be represented as $H = \{D_\Lambda h_o(I_r \otimes \Lambda^{-1})L, D_\Lambda G_o\}$. The Fisher information J under the copula response envelope model can be obtained by replacing $\boldsymbol{\Sigma}_o$ with $\boldsymbol{\Sigma}$ in J_o, $J = \text{bdiag}\{\boldsymbol{\Sigma}_\mathbf{X} \otimes \boldsymbol{\Sigma}^{-1}, 0.5E_r^T(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1})E_r\}$. Now we establish the asymptotic normality of the proposed Gaussian copula response

29

envelope model.

Theorem 2.5.2 *Under model (2.15) with normal errors, assume $\max_{i \leq n} p_{ij} \to 0$ as $n \to \infty$. Then $\sqrt{n}[\{vec(\widehat{\boldsymbol{\beta}} - vec(\boldsymbol{\beta}))\}^T, \{vech(\widehat{\boldsymbol{\Sigma}} - vech(\boldsymbol{\Sigma}))\}]^T$ converges in distribution to a normal random vector with mean zero and covariance matrix*

$$V = H(H^T J H)^\dagger H^T = D_\Lambda \{A_o(A_o^T J_o A_o)^\dagger\} D_\Lambda + D_\Lambda \{G_o(G_o^T J_o G_o)^\dagger\} D_\Lambda = V_1 + V_2,$$

where $V_1 = D_\Lambda \{A_o(A_o^T J_o A_o)^\dagger\} D_\Lambda$ and $V_2 = D_\Lambda \{G_o(G_o^T J_o G_o)^\dagger\} D_\Lambda$.

Consequently, we have

Corollary 2.5.2.1 *Assume that the conditions in Theorem (2.5.2) hold. Then the copula envelope model (2.15) is asymptotically more efficient or as efficient as the standard model in estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.*

The component V_1 in the asymptotic covariance matrix can be interpreted as the asymptotic cost of estimating Λ.

2.6   Simulations and Data Analysis

2.6.1   Simulations

Similar to the scaled response envelope model, we adopt an alternating algorithm when estimating $\boldsymbol{\Gamma}$ and Λ.

To evaluate the finite sample size performance of the proposed Gaussian copula response envelope model, we conducted simulation studies under different circumstances.

Following the settings as in Cook and Su (2013)[12], we set $r = 10, u = 5$ and $p = 5$. Each element in $\mathbf{X}$ were generated independently from a $N(0, 1)$.

For the covariance matrix, we took $\mathbf{\Omega} = \sigma^2 I_5$ and $\mathbf{\Omega}_0 = \sigma_0^2 I_5$. The matrix $\boldsymbol{\eta}$ was generated as a 5×5 matrix of independent $N(0,2)$ random variables, and $\boldsymbol{\Gamma}$ was obtained by orthogonalizing a 10×5 matrix of independent $U(0,1)$ random variables. We took σ^2 as 0.25 and σ_0^2 as 5 and 25. The scale matrix $\boldsymbol{\Gamma}$ as taken as $\text{diag}(\boldsymbol{\Sigma})^{-0.5}$. The sample sizes were 100, 200, 300, 500, 800, 1200 and 200 replicates were used for each sample size. After generating Σ based on the above descriptions, we then obtain n samples of $\widetilde{\mathbf{Y}}_i \sim N_r(0, \boldsymbol{\Sigma})$. We consider two settings, for the first setting, we set $Y_{ij} = exp(\tilde{Y}_{ij})$ and $X_{ij} = 3 * \tilde{X}_{ij} - 10$ for $j = 1, \ldots, 5$; $Y_{ij} = 2 * \tilde{Y}_{ij} + 5$ and $X_{ij} = exp(\tilde{X}_{ij})$ for $j = 6, \ldots, 10$. For the second setting, we set $Y_{ij} = \Phi(\tilde{Y}_{ij})$ and $X_{ij} = \tilde{X}_{ij}^2 + 5$ for $j = 1, \ldots, 5$; $Y_{ij} = \tilde{Y}_{ij}^3 - 10$ and $X_{ij} = \Phi(\tilde{X}_{ij})$ for $j = 6, \ldots, 10$.

We compared the performance of (1) original response envelope, (2) scaled response envelope and (3) the proposed Gaussian copula response envelope using either original data $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ or the observed data $(\mathbf{Y}, \mathbf{X})$. We first report the angle between the true envelope space and the estimated envelope space using different envelope models. The angle $\angle\{\text{span}(\mathbf{A}_1), \text{span}(\mathbf{A}_2)\}$ between the subspaces spanned by columns of the semi-orthogonal basis matrices $\mathbf{A}_1 \in \mathbb{R}^{r \times u}$ and $\mathbf{A}_2 \in \mathbb{R}^{r \times u}$ was computed in degrees as the arc cosine of the smallest absolute singular value of $\mathbf{A}_1^T \mathbf{A}_2$.

Tables (2.1) and (2.2) report the results related to the angles under setting 1, and Tables (2.3) and (2.4) report the results related to the angles under setting 2. We have the following findings: 1) both the original and scaled response envelope model fail to detect the envelope structure given the transformed data, while the proposed copula response envelope model successfully detect it; 2) with the increase of sample size, unsurprisingly the proposed cop-

Table 2.1: Response Envelope Models Setting 1 with $\sigma_0^2 = 5$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
	$n = 100$	79.35	6.21	6.34
	$n = 200$	75.45	4.34	4.56
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 300$	70.22	3.21	3.19
	$n = 500$	73.32	2.22	2.33
	$n = 800$	68.43	1.05	1.24
	$n = 1200$	69.32	0.93	0.98
	$n = 100$	78.54	78.45	6.36
	$n = 200$	79.32	68.43	4.55
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 300$	68.34	56.43	3.22
	$n = 500$	66.24	42.34	2.35
	$n = 800$	64.15	41.22	1.21
	$n = 1200$	69.32	39.24	0.99

Table 2.2: Response Envelope Models Setting 1 with $\sigma_0^2 = 25$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
	$n = 100$	74.15	2.94	3.14
	$n = 200$	70.12	2.14	2.16
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 300$	68.29	1.15	2.26
	$n = 500$	71.81	1.02	1.63
	$n = 800$	69.27	0.85	0.94
	$n = 1200$	68.36	0.76	0.78
	$n = 100$	74.13	71.39	3.16
	$n = 200$	72.13	69.23	2.17
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 300$	71.52	64.41	2.31
	$n = 500$	69.58	53.14	1.63
	$n = 800$	66.21	45.27	0.98
	$n = 1200$	67.38	38.14	0.78

Table 2.3: Response Envelope Models Setting 2 with $\sigma_0^2 = 5$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 100$	71.37	4.16	4.45
	$n = 200$	68.55	3.34	3.46
	$n = 300$	69.42	2.18	2.17
	$n = 500$	65.78	1.51	1.58
	$n = 800$	68.19	1.01	1.03
	$n = 1200$	67.27	0.87	0.89
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 100$	78.54	78.45	4.64
	$n = 200$	69.32	68.43	3.45
	$n = 300$	65.34	56.43	2.22
	$n = 500$	66.24	42.34	1.59
	$n = 800$	64.15	41.22	1.01
	$n = 1200$	69.32	39.24	0.90

Table 2.4: Response Envelope Models Setting 2 with $\sigma_0^2 = 25$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 100$	70.67	3.27	3.56
	$n = 200$	67.25	2.64	2.66
	$n = 300$	66.12	1.99	2.01
	$n = 500$	58.12	1.78	1.79
	$n = 800$	61.23	1.03	1.09
	$n = 1200$	60.87	0.68	0.68
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 100$	80.46	73.46	3.55
	$n = 200$	78.43	69.13	2.68
	$n = 300$	71.35	65.36	2.01
	$n = 500$	76.21	62.46	1.77
	$n = 800$	74.36	56.37	1.10
	$n = 1200$	78.12	59.14	0.68

Table 2.5: Response Envelope Models Setting 1: Comparisons of the standard deviation of $\widehat{\boldsymbol{\beta}}$ between copula response envelope estimator and standard model estimator. The results are based on 200 replications.

$\log(\mathrm{SD}(\widehat{\beta}))$	n	Standard model	Actual Copula	Bootstrap
	$n = 100$	-1.9	-2.5	-2.6
	$n = 200$	-2.2	-3.1	-3.2
$\sigma_0^2 = 5$	$n = 300$	-2.6	-3.4	-3.5
	$n = 500$	-2.9	-3.9	-4.0
	$n = 800$	-3.2	-4.1	-4.2
	$n = 1200$	-3.6	-4.3	-4.4
	$n = 100$	-1.3	-2.0	-2.1
	$n = 200$	-1.5	-2.1	-2.2
$\sigma_0^2 = 25$	$n = 300$	-1.6	-2.3	-2.4
	$n = 500$	-2.1	-2.6	-2.6
	$n = 800$	-2.5	-3.2	-3.3
	$n = 1200$	-3.1	-3.6	-3.7

ula response envelope model has better performance; 3) when the difference between the material part and immaterial part is bigger ($\sigma_0^2 = 25$), it is easier to uncover the envelope structure; 4) as long as the transformations introduced here are strictly monotone, the proposed copula envelope model can detect the envelope structure without any difficulty. For the next part, we evaluated the amount of efficiency gains that can be achieved using the proposed copula envelope model compared with the standard least squares estimator. For each realization, the standard deviation of each element in $\widehat{\boldsymbol{\beta}}$ over the replicates is computed, which we call the actual standard deviation of the elements in $\widehat{\boldsymbol{\beta}}$. We also computed the bootstrapped standard deviations by bootstrapping the residuals 200 times.

Tables (2.5) and (2.6) report the log-scale comparisons of the copula response envelope estimator and the standard least squares estimator for one

Table 2.6: Response Envelope Models Setting 2: Comparisons of the standard deviation of $\widehat{\boldsymbol{\beta}}$ between copula response envelope estimator and standard model estimator. The results are based on 200 replications.

$\log(\mathrm{SD}(\widehat{\beta}))$	n	Standard model	Actual Copula	Bootstrap
	$n = 100$	-1.8	-2.5	-2.6
	$n = 200$	-2.2	-3.1	-3.2
	$n = 300$	-2.7	-3.2	-3.3
$\sigma_0^2 = 5$	$n = 500$	-2.9	-3.8	-3.9
	$n = 800$	-3.1	-4.1	-4.2
	$n = 1200$	-3.7	-4.2	-4.4
	$n = 100$	-1.4	-2.0	-2.1
	$n = 200$	-1.5	-2.1	-2.2
	$n = 300$	-1.7	-2.4	-2.5
$\sigma_0^2 = 25$	$n = 500$	-2.1	-2.6	-2.6
	$n = 800$	-2.6	-3.3	-3.5
	$n = 1200$	-3.1	-3.6	-3.7

element in $\widehat{\boldsymbol{\beta}}$. We report the following findings: (1) the proposed copula response envelope estimator achieves significant efficiency gain compared to the standard least squares estimator; (2) the efficiency gain remains roughly constant as the sample size increases; (3) the bootstrapped standard deviation is a good estimator of the actual standard deviation; (4) settings (1) and (2) result in very similar numerical performances, suggesting that the actual transformation used here are irrelevant as long as they are monotone and continuous.

Even though the envelope structure posed on the model does not depend on the normality assumption of the error, the likelihood we used in the estimation process relies on the assumption of normality. Here we examined the performance of the proposed copula response envelope model under different error distributions, specifically, we considered centered and consistently scaled t_6, $U(0,1)$ and χ_4^2 distributions to represent cases with longer tails, shorter

Table 2.7: Response Envelope Models Setting 1: Comparisons of the actual standard deviation of $\widehat{\boldsymbol{\beta}}$ under different error distribution. The results are based on 200 replications.

$\log(\mathrm{SD}(\widehat{\beta}))$	n	Normal	t_6	Uniform	χ_4^2
	$n = 100$	-2.5	-2. 6	-2.7	-2.8
	$n = 200$	-3.1	-3.2	-3.3	-3.2
	$n = 300$	-3.2	-3.4	-3.1	-3.1
$\sigma_0^2 = 5$	$n = 500$	-3.8	-4.0	-3.9	-3.9
	$n = 800$	-4.1	-4.2	-4.2	-4.3
	$n = 1200$	-4.2	-4.4	-4.4	-4.2
	$n = 100$	-2.0	-2.1	-2.4	-2.5
	$n = 200$	-2.1	-2.3	-2.2	-2.1
	$n = 300$	-2.4	-2.5	-2.6	-2.5
$\sigma_0^2 = 25$	$n = 500$	-2.6	-2.5	-2.2	-2.1
	$n = 800$	-3.3	-3.2	-3.6	-3.4
	$n = 1200$	-3.6	-3.7	-3.8	-3.9

tails and skewness respectively.

Table (2.7) and (2.8) report the actual and bootstrapped standard deviation of an element of $\boldsymbol{\beta}$ using the proposed copula response envelope model under different error distributions . We have the following findings: (1) overall there exist no significant differences for different error distributions, which means the proposed copula response envelope model is robust to moderate deviation from the normality assumption; (2) the bootstrapped standard deviation still a reasonable estimate of the actual standard deviation under moderate deviation from the normality.

In summary, comparing these different envelope response models, we have the following conclusions: using either the original data $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ or the observed data $(\mathbf{Y}, \mathbf{X})$, the original response envelope model failed to pick up the right envelope structure and reduced to the standard estimator, which means no

Table 2.8: Response Envelope Models Setting 1: Comparisons of the boot-strapped standard deviation of $\widehat{\boldsymbol{\beta}}$ under different error distribution. The results are based on 200 replications.

$\log(\text{SD}(\widehat{\beta}))$	n	Normal	t_6	Uniform	χ_4^2
	$n = 100$	-2.6	-2. 6	-2.7	-2.7
	$n = 200$	-3.2	-3.3	-3.1	-3.3
$\sigma_0^2 = 5$	$n = 300$	-3.3	-3.5	-3.2	-3.3
	$n = 500$	-3.9	-4.1	-3.8	-4.0
	$n = 800$	-4.2	-4.1	-4.3	-4.4
	$n = 1200$	-4.4	-4.3	-4.2	-4.3
	$n = 100$	-2.1	-2.2	-2.4	-2.1
	$n = 200$	-2.2	-2.5	-2.2	-2.4
$\sigma_0^2 = 25$	$n = 300$	-2.5	-2.4	-2.6	-2.5
	$n = 500$	-2.6	-2.7	-2.2	-2.5
	$n = 800$	-3.5	-3.4	-3.6	-3.5
	$n = 1200$	-3.7	-3.8	-3.8	-3.9

efficiency gain was obtained. The scaled response envelope model was able to estimate the right envelope structure using the original data $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ (not observed) but not using the observed data $(\mathbf{Y}, \mathbf{X})$. The proposed Gaussian copula response envelope model was able to pick up the correct envelope structure using either the original data $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ or the observed data $(\mathbf{Y}, \mathbf{X})$.

2.6.2 Data Analysis

We apply the proposed copula response envelope model to an ADHD study to examine the relationship between brain connectivity and a few demographic variables including age, gender and the ADHD status.

Attention deficit hyperactivity disorder (ADHD) is one of the most common childhood neuropsychiatric disorders. The psychopathology of ADHD is marked by developmentally inappropriate and pervasive expressions of inattention, overactivity, and impulsiveness, and it often persists into adulthood. The

understanding of the underlying pathophysiology of neuropsychiatric illnesses remains insufficient [55], and clinically useful biomarkers are rarely attained for ADHD [56]. Recent studies have demonstrated the potential of medical imaging such as functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (DTI) in predicting patient outcomes and understanding the underlying pathophysiology of diseases [57, 58, 59].

The data used here is the publicly available resting-state fMRI (rs-fMRI) data from the ADHD-200 Consortium [60]. fMRI is a neuroimaging procedure that measures brain activity by detecting changes associated with blood flow, and rs-fMRI is acquired when a subject is not performing an explicit task. rs-fMRI is useful for exploring the brain's functional organization and determining whether it is altered in neurological or psychiatric diseases. The data set contains 120 subjects ($n = 120$) from the NYU site (New York University Child Study Center) of the ADHD-200 Consortium. The data were preprocessed through the Athena pipeline [61] and are region of interest (ROI) based. The Anatomical Automatic Labeling (AAL) atlas [62] was used for the parcellation. For each subject, there are 172 time courses and the AAL has 116 ROIs. The cerebra include 90 regions (45 in each hemisphere), and the cerebellar include 26 regions (9 in each cerebellar hemisphere and 8 in the vermis). Of the 120 subjects, 42 are typically developing children and 78 are diagnosed as ADHD. The ADHD group is further separated into ADHD-Combined ($n = 33$) and ADHD-Inattentive ($n = 45$). Table 2 gives the demographic characteristics and neuropsychological scores of the subjects analyzed in this study. The verbal IQ scores measure general knowledge, language, reasoning, and memory skills, while the performance IQ scores measure spatial, sequencing, and

Table 2.9: Characteristics of ADHD subjects and healthy controls

Characteristics	ADHD (n = 78)	Controls (n = 42)	p-value
Gender (female/male)	41/37	27/15	0.297 [a]
Age (year)	9.75 ± 0.37	10.93 ±0.55	<0.0005 [b]
Verbal IQ	107.37± 2.31	114.31 ± 3.11	<0.0005 [b]
Performance IQ	110.85 ± 3.31	106.81 ± 4.07	0.1255 [b]
Full-scale IQ	110.14 ±2.56	111.69 ±3.39	0.4656 [b]

[a] The p value was obtained by χ^2 test.
[b] The p value was obtained by two-sample two-tailed t-test.

problem-solving skills. The verbal and performance IQ scores are summed and converted to obtain the full-scale IQ scores. No significant differences in gender, performance IQ, or full-scale IQ were found, and a significantly higher age and verbal IQ were found in the healthy control subjects.

For each subject, we obtained the mean time series for each of the 116 regions by averaging the fMRI time series over all voxels in the region. We computed partial correlation coefficients between each pair of ROIs. Each partial correlation measures the degree of association between two regions while controlling the effect of the remaining regions. In the variable selection, the goal is to find the significant brain functional connectivity that contribute to the level of ADHD. Each partial correlation is considered to be a predictor, so initially we have $p = (116 \times 116 - 116)/2 = 6670$. Figure 1 shows the histograms of the mean partial correlations between ROIs in ADHD subjects and controls.

Because of the large p small n scenario ($n = 120$, $p = 6670$), we first applied a variable screening to remove some partial correlations that are not significant. We used a Fisher's r-to-z transformation to improve the normality

of these partial correlation coefficients. We used a two-tailed t test between the z values of the ADHD group and the control group to determine whether the functional connections are different. The selected significant functional correlations between the ADHD subjects and the control must satisfy two criteria: (1) significantly different z values at the threshold of $p < 0.001$; (2) z values for the correlations that are significantly different from zero in at least one group at the threshold of $p < 0.001$. We did not apply a multiple-testing p-value correction because the screening step is a preliminary step, and a more complicated analysis will be applied using the proposed variable selection algorithms. After the screening, we selected $p = 12$ functional connections for the robust variable selection.

Due to the strong correlation existing among the functional connectivity between the regions, there are potentially lots of room for efficiency gain. We take the selected 12 functional connections as the response variables, for the covariates we consider gender, age, verbal IQ, performance IQ, and ADHD index, which is a measurement of the overall level of ADHD symptoms. It is a continuous variable ranging from 40 to 90, and typically developing children usually have a score below 50. This variable is more informative than the ordinal ADHD diagnosis result. Typically the ADHD index is log transformed, however, in our case, since the proposed method is equivariant towards log transformation, we do not need to do that here. We compare the standard errors of the ordinary least squares estimator $\tilde{\boldsymbol{\beta}}$ to the standard errors of the scaled envelope estimator $\widehat{\boldsymbol{\beta}}$ by using the fractions $f_{ij} = 1 - \widehat{\text{avar}}^{1/2}(\sqrt{n}\tilde{\boldsymbol{\beta}}_{ij})/\widehat{\text{avar}}^{1/2}(\sqrt{n}\widehat{\boldsymbol{\beta}}_{ij})$, where the subscripts i, j indicate the elements of the estimator of $\boldsymbol{\beta}$.

We first fit an ordinary envelope model to the data and BIC suggested that

$u = 9$. Compared to $\tilde{\boldsymbol{\beta}}$, the standard deviations of the elements in the ordinary envelope estimator were 2.0% to 15.6% smaller. A sample size of about $n = 100$ observations would be needed to reduce the standard error of the ordinary least squares estimator by 15.6%. Then we fit a scaled envelope model to the data and BIC suggested that $u = 7$. Compared to $\tilde{\boldsymbol{\beta}}$, the standard deviations of the elements in the scaled envelope estimator were 5.0% to 35.4% smaller. A sample size of about $n = 100$ observations would be needed to reduce the standard error of the ordinary least squares estimator by 35.4%. At last we fit the proposed copula envelope model to the data and BIC suggested that $u = 5$. Compared to $\tilde{\boldsymbol{\beta}}$, the standard deviations of the elements in the scaled envelope estimator were 7.0% to 64.7% smaller. A sample size of about $n = 100$ observations would be needed to reduce the standard error of the ordinary least squares estimator by 64.7%. Based on these empirical results, it seems the proposed copula response envelope model has the potential to offer more efficiency gains.

2.7 Discussion

Multivariate linear regression is one of the most fundamental tools in data analysis. Response envelope model offers a great framework to potentially utilize the dependencies between the response variables to obtain more efficient estimation for the coefficient $\boldsymbol{\beta}$. The linear combinations of Y that are useful to the regression are called material parts. The efficiency gains can be quite substantial. However, the original envelope model is quite strict on the data generating process. For example, the original response envelope model is

not invariant or equivariant on scaling, which can be a challenge in practice, since this means that a simple rescaling can change the envelope structure. To address this issue, scaled envelope model was proposed by Cook and Su (2013)[12]. The proposed scaled envelope model can handle the rescaling of response variables.

Another strong assumption imposed on the model is the linearity between response and covariates. The linearity assumption is often too restrictive in practice. To deal with the nonlinearity, one can either transform the data so that there is linear relationship between the transformed variables or use a non-linear regression method. The use of linear models is greatly enhanced through the use of various transformations of the data. There are a few reasons why a transformation of the data makes sense in practice. For example, sometimes there exists prior knowledge that the data generating process corresponds to a theoretical model, which has a specific parametric form; sometimes the data itself appears a nonlinear pattern between the response and covariates. What kind of data can be transformed into a linear relationship is often hard to determine. Data transformations, such as the Box-Cox transformation, Fisher's z transformation and variance stabilization transformation, have been frequently used to improve a linear fit. However, typically these transformations need to be applied before the regression, and various transformations have to be applied in order to find the right one.

In this chapter, we presented the Gaussian copula response envelope model, which can be seen as a hybrid of the Gaussian copula regression and scaled response envelope model. The proposed model can handle continuous and monotone transformations of both response variables and covariates. The

proposed model is highly useful in practice since it is not necessary to try different transformations now in order to find suitable envelope structures to the data for more efficient estimation. Under the current framework, the transformation has to be continuous and strictly monotone. Possible relaxation of the current method is a generalization to continuous piecewise monotone functions, or even piece-wise continuous functions. For example, when we have a binary response, the current framework can be generalized to envelope logistic regression model. Also, when the response or covariates are censored, the censoring can be expressed using a transformation function or as a nested model. Another example is the use of splines to accommodate more nonlinear effects. If the current class of transformation functions can be generalized to a broader class, many other useful models can be further incorporated, and thus more complex types of data can be handled by response envelope models.

Chapter 3

Gaussian Copula Predictor Envelope Model and Partial Least Squares

3.1 Introduction

In the last chapter we considered the response envelope model and its several variants, where the dimension reduction is applied to $\mathbf{Y}$ relying on the notion of material and immaterial components. However, for most applications in practice, it is the dimension of the covariates $\mathbf{X}$ that needs to reduced. For example, in clinical trails, different information of the trial participants have been recorded including the demographic, clinical, drug-related and psychological related variables. All of them could be potentially related to the many different response variables we are considering in the study. Nowadays especially with the popularity of the use of electronic patient record data, there might be thousands or even hundreds of thousands of variables available for each trial participant, including past drug intakes, hospital admittance and family medical history.

Traditional variable selection methods such as forward selection, backward elimination and best subset selection become computationally expensive or even infeasible at these conditions. To address these problems, a family of penalized least squares based methods has been developed. Examples include Lasso and Adaptive Lasso ([41, 49]), SCAD [19], elastic net [50], and MCP [46]. However, when the dimensionality p is much larger than the sample size n or even grows exponentially with n, the aforementioned penalization methods can perform poorly or even become infeasible due to the simultaneous challenges of computational expediency, statistical accuracy and algorithm stability [21]. For example, in MRI studies, images with dimension $1024 \times 1024 \times 200$ can be acquired for each subject, and due to the high cost of the MRI scanning, studies might only contain less than 100 subjects. If we treat the signal from each voxel as a feature, the dimension of feature space p is much higher than the sample size n.

A natural idea to address these challenges is to reduce the dimensionality p from a large scale to a relatively small scale d using a fast screening algorithm, and then the ultrahigh-dimensional problem can be greatly simplified into a moderately high-dimensional one. Subsequently, the standard penalized variable selection methods can be applied to the remaining variables. Fan and Lv [20] first introduced the sure independence screening (SIS) by ranking the marginal correlation of each covariates and the response. The good numerical performance and novel theoretical properties have made SIS popular in ultrahigh dimensional reduction. As a result, SIS and its extensions have been generalized to many important settings including generalized linear models [22], multi-index semi-parametric models [48], nonparametric regres-

sion [18], quantile regression [24], among others. Other marginal screening methods based on different measure of association between predictors and response have also been studied, such as Kendall's τ [32] and distance correlation [33]. We refer to [34] for a more comprehensive list of references.

Besides the methods mentioned above in variable selection and screening, another type of method to reduce the dimensionality in the predictor space is through sufficient dimension reduction. Specifically, we focus on the use of predictor envelope model in this chapter. Consider the following multivariate linear regression

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\beta}(\mathbf{X} - \mu_{\mathbf{X}}) + \boldsymbol{\epsilon} \tag{3.1}$$

where the error vector $\boldsymbol{\epsilon}$ has mean 0 and covariance matrix $\boldsymbol{\Sigma}$, the random predictor vector $\mathbf{X}$ has mean $\mu_{\mathbf{X}}$ and variance $\boldsymbol{\Sigma}_{\mathbf{X}}$, and $\boldsymbol{\epsilon} \perp \mathbf{X}$. Similar to the response envelope model, the predictor envelope model seeks to find the immaterial information in the predictor space that are not relevant to the estimation of $\boldsymbol{\beta}$. The predictor envelope model is actually tightly connected to the popular partial least squares (PLS) model.

The originally proposed predictor envelope models are not invariant or equivariant under the rescaling of variables, let along other strictly increasing transformations. This brings inconveniences to the practical use of envelope models. For example, it is difficult to ensure all the covariates are measured by the same or similar scales, and sometimes the responses might measure different things, which makes them not comparable. We have observed in practice that such transformations often resulting in envelope based estimators reduce back to the ordinary least square estimators, which means no efficiency gains

can be achieved. The use of envelope models will be affected greatly as certain variables with large variance might dominate the results, just as what we see in methods like the principal component regression, ridge regression and partial least squares. To address this issue, Cook and Su (2016)[13] proposed scaled predictor envelope model, which is invariant to a scaling of the predictors. Similar to the scaled response envelope model, the basic idea is to introduce scaling parameters to estimate the best rescaling of the variables under consideration. Unlike the scaled response envelope model, Cook and Su (2016)[13] allowed groups of predictors to be scaled in the same way, but not individually. This brings some constraints on the applicability of the scaled predictor envelope model. For example, there is no point in rescaling all the predictors in univariate linear regression since the asymptotic variance of the scaled predictor envelope estimator reduces to that of the ordinary least squares estimator. However, the gains can still be substantial in multivariate linear regression and when there are natural intuition to rescale the predictors in group, for example, when several predictors are measured in similar or the same scale.

In this chapter, we further broaden the scope of predictor envelope model and partial least squares. Specifically, we allow the estimators to be equivariant against continuous increasing transformations, which includes a large class of the popularly used transformations such as log transformation, box-cox transformation and variance stabilization transformation. Similar to the proposed Gaussian copula response envelope model in Chapter 2, we combine the framework of Gaussian copula model and scaled predictor envelope model. The proposed Gaussian copula predictor envelope model relaxes the linear assumption between covariates and responses, and it can pick up the envelope

structures when scaled predictor envelope reduces to ordinary least squares. This is also different from the work in both Cai and Zhang (2018)[4] and Zhao and Genest (2019)[47], where a penalized regression framework has been used.

The rest of this chapter is organized as follows. In Sections 3.2 and 3.3, we first introduce the background of predictor envelope model and scaled response envelope model. In Section 3.4, we introduce the proposed Gaussian copula predictor envelope model. The model formulation, estimation, selection of envelope dimension have been established. In Section 3.5, the theoretical properties have been established, including the uniqueness, consistency and asymptotic normality. Simulation studies and data analysis can be found in Section 3.6.

3.2 Predictor Envelope Model and Partial Least Squares

PLS is a popular method in dimension reduction, and it has been widely used in chemistry, agriculture and medicine. Historically, it was defined in terms of the iterative algorithms NIPALS (Wold, 1966[45]) and SIMPLS (de Jong, 1993[16]). Nowadays it has been used as a method to improve the predictive performance of ordinary least square regression. The idea behind PLS is to find a few linear combinations $\mathbf{X} \to \mathbf{\Gamma}^T \mathbf{X}$ that maximizes the covariance with the responses subject to certain constraints. Here $\mathbf{\Gamma} \in \mathbb{R}^{p \times u}$, is a semi-orthogonal matrix that we temporarily assume to be known, and u is called a number of components. Estimation and prediction are then based on the OLS fit of the

reduced model

$$\mathbf{Y} = \boldsymbol{\alpha} + \boldsymbol{\eta}^T \boldsymbol{\Gamma}^T (\mathbf{X} - \mu_{\mathbf{X}}) + \boldsymbol{\epsilon}, \tag{3.2}$$

where the coefficient $\boldsymbol{\eta} \in \mathbb{R}^{u \times r}$ .The PLS estimator of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}_{PLS} = \boldsymbol{\Gamma}\widehat{\boldsymbol{\eta}} = \boldsymbol{\Gamma}(\boldsymbol{\Gamma}^T S_{\mathbf{X}} \boldsymbol{\Gamma})^{-1} \boldsymbol{\Gamma}^T S_{\mathbf{XY}} = \mathbf{P}_{\varepsilon_{S_{\mathbf{X}}}} \widehat{\boldsymbol{\beta}}_{ols}, \tag{3.3}$$

where $\widehat{\boldsymbol{\beta}}_{ols}$ is the OLS estimator of $\boldsymbol{\beta}$. Compared to OLS, PLS has a dimension reduction step, which reduces p predictors $\mathbf{X}$ to u components $\boldsymbol{\Gamma}^T\mathbf{X}$. When $u = p$, $\boldsymbol{\Gamma} = \mathbf{I}_p$ and PLS degenerates to OLS. When $u < p$, PLS often shows better prediction performance over OLS, especially when there is collinearity among the predictors.

The SIMPLS version ([16]) of PLS uses the following algorithm to construct an estimator of $\boldsymbol{\Gamma}$. Set $\widehat{\gamma}_1$ equal to the eigenvector of $S_{\mathbf{XY}} S_{\mathbf{XY}}^T$ corresponding to its largest eigenvalue, and let $\widehat{\boldsymbol{\Gamma}}_k = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_k), k = 1, \ldots, p$. Given $\widehat{\boldsymbol{\Gamma}}_k$ and $k < p$,

$$\widehat{\gamma}_{k+1} = \arg\max_{g} g^T S_{\mathbf{XY}} S_{\mathbf{XY}}^T g, \ \text{s.t.} g^T S_{\mathbf{X}} \widehat{\boldsymbol{\Gamma}}_k = 0 \text{ and } g^T g = 1. \tag{3.4}$$

Then $\widehat{\boldsymbol{\Gamma}}_{PLS} = \widehat{\boldsymbol{\Gamma}}_u$ is the SIMPLS estimator of $\boldsymbol{\Gamma}$.

The predictor envelope model is derived under the linear regression model, but $\mathbf{Y}$ can be either univariate or multivariate. The predictor envelope model is constructed based on a dimension reduction of the predictor space $\mathbf{X}$. Cook et al. (2013)[9] used envelopes to account for immaterial variation in the predictor space, resulting in a different envelope estimator that outperforms OLS and PLS estimators.

Denote $\mathcal{S}$ as a u-dimensional subspace of $\mathbb{R}^p$ with $u < p$. Let $\mathbf{G} \in \mathbb{R}^{p \times u}$ be an orthonormal basis of $\mathcal{S}$ and $\mathbf{G}_0 \in \mathbb{R}^{p \times (p-u)}$ be an orthonormal basis of $\mathcal{S}^{\perp}$,

the orthogonal complement of $\mathcal{S}$. The envelope model arises by imposing the following two conditions on $\mathbf{G}^T\mathbf{X}$ and $\mathbf{G}_0^T\mathbf{X}$:

$$\text{(i) } \mathbf{G}_0^T\mathbf{X}|\mathbf{Y} \sim \mathbf{G}_0^T\mathbf{X} \quad \text{and} \quad \text{(ii) } \text{cov}(\mathbf{G}^X\mathbf{Y}, \mathbf{G}_0^T\mathbf{X}|\mathbf{Y}) = 0. \qquad (3.5)$$

Condition (i) implies that the marginal distribution of $\mathbf{G}_0^T\mathbf{X}$ must be unaffected by changes in $\mathbf{Y}$, and condition (ii) requires that $\mathbf{G}_0^T\mathbf{X}$ must be unaffected by changes in $\mathbf{Y}$ through an association with $\mathbf{G}^T\mathbf{X}$. The above two conditions combined imply that any dependence of $\mathbf{Y}$ on $\mathbf{X}$ must be concentrated on $\mathbf{G}^T\mathbf{X}$, and $\mathbf{G}_0^T\mathbf{X}$ carries no material information for the estimation of $\boldsymbol{\beta}$. The two conditions in (3.5) hold if and only if (a) $\mathcal{B} := \text{span}(\boldsymbol{\beta}) \in \mathcal{S}$ and (b) $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}_\mathbf{X}$. The definition of reducing subspace has been given in Chapter 2.

There are many subspaces that reduce $\boldsymbol{\Sigma}_\mathbf{X}$ and contain $\mathcal{B}$. We define the $\boldsymbol{\Sigma}$-envelope of $\mathcal{B}$ as the smallest reducing subspace of $\boldsymbol{\Sigma}_\mathbf{X}$ that contains $\mathcal{B}$. The smallest here means the intersection of all such subspaces. The $\boldsymbol{\Sigma}$-envelope of $\boldsymbol{\beta}$ is often denoted as $\varepsilon_{\boldsymbol{\Sigma}_\mathbf{X}}(\mathcal{B})$. The multivariate linear regression model (3.1) now can be reparametrized in terms of $\varepsilon_{\boldsymbol{\Sigma}_\mathbf{X}}(\mathcal{B})$ by using a basis. Let $u = \dim(\boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}_\mathbf{X}}(\mathcal{B}))$, and let $\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0 \in \mathbb{R}^{p \times p}$ be orthogonal matrices with $\boldsymbol{\Gamma} \in \mathbb{R}^{p \times p}$, $\text{span}(\boldsymbol{\Gamma}) = \boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}_\mathbf{X}}(\mathcal{B})$, and $\text{span}(\boldsymbol{\Gamma}_0) = \boldsymbol{\mathcal{E}}_{\boldsymbol{\Sigma}_\mathbf{X}}^{\perp}(\mathcal{B})$. Then the predictor envelope copula model can be written as

$$\mathbf{Y} = \alpha + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T(\mathbf{X} - \mu_\mathbf{X}) + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\Sigma}_\mathbf{X} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T, \qquad (3.6)$$

where $\boldsymbol{\beta} = \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T$, and $\boldsymbol{\eta} \in \mathbb{R}^{u \times r}$ is the coordinate of $\boldsymbol{\beta}$ with respect to $\boldsymbol{\Gamma}$. The matrices $\boldsymbol{\Omega} \in \mathbb{R}^{u \times u}$ and $\boldsymbol{\Omega}_0 \in \mathbb{R}^{(p-u) \times (p-u)}$ are the coordinates of $\boldsymbol{\Sigma}$ with respect to $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}_0$, respectively. The variable of interest is still $\boldsymbol{\beta}$,

and we normally do not infer about the constitute parameters $\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$. The values of the $\boldsymbol{\eta}, \boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$ depend on the choice of $\boldsymbol{\Gamma}$. It should also be noted that the basis matrix $\boldsymbol{\Gamma}$ is not identifiable because $\boldsymbol{\Gamma O}$ for any orthogonal matrix $\mathbf{O}$ will lead to an equivalent model. However, the envelope $\varepsilon_\Sigma(\mathcal{B}) = \text{span}(\boldsymbol{\Gamma})$ is identifiable, which is the key for the estimation of $\boldsymbol{\beta}$. Under this formulation, it is also straightforward to see the covariance matrix $\boldsymbol{\Sigma_X}$ can be decomposed as a sum of the covariance matrix of the material part of the response $\text{var}(\mathbf{P}_\varepsilon \mathbf{X}) = \boldsymbol{\Gamma \Omega \Gamma}^T$ and immaterial part of the response $\text{var}(\mathbf{Q}_\varepsilon \mathbf{X}) = \boldsymbol{\Gamma}_0 \boldsymbol{\Omega} \boldsymbol{\Gamma}_0^T$.

The estimation problem of the parameters in (3.6) depends on the dimension of the envelope u. Assuming that u is known and a normal likelihood function, Cook et al. (2013)[9] derived estimation procedures. The log-likelihood for fixed u can be written as

$$L_u(\boldsymbol{\Gamma}) = \log |\boldsymbol{\Gamma}^T S_{\mathbf{X}|\mathbf{Y}} \boldsymbol{\Gamma}| + \log |\boldsymbol{\Gamma} S_{\mathbf{X}}^{-1} \boldsymbol{\Gamma}|. \tag{3.7}$$

After obtaining a minimizer of the above objective function, the remaining parameters are determined as

$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^T S_{\mathbf{X}} \widehat{\boldsymbol{\Gamma}}$$

$$\widehat{\boldsymbol{\Omega}}_0 = \widehat{\boldsymbol{\Gamma}}_0^T S_{\mathbf{X}} \widehat{\boldsymbol{\Gamma}}_0$$

$$\widehat{\boldsymbol{\Sigma_X}} = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Omega}} \widehat{\boldsymbol{\Gamma}}^T + \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Omega}}_0 \widehat{\boldsymbol{\Gamma}}_0^T$$

$$\widehat{\boldsymbol{\eta}} = \widehat{\boldsymbol{\Omega}}^{-1} \widehat{\boldsymbol{\Gamma}}^T S_{\mathbf{XY}}$$

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\eta}}.$$

Assuming that $(\mathbf{X}, \mathbf{Y})$ is multivariate normal, Cook et al. (2013)[9] showed that the likelihood based estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is more efficient or at least as efficient as the OLS estimator asymptotically, and that the efficiency gain can be

substantial when $||\mathbf{\Omega}|| > ||\mathbf{\Omega_0}||$, where $||.||$ is the spectral norm. Additionally, they proved that $\widehat{\boldsymbol{\beta}}$ is a $\sqrt{n}-$consistent estimator under model (3.6) without normality.

PLS and predictor envelope models are closely related. Cook et al. (2013)[9] showed that $\widehat{\mathbf{\Gamma}}_{\mathrm{PLS}}$ is a $\sqrt{n}$- consistent estimator of a basis for $\varepsilon_{\mathbf{\Sigma_X}}(\mathcal{B})$ and that the number of PLS components corresponds to the dimension u of $\varepsilon_{\mathbf{\Sigma_X}}(\mathcal{B})$. The envelope and SIMPLS estimators, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, have the same form and are based on the same population construct $\varepsilon_{\mathbf{\Sigma_X}}(\mathcal{B})$, but differ in their methods of estimating a basis for $\varepsilon_{\mathbf{\Sigma_X}}(\mathcal{B})$. Additionally, $\widehat{\boldsymbol{\beta}}$ typically dominates $\widehat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ in both estimation and prediction and is less sensitive to the number of components selected [9].

The maximum likelihood estimation procedure described above are based on a known u. The dimension of the envelope subspace, u, is a model selection parameter. In practice, the popular ways of selecting u include sequential likelihood ratio test, information criterion such as AIC or BIC, and cross-validation.

3.3 Scaled Predictor Envelope Model

Like PLS and the principal component regression, the predictor envelope model presented in Section 3.2 is not invariant or equivariant under scale transformations. Figure (3.1) shows an example of how rescaling the response can affect an envelope analysis. After the rescaling, all linear combinations of $\mathbf{Y}$ are material to the regression and the envelope model is the same as the standard model and no efficiency gains have been achieved. This could potentially
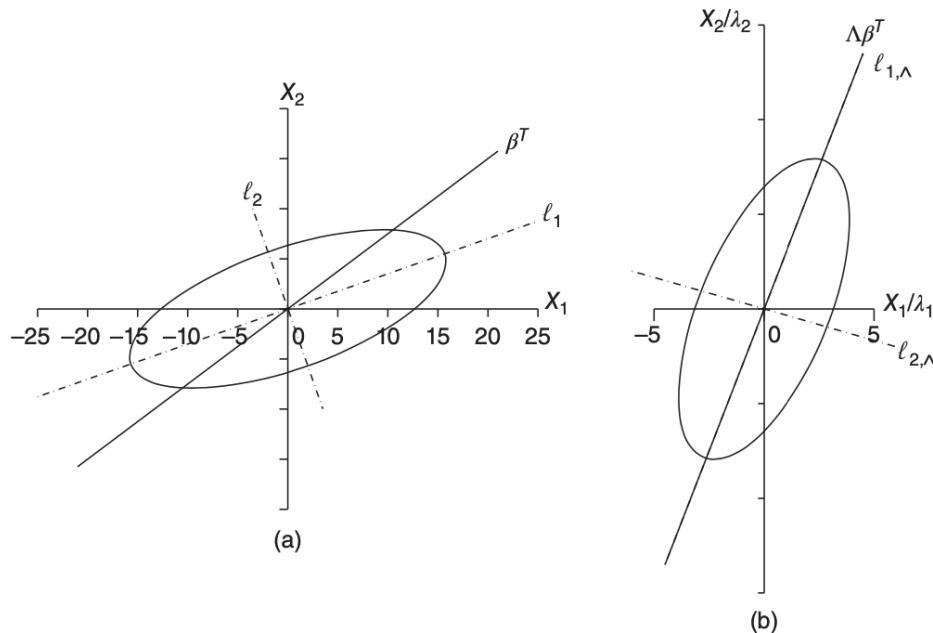
Figure 3.1: Schematic illustration of how rescaling the predictor can affect an envelope analysis. (a) original distributions, (b) rescaled distributions.

greatly limit the use of envelope models in practice since the measuring units might vary greatly for different response variables. Cook and Su (2016)[13] proposed a scaled predictor envelope model to achieve scale invariance for response envelope model, and the details can be seen in Section 2.3. The basic idea for the scaled predictor envelope model is similar to the one in the scaled response envelope model. However, the problem itself is quite different, and the techniques used in both models are different.

In the scaled response envelope model, each response variable is assigned a scaling parameter. In the scaled predictor envelope model, we adopt a different strategy. Consider $\Lambda \in \mathbb{R}^{p \times p}$ to be a diagonal matrix with repeated diagonal elements in blocks $1, \ldots, 1, \lambda_1, \ldots, \lambda_1, \ldots, \lambda_{q-1}, \ldots, \lambda_{q-1}$, where $1, \lambda_1, \ldots, \lambda_{q-1}$

are q positive numbers. Suppose that the i-th of these q scalings has r_i replications, $\sum_{i=1}^{q} r_i = p$. The range of q can go from 1 to p, and the implications of different q will be discussed later in the theoretical results section. Assume that the scaled response vector $\Lambda^{-1}\mathbf{X}$ follows the predictor envelope model (3.6), so the scaled predictor envelope model for the original responses $\mathbf{X}$ becomes

$$\mathbf{Y} = \alpha + \boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\Lambda^{-1}(\mathbf{X} - \mu_\mathbf{X}) + \boldsymbol{\epsilon}, \ \text{ with } \ \boldsymbol{\Sigma}_\mathbf{X} = \Lambda\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^T\Lambda + \Lambda\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T\Lambda. \ (3.8)$$

The scaled response $\Lambda^{-1}\mathbf{X}$ follows an envelope model with u-dimensional envelope $\varepsilon_{\Lambda^{-1}\boldsymbol{\Sigma}\Lambda^{-1}}(\Lambda^{-1}\mathcal{B})$ and semi-orthogonal basis matrix $\boldsymbol{\Gamma}$. The material and immaterial pairs of $\mathbf{X}$ are $\Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$, respectively. To show that model (3.8) follows the same envelope structure as (3.5), we denote $\mathbf{U} = \Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\mathbf{V} = \Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$. It is easy to verify that

$$\text{(i) } \mathbf{VX}|\mathbf{Y} \sim \mathbf{VX} \quad \text{and} \quad \text{(ii) } \text{cov}(\mathbf{UX}, \mathbf{VX}|\mathbf{Y}) = 0.$$

The SPE model has $N(u) = r + p + q - 1 + ur + p(p+1)/2 + r(r+1)/2$ parameters. The maximum likelihood estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ of Λ and Γ can be obtained by minimizing the objective function

$$L_u(\Lambda, \boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^T\Lambda^{-1}(S_\mathbf{X} - S_{\mathbf{XY}}S_\mathbf{Y}^{-1}S_{\mathbf{YX}})\Lambda^{-1}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^T\Lambda S_\mathbf{X}^{-1}\Lambda\boldsymbol{\Gamma}|. \quad (3.9)$$

The optimization of (3.9) can be performed in an alternating fashion. Given a fixed value of Λ, we can obtain the optimal value for $\boldsymbol{\Gamma}$; given the obtained value $\boldsymbol{\Gamma}$, we then obtain the optimal value for Λ. We iterate between $\boldsymbol{\Gamma}$ and Λ until the difference between the objective functions in two adjacent iterations is smaller than a pre-specified value. Once we obtained $\widehat{\boldsymbol{\Gamma}}$ and $\widehat{\Lambda}$, compared with the standard response envelope model, the scaled response envelope model

introduces the new scaling parameter Λ. The maximum likelihood estimators of the remaining parameters are as follows:

$$\widehat{\mu}_{\mathbf{X}} = \bar{\mathbf{X}};$$

$$\widehat{\boldsymbol{\alpha}} = \bar{\mathbf{Y}};$$

$$\widehat{\boldsymbol{\eta}} = (\widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} S_{\mathbf{X}} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}})^{(} - 1) \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} S_{\mathbf{XY}};$$

$$\widehat{\boldsymbol{\beta}} = \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\eta}},$$

$$\widehat{\boldsymbol{\Omega}} = \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^{-1} S_{\mathbf{X}} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}},$$

$$\widehat{\boldsymbol{\Gamma}}_0^T = \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^{-1} S_{\mathbf{X}} \widehat{\Lambda}^{-1} \widehat{\boldsymbol{\Gamma}}_0,$$

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Omega}} \widehat{\boldsymbol{\Gamma}}^T \widehat{\Lambda}^T + \widehat{\Lambda} \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Omega}}_0 \widehat{\boldsymbol{\Gamma}}_0^T \widehat{\Lambda}^T.$$

The algorithms used in minimizing (3.9) are similar to the ones used in the standard response envelope model. The strategies used in choosing the envelope dimension u here are also similar, including likelihood-based methods such as AIC and BIC, and nonparametric methods such as cross validation and permutation test.

A global minimizer in (3.9) is not unique; however, $\text{span}(\widehat{\boldsymbol{\Gamma}})$ is typically unique. Also, both Λ and $\boldsymbol{\Gamma}$ are constituents of the parameters of interest $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$, which are identifiable. The efficiency gain depends on the relative magnitude of $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$.

3.4   Gaussian Copula Predictor Envelope Model

3.4.1   Motivation

One of the key assumptions in the previously introduced predictor envelope model and scaled predictor envelope models is the linearity between the $\mathbf{X}$ and $\mathbf{Y}$, which is often too restrictive and unrealistic. For example, there might not

be an efficient envelope structure among $\mathbf{X}$. However, such structures might exist in transformed $f(\mathbf{X})$, where f includes a large class of popularly used functions such as log function. This greatly enhances the power of predictor envelope model and partial least squares since potential extra dimension reduction can be discovered.

We consider the following model, which has been widely used in a range of applications,

$$g(\mathbf{Y}) = \boldsymbol{\alpha} + \boldsymbol{\beta} f(\mathbf{X}) + \boldsymbol{\epsilon}, \tag{3.10}$$

where $f(\mathbf{X}) = (f_1(X_1), \ldots, f_p(X_p)^\top$ and $g(\mathbf{Y}) = (g_1(Y_1), \ldots, g_r(Y_r)^\top$ are unknown functions. Examples of model (3.10) include the additive regression model, single index model, copula regression model, and semiparametric proportional hazards models, among others. Typically, when we apply a transformation to data, the transformation functions are continuous and one-to-one, for example, log transformation, Box-cox transformation and Fisher's z transformation. The functions that satisfy these two assumptions must be strictly monotone.

Under model (3.10), the linearity is assumed between $g(\mathbf{Y})$ and $f(X)$, and if the transformation functions f and g are known, we can use envelope models to efficiently estimate $\boldsymbol{\beta}$. Unlike in Chapter 2, we consider the dimension reduction in $f(\mathbf{X})$. This problem is often addressed under the name of penalized least squares, where a penalty term of $f(\mathbf{X})$ is added to the estimation equation to achieve sparsity in $\boldsymbol{\beta}$. Here we take a different perspective: instead of selecting the predictors that influence the response the most, we seek linear combinations of the predictors that carries all the essential useful information

in explaining the responses. Specifically, we allow the predictors to be transformed into $f(\mathbf{X})$ and this greatly broaden the scope of the scaled predictor model.

3.4.2 Model Formulation

Let us further formulate (3.10) in more details. Suppose we have an independent and identically distributed random sample $\mathbf{Z}_1 = (\mathbf{Y}_1, \mathbf{X}_1), \ldots, \mathbf{Z}_n = (\mathbf{Y}_n, \mathbf{X}_n) \in \mathbb{R}^{p+r}$. The observations $(\mathbf{Y}_i, \mathbf{X}_i)$ satisfies the Gaussian copula regression model, if there exists a set of strictly increasing functions $g = \{g_1, \ldots, g_r\}$ and $f = \{f_1, \ldots, f_p\}$ such that the marginally transformed random vectors $\widetilde{\mathbf{Z}}_i = (\widetilde{\mathbf{Y}}_i, \widetilde{\mathbf{X}}_i) = (g_1(Y_1), \ldots, g_r(Y_r), f_1(X_1), \ldots, f_p(X_p))$ satisfy $\widetilde{\mathbf{Z}}_i \sim N_{p+r}(0, \boldsymbol{\Sigma}_Z)$ for some positive-definite covariance matrix $\boldsymbol{\Sigma}_Z$. Under a Gaussian copula regression model, one has the following linear relationship for the transformed data:

$$\widetilde{\mathbf{Y}}_i = \boldsymbol{\beta}\widetilde{\mathbf{X}}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{3.11}$$

where $\boldsymbol{\beta} \in \mathbb{R}^{r \times p}$ and ϵ_i are i.i.d. zero-mean Gaussian variables. The fundamental difference between the Gaussian copula regression model (3.11) and conventional linear regression model is that one observes $\{(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n)\}$, instead of $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$ as the transformations f and g are unknown.

For model (3.11), the linear relationship is assumed between the unobserved $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$. Similarly, we assume an envelope structure exists for the unobserved $\{(\widetilde{\mathbf{Y}}_1, \widetilde{\mathbf{X}}_1), \ldots, (\widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{X}}_n)\}$, and we can use the observed $\{(\mathbf{Y}_1, \mathbf{X}_1), \ldots, (\mathbf{Y}_n, \mathbf{X}_n)\}$ to estimate the envelope structure, which means the

57

proposed Gaussian copula response envelope is invariant towards any strictly increasing transformation of the responses and predictors. We assume the following predictor envelope exist between the unobserved $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{X}}$.

$$\widetilde{\mathbf{Y}} = \boldsymbol{\eta}^T \boldsymbol{\Gamma}^T \Lambda^{-1} \widetilde{\mathbf{X}} + \boldsymbol{\epsilon}, \quad \text{with} \quad \boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}} = \Lambda \boldsymbol{\Gamma} \boldsymbol{\Omega} \boldsymbol{\Gamma}^T \Lambda + \Lambda \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^T \Lambda, \qquad (3.12)$$

The scaled response $\Lambda^{-1}\widetilde{\mathbf{X}}$ follows an envelope model with u-dimensional envelope $\varepsilon_{\Lambda^{-1}\boldsymbol{\Sigma}\Lambda^{-1}}(\Lambda^{-1}\mathcal{B})$ and semi-orthogonal basis matrix $\boldsymbol{\Gamma}$. The material and immaterial pairs of $\mathbf{Y}$ are $\Lambda \mathbf{P}_{\boldsymbol{\Gamma}} \Lambda^{-1}$ and $\Lambda \mathbf{Q}_{\boldsymbol{\Gamma}} \Lambda^{-1}$, respectively.

As one can notice, the structure of the proposed Gaussian copula predictor envelope (3.12) is very similar to the ones in scaled predictor envelope (3.8), with slight differences where one only observes the transformed $\{\mathbf{Y}, \mathbf{X}\}$, instead of $\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$. The key obstacle now is to estimate the parameters in (3.12) without observing $\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$. If we take a closer look at the likelihood function in the objection function in scaled predictor envelope (3.9), the essential information we need is the covariance matrix $\widetilde{\mathbf{X}}$ and the residuals, so the question now becomes how do we estimate $S_{\widetilde{\mathbf{X}}}$ and $S_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}}}$ using only the observed $\mathbf{X}$ and $\mathbf{Y}$. Similarly, a key to this question is the use of rank-based correlation estimator. To keep the autonomy of this chapter, we introduce the rank-based estimator of correlation again, which is essentially similar to Section 2.4.3.

3.4.3 Rank-based Estimator of Correlation Matrix

The goal here is to estimate the $\boldsymbol{\Sigma}_Z$, the covariance matrix of $\widetilde{\mathbf{Z}}$, using only the observed $(\mathbf{Y}, \mathbf{X})$. Since the marginal transformations f and g are unknown, we use rank-based correlation of the observed data $\mathbf{Z}$ to estimate the covariance/correlation matrix $\boldsymbol{\Sigma}_Z$. Given that f and g are strictly monotone, $\mathbf{Z}$ and

$\widetilde{\mathbf{Z}}$ have the same elliptical copula. Set $d = p + r$. If $\widetilde{\mathbf{Z}}_i \sim N_d(0, \boldsymbol{\Sigma}_Z)$ with $\boldsymbol{\Sigma}_Z = (\sigma_{jk})_{1 \le j,k \le d}$, then

$$\sigma_{jk} = \sin(\frac{\pi}{2}\tau_{jk}),$$

where τ_{jk} is Kendall's tau and is defined as

$$\tau_{jk} = E[\mathrm{sgn}(\tilde{Z}_{1j} - \tilde{Z}_{2j})\mathrm{sgn}(\tilde{Z}_{1k} - \tilde{Z}_{2k})]$$

with $\widetilde{\mathbf{Z}}_i = (\tilde{Z}_{i1}, \dots, \tilde{Z}_{id})^T, i = 1, 2$ being two independent copies of $N_d(0, \boldsymbol{\Sigma}_Z)$.

The Kendall's tau τ_{jk} is invariant under strictly increasing marginal transformations. This promises that using the observed data $\mathbf{Z}$ will return the same correlation matrix as using the observed $\widetilde{\mathbf{Z}}$. Specifically, we have

$$\begin{aligned}
\widehat{\tau}_{jk} &= \frac{2}{n(n-1)} \sum_{1 \le i_1 \le i_2 \le n} \mathrm{sgn}(\tilde{Z}_{i_1,j} - \tilde{Z}_{i_2,j})\mathrm{sgn}(\tilde{Z}_{i_1,k} - \tilde{Z}_{i_2,k}) \\
&= \frac{2}{n(n-1)} \sum_{1 \le i_1 \le i_2 \le n} \mathrm{sgn}(Z_{i_1,j} - Z_{i_2,j})\mathrm{sgn}(Z_{i_1,k} - Z_{i_2,k}), \quad 1 \le j, k \le d.
\end{aligned}$$
(3.13)

Using the Kendall's tau correlation estimator, we obtain the following estimator for the correlation matrix $\boldsymbol{\Sigma}_Z$,

$$\boldsymbol{\Sigma}_Z = (\widehat{\sigma}_{jk}) \text{ with } \widehat{\sigma}_{jk} = \sin(\frac{\pi}{2}\widehat{\tau}_{jk}). \tag{3.14}$$

We divide $\boldsymbol{\Sigma}_Z$ into four sub-matrices, denoted by $\Sigma_{XX}, \Sigma_{XY}, \Sigma_{YX}, \Sigma_{YY}$, and their corresponding Kendall's tau based estimators are $\widehat{\Sigma}_{XX}, \widehat{\Sigma}_{XY}, \widehat{\Sigma}_{YX}, \widehat{\Sigma}_{YY}$. These covariance estimators will be essential in the estimation of $\boldsymbol{\beta}$. However, keep in mind that here we assume $\mathrm{diag}(\boldsymbol{\Sigma}_Z) = \mathbf{1}$, so the correlation matrix is the same as covariance matrix. Since the original response envelope model is not invariant to the rescaling of the variables, we use the scaled envelope response model.

Note that Kendall's tau based correlation matrix may not be positive semidefinite and we do need to invert some of these matrices. Take $\widehat{\Sigma}_{XX}$ for example, we can project $\widehat{\Sigma}_{XX}$ onto the cone of the positive semidefinite matrices. This can be done using the following convex optimization problem

$$\widehat{\Sigma}_{XX}^{+} = \arg\min_{\Sigma \geq 0} ||\widehat{\Sigma}_{XX} - \Sigma||_{2,s}. \tag{3.15}$$

3.4.4 Estimation of $\boldsymbol{\beta}$

In order to estimate the parameters in (3.12), we use the maximum likelihood estimators. The derivation of the maximum likelihood estimators can be obtained similar to the ones in the scaled predictor model. Under (3.12), the scaled predictor $\Lambda^{-1}\widetilde{\mathbf{X}}$ follows an envelope model with u-dimensional envelope $\varepsilon_{\Lambda^{-1}\boldsymbol{\Sigma}\Lambda^{-1}}(\Lambda^{-1}\mathcal{B})$ and semi-orthogonal basis matrix $\boldsymbol{\Gamma}$. The material and immaterial pairs of $\mathbf{X}$ are $\Lambda\mathbf{P}_{\boldsymbol{\Gamma}}\Lambda^{-1}$ and $\Lambda\mathbf{Q}_{\boldsymbol{\Gamma}}\Lambda^{-1}$, respectively. The maximum likelihood estimators $\widehat{\Lambda}$ and $\widehat{\Gamma}$ of Λ and Γ, respectively, can be obtained by minimizing the objective function

$$L_u(\Lambda, \boldsymbol{\Gamma}) = \log|\boldsymbol{\Gamma}^T\Lambda^{-1}(\boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}} - \boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}}\Sigma_{\widetilde{\mathbf{Y}}}^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{Y}}\widetilde{\mathbf{X}}})\Lambda^{-1}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^T\Lambda\boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}}^{-1}\Lambda\boldsymbol{\Gamma}|. \tag{3.16}$$

The covariance related quantities such as $\boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}}$ and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}}}$ can be estimated using the Kendall's τ estimator as discussed in Section 3.4.3. Specifically, we use $\widehat{\Sigma}_{XX}$ and $\widehat{\Sigma}_{XX} - \widehat{\Sigma}_{XY}\widehat{\Sigma}_{XX}^{-1}\widehat{\Sigma}_{YX}$ as estimators for $\Sigma_{\widetilde{\mathbf{X}}}$ and $\Sigma_{\widetilde{\mathbf{X}}|\widetilde{\mathbf{Y}}}$ respectively. Substituting them into the objective function, we have

$$\Lambda = \arg\min_{\Lambda, \boldsymbol{\Gamma}} \log|\boldsymbol{\Gamma}^T\Lambda^{-1}(\widehat{\Sigma}_{XX} - \widehat{\Sigma}_{XY}\widehat{\Sigma}_{YY}^{-1}\Sigma_{YX})\Lambda^{-1}\boldsymbol{\Gamma}| + \log|\boldsymbol{\Gamma}^T\Lambda\widehat{\Sigma}_{XX}^{-1}\Lambda\boldsymbol{\Gamma}|. \tag{3.17}$$

The optimization problem in (3.17) can be performed in an alternation fashion. We iterate between Λ and $\boldsymbol{\Gamma}$ until the difference between the objective functions

60

in two adjacent iterations is smaller than a pre-specified value. Once we have $\widehat{\Lambda}$ and $\widehat{\Gamma}$, the maximum likelihood estimators for the rest of the parameters can be obtained as follows:

$$\widehat{\boldsymbol{\eta}} = (\widehat{\Gamma}^T\widehat{\Lambda}^{-1}\widehat{\Sigma}_{XX}\widehat{\Lambda}^{-1}\widehat{\Gamma})^{-1}\widehat{\Gamma}^T\widehat{\Lambda}^{-1}\widehat{\Sigma}_{XY};$$

$$\widehat{\boldsymbol{\beta}} = \widehat{\Lambda}^{-1}\widehat{\Gamma}\widehat{\boldsymbol{\eta}},$$

$$\widehat{\boldsymbol{\Omega}} = \widehat{\Gamma}^T\widehat{\Lambda}^{-1}\widehat{\Sigma}_{XX}\widehat{\Lambda}^{-1}\widehat{\Gamma}, \tag{3.18}$$

$$\widehat{\boldsymbol{\Gamma}}_0^T = \widehat{\Gamma}_0^T\widehat{\Lambda}^{-1}\widehat{\Sigma}_{XX}\widehat{\Lambda}^{-1}\widehat{\Gamma}_0,$$

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\Lambda}\widehat{\Gamma}\widehat{\Omega}\widehat{\Gamma}^T\widehat{\Lambda}^T + \widehat{\Lambda}\widehat{\Gamma}_0\widehat{\Omega}_0\widehat{\Gamma}_0^T\widehat{\Lambda}^T.$$

3.4.5 Selection of u

The dimension of the envelope u is an important parameter in determining the right envelope structure. Likelihood-based methods such as AIC, BIC or other information criteria can be used to select the dimension u. BIC is often preferred for parameter estimation and cross validation is preferred for prediction.

The number of free parameters in model (3.11) can be counted as

$$N_{q,s} = s - 1 + r + p + rq + p(p+1)/2 + r(r+1)/2.$$

The AIC estimator of u is $\arg\min -2\widehat{L}(u) + 2N_{q,s}$. The maximized log likelihood under the copula predictor envelope model with dimension u take the following form

$$\widehat{L}(u) = -\frac{n(p+r)}{2}\log(2\pi) - \frac{nr}{2} - \frac{n}{2}\log|\widehat{\boldsymbol{\Sigma}}_{\mathbf{x}}| - \frac{n}{2}\mathrm{tr}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}S_{XX})$$
$$- \frac{n}{2}\log|\widehat{\Sigma}_{XX} - \widehat{\Sigma}_{XY}\widehat{\Sigma}_{XX}^{-1}\widehat{\Sigma}_{YX}|. \tag{3.19}$$

The BIC estimator of u is $\arg\min -2\widehat{L}(u) + \log(n)N(u)$. The properties of BIC under Gaussian copula model is similar to the ones in predictor envelope

models. If the true model is in the candidate set then, as $n \to \infty$, BIC will select the true model with probability tending to 1, AIC will select a model that at least contains the true model and LRT will select the true model with probability $1 - \alpha$, where α is a significance level.

3.5 Theoretical Results

Here we establish the consistency and asymptotic normality of the proposed estimator.

We introduce the operator vec : $\mathbb{R}^{a \times b} \to \mathbb{R}^{ab}$ stacks the columns of a matrix, and the operator vech : $\mathbb{R}^{a \times a} \to \mathbb{R}^{a(a+1)/2}$ stacks the lower triangular part of a symmetric matrix.

First we establish the consistency and asymptotic normality of the proposed copula predictor envelope (cpe) estimator. Although the CPE estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma_X}$ are derived using the normal likelihood, they are $\sqrt{n}$ consistent without the normality assumption.

Theorem 3.5.1 *Assume model (3.12) holds and that* $(\mathbf{Y}, \mathbf{X})$ *has finite fourth moments. Then*

$$\sqrt{n}[\{vec^T(\widehat{\boldsymbol{\beta}}_{cpe}), vech^T(\widehat{\Sigma}_{\mathbf{X},cpe})\}^T - \{vec^T(\boldsymbol{\beta}), vech^T(\Sigma_{\mathbf{X}})\}^T]$$

is asymptotically normally distributed, and $\widehat{\boldsymbol{\beta}}_{cpe}$ *and* $\widehat{\Sigma}_{\mathbf{X},cpe}$ *are* $\sqrt{n}$ *consistent estimators of* $\boldsymbol{\beta}$ *and* $\Sigma_{\mathbf{X}}$, *respectively.*

Without the normality assumption, it is difficult to give a useful expression for the asymptotic variance of the proposed estimator. However, if we are wiling to assume the normality, we can obtain the asymptotic variance of the proposed copula predictor envelope estimators.

If a quantity is derived from the ordinary predictor envelope, it is designated with a subscript o. For instance $\widetilde{\mathbf{Y}}$ and $\Lambda^{-1}\widetilde{\mathbf{X}}$ follow an ordinary envelope model and thus we write $\boldsymbol{\beta}_o = \Lambda\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}_o = \Lambda^{-1}\boldsymbol{\Sigma}_{\widetilde{\mathbf{X}}}\Lambda^{-1}$. The gradient matrix under model (3.6) is then

$$H_o = \partial\{\mathrm{vec}^T(\boldsymbol{\beta}_o), \mathrm{vech}^T(\boldsymbol{\Sigma}_o)\}^T / \partial\{\mathrm{vec}^T(\boldsymbol{\eta}), \mathrm{vec}^T(\boldsymbol{\Gamma}), \mathrm{vech}^T(\boldsymbol{\Omega}), \mathrm{vech}^T(\boldsymbol{\Omega}_0)\}.$$

Let bdiag(.) denote a block diagonal matrix with diagonal blocks as arguments. The column vector $\lambda = (\lambda_1, \ldots, \lambda_{q-1})^T$ contains the $q-1$ unique elements of Λ, so that $\lambda^T = \mathrm{vec}^T(\Lambda)L$, where $L = (e_{r1+1} \otimes e_{r1+1}, \ldots, e_{p-r_q+1} \otimes e_{p=r_q+1}) \in \mathbb{R}^{p^2 \times (q-1)}$ extracts the $q-1$ scaling parameters from $\mathrm{vec}(\lambda)$, $\otimes$ denotes Kronecker product, $e_i \in \mathbb{R}^{p \times 1}$ contains a 1 in the i-th position and 0 elsewhere.

The Fisher information for $\{\mathrm{vec}^T(\boldsymbol{\beta}_o), \mathrm{vech}(\boldsymbol{\Sigma}_o)\}$ is

$$J_o = \mathrm{bdiag}\{\boldsymbol{\Sigma}_{\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}}}^{-1} \otimes \boldsymbol{\Sigma}_o, E_p^T(\boldsymbol{\Sigma}_o^{-1} \otimes \boldsymbol{\Sigma}_o^{-1})E_p/2\} \in \mathbb{R}^{\{rp+p(p+1)/2\} \times \{rp+p(p+1)/2\}}.$$

Let $\mathbf{K} = \mathrm{bdiag}\{-\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T \otimes \mathbf{I}_p, 2C_p(\boldsymbol{\Sigma}_o^{-1} \otimes \mathbf{I}_p)\}(L^T, L^T)^T$, $G = Q_{H_o(J_o)}\mathbf{K}$ and $D = \mathrm{bdiag}\{\mathbf{I}_r \otimes \Lambda^{-1}, C_p(\Lambda \otimes \Lambda)E_p\}$. Then the Fisher information for $\{\mathrm{vec}^T(\boldsymbol{\beta}), \mathrm{vech}^T(\boldsymbol{\Sigma})\}^T$ is $D^{-1}J_0D^{-T}$.

Theorem 3.5.2 *Under the normal Gaussian copula predictor envelope model (3.12),*

$$\sqrt{n}[\{vec^T(\widehat{\boldsymbol{\beta}}_{cpe}), vech^T(\widehat{\Sigma}_{\mathbf{X},cpe})\}^T - \{vec^T(\boldsymbol{\beta}), vech^T(\Sigma_{\mathbf{X}})\}^T]$$

converges in distribution to a normal random vector with mean zero and covariance matrix

$$V = DG(G^TJ_oG)^\dagger G^TD^T + DH_o(H_o^TJ_oH_o)^\dagger H_o^TD^T = V_1 + V_2.$$

The estimators $\widehat{\boldsymbol{\beta}}_{cpe}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X},cpe}$ are more efficient or are at least as efficient as the OLS estimators asymptotically; that is, $DJ_o^{-1}D^T - V$ is a positive semi-definite matrix.

The asymptotic covariance matrix V is decomposed into two parts: V_2 is the asymptotic variance when the scaling Λ is known, and V_1 can be thought as the asymptotic cost of estimating Λ. The above theorem also states that the proposed copula predictor envelope estimator is asymptotically at least as efficient as the OLS estimators. Specifically, let us discuss when the proposed estimator is more efficient.

Theorem 3.5.3 *Under the normal CPE model (3.12), when $u \leq p-(q-1)/r$, the estimators $\widehat{\boldsymbol{\beta}}_{cpe}$ and $\widehat{\boldsymbol{\Sigma}}_{\mathbf{X},cpe}$ have the same asymptotic covariance as the OLS estimators.*

Let u_0 be the ceiling of $p - (q - 1)/r$. The above theorem states that when $u \leq u_0$, the CPE estimators and the OLS estimators have the same asymptotic variances. This means that there is no point to rescale all of the predictors in the univariate linear regression and the the proposed model works better when considering multiple responses.

3.6 Simulations and Data Analysis

3.6.1 Simulations

Following the settings as in Cook and Su (2016)[13], we set $r = 8, u = 5$ and $p = 10$. The elements in $\boldsymbol{\eta}$ were generated from the $U(0, 2)$ distribution, and $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ was obtained by normalizing a $p \times p$ matrix whose elements were

generated as independent and from the $U(0,1)$ distribution. For the covariance matrices, we take $\mathbf{\Omega} = \sigma^2 I_5$ and $\mathbf{\Omega}_0 = \sigma_0^2 I_5$ with $\sigma = 5$ and $\sigma_0 = \sqrt{5}$ or 5. The scaling parameter Λ takes diagonal elements $2^0, 2^0.5, 2^1, \ldots, 2^4.5$. We simulated the error term $\boldsymbol{\epsilon}$ from the multivariate normal distribution with mean 0 and covariance matrix $\mathbf{\Sigma}_{\mathbf{Y}|\mathbf{X}} = ADA^T$, where A is an orthogonal matrix obtained by normalizing an $r \times r$ matrix of $U(0,1)$ random variables, and D is a diagonal matrix with diagonal elements $1, 2, \ldots, r$. The sample sizes were 100, 200, 300, 500, 800, 1200, and 200 replicates were used for each sample size. We consider two settings. For the first setting, we set $Y_{ij} = exp(\tilde{Y}_{ij})$ and $X_{ij} = 3 * \tilde{X}_{ij} - 10$ for $i = 1, \ldots, 5$, $Y_{ij} = 2 * \tilde{Y}_{ij} + 5$ and $X_{ij} = exp(\tilde{X}_{ij})$ for $i = 6, \ldots, 10$; and for the second setting, we set $Y_{ij} = \Phi(\tilde{Y}_{ij})$ and $X_{ij} = \tilde{X}_{ij}^2 + 5$ for $i = 1, \ldots, 5$, $Y_{ij} = \tilde{Y}_{ij}^3 - 10$ and $X_{ij} = \Phi(\tilde{X}_{ij})$ for $i = 6, \ldots, 10$.

We compared the performance of the original predictor envelope [9] the scaled response envelope [13] and the proposed Gaussian copula response envelope using either original data $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ or the observed data $(\mathbf{Y}, \mathbf{X})$. To evaluate the performances of these models, we considered two important metrics: the angle between the true envelope space and the estimated envelope space using different envelope models and the standard deviation of $\widehat{\boldsymbol{\beta}}$. The angle $\angle\{\text{span}(\mathbf{A}_1), \text{span}(\mathbf{A}_2)\}$ between the subspace spanned by columns of the semi-orthogonal basis matrices $\mathbf{A}_1 \in \mathbb{R}^{r \times u}$ and $\mathbf{A}_2 \in \mathbb{R}^{r \times u}$ was computed in degrees as the arc cosine of the smallest absolute singular value of $\mathbf{A}_1^T \mathbf{A}_2$.

Tables (3.1) to (3.4) report the comparison of the angles between the estimated and true envelope space for different simulation settings. Again we point out that the range of the angle is between 0 and 90 degrees. Usually if the angle is larger than 40 degrees (depending on the size of the space), it means

Table 3.1: Predictor Envelope Models Setting 1 with $\sigma_0^2 = 5$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 100$	84.56	10.35	10.23
	$n = 200$	89.24	8.45	8.56
	$n = 300$	84.56	8.56	8.45
	$n = 500$	85.45	8.32	8.24
	$n = 800$	82.35	7.84	7.38
	$n = 1200$	89.34	7.34	7.37
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 100$	85.67	75.24	10.26
	$n = 200$	88.34	78.56	8.45
	$n = 300$	85.65	74.34	8.89
	$n = 500$	86.34	78.23	8.56
	$n = 800$	83.24	77.21	7.29
	$n = 1200$	89.12	74.56	7.40

that the envelope models failed to find the right structure. As expected, we have the following findings: 1) the original predictor envelope did not pick up the envelope structure on all occasions; 2) the scaled predictor envelope model is able to pick up the envelope structure using $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$, which is not observed in practice, but not the observed $(\mathbf{Y}, \mathbf{X})$; 3) the proposed copula predictor envelope is able to pick up the correct envelope structure using either $(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}})$ or $(\mathbf{Y}, \mathbf{X})$, and the performances are very similar; 4) with the increase of sample size, the performance of the proposed copula predictor envelope estimators improves; 5) with the increase of the contract between material and immaterial parts (increase of σ_0), the performance of the proposed copula predictor envelope estimators improves.

Table 3.2: Predictor Envelope Models Setting 1 with $\sigma_0^2 = 25$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 100$	88.45	8.45	9.01
	$n = 200$	84.56	7.45	7.54
	$n = 300$	82.34	6.78	6.79
	$n = 500$	83.45	6.03	6.05
	$n = 800$	83.45	5.84	5.88
	$n = 1200$	81.23	4.85	4.89
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 100$	85.67	74.56	9.05
	$n = 200$	84.32	73.24	7.50
	$n = 300$	87.45	74.56	6.89
	$n = 500$	89.34	74.56	6.06
	$n = 800$	84.56	72.91	5.88
	$n = 1200$	83.45	68.45	4.89

Table 3.3: Predictor Envelope Models Setting 2 with $\sigma_0^2 = 5$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 100$	85.67	10.38	10.24
	$n = 200$	88.34	8.44	8.53
	$n = 300$	85.43	8.53	8.46
	$n = 500$	84.25	8.31	8.22
	$n = 800$	82.24	7.89	7.41
	$n = 1200$	84.45	7.40	7.37
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 100$	84.34	74.53	10.26
	$n = 200$	84.56	79.54	8.56
	$n = 300$	86.65	73.45	8.45
	$n = 500$	81.23	78.32	8.25
	$n = 800$	84.24	71.23	7.40
	$n = 1200$	86.24	72.34	7.39

Table 3.4: Predictor Envelope Models Setting 2 with $\sigma_0^2 = 25$: Comparisons of the angle between the true envelope subspace and the estimated subspace using the correct envelope dimension $u = 5$. The results are based on 200 replications.

Angle	n	Original Envelope	Scaled Envelope	Copula Envelope
	$n = 100$	84.56	8.46	9.02
	$n = 200$	81.56	7.51	7.56
$\{\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}\}$	$n = 300$	82.34	6.81	6.81
	$n = 500$	83.34	6.02	6.07
	$n = 800$	85.45	5.89	5.87
	$n = 1200$	86.78	4.84	4.88
	$n = 100$	84.56	73.45	9.04
	$n = 200$	83.45	71.12	7.61
$\{\mathbf{Y}, \mathbf{X}\}$	$n = 300$	82.34	74.34	6.83
	$n = 500$	84.56	76.56	6.04
	$n = 800$	89.24	78.34	5.67
	$n = 1200$	81.24	72.31	4.85

3.6.2 Data Analysis

To evaluate the usefulness of the proposed copula predictor envelope model, we use the Alzheimer's Disease Neuroimaging Initiative (ADNI) [2] [1] data. In Chapter 2, we used the ADHD-200 data to predict brain connectivity using some demographic and clinical variables. In this data analysis, we take a different perspective. The goal is to use demographic, clinical variables and brain connectivities to predict the Alzheimer's Disease assessment scores of the participants. This is an excellent data to test the proposed method since the response variables are multivariate (different aspects of the psychological

---

[1]Data used in preparation of this thesis were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

testing scores), and there are a large number of predictors with different scales collected. Also, preliminary analysis show that the linearity does not hold well between some predictors and responses. First, let us explain a bit about the background of the data set.

ADNI is a global longitudinal study for AD through the enrollment and follow-up of cohorts of individuals who have mild cognitive impairment (MCI) and mild Alzheimer's disease. The study is designed for the detection at the earliest possible stage and tracking the progression of Alzheimer's disease with biomarkers to assess the brain structure and the brain function. The participants enrolled by ADNI were between 55 to 90 years of age, selected based on the particular criteria, and recruited at the 57 ADNI acquisition sites located in the United States and Canada. The five cohorts in this study are Normal Control (CN), Significant Memory Concern (SMC), Early Mild Cognitive Impairment (EMCI), Late Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD), respectively. The selected subjects undergo clinical, imaging, genetic, and also biochemical biomarkers at multiple time points [1, 2].

The general goals of the ADNI study are for validation of biomarkers' data for the trials' use in the illness clinical treatment and assessments, for exploring methods for obtaining data and analyzing neuroimaging data in longitudinal studies for clinical trials on patients with normal controls, mild cognitive impairment, and Alzheimer's disease, for making data repository accessible for other researchers and communities, and for developing technical standards of imaging in longitudinal studies [2, 26].

In this study, we employed a subset of the resting-state fMRI ADNI data

which includes 57 subjects from two cohorts, one is 33 subjects from Normal Control group, and the other one is 24 subjects from Alzheimer's Disease group. The resting-state fMRI data were preprocessed using Automated Anatomical Labeling (AAL) template [43]. AAL is an anatomical atlas of total regions of interest obtained on one subject. The AAL template is broadly employed in functional neuroimaging research, including resting-state fMRI. It aims at deriving neuroanatomical labels in a space where the measurements of brain function were captured [43]. The non-overlapping regions of interest were then extracted for each subject. For each subject, each time-series and ROI were computed through averaging all the voxels' time series within the ROIs [37]. Hence each subject has BOLD signal data at 116 ROIs through 134 equal spaced time courses. Also the demographic and clinical information of 57 subjects were collected, which consist of ID, gender, age and diagnostic information. For the variable gender, 0 represents female and 1 male. For diagnosis (DX), 0 stands for Normal Control(CN), and 4 Alzheimer's Disease (AD). All subjects had 1.5 Tesla and 3 Tesla scans by Philips scanners, having their eyes open when receiving the scanning [35]. For the responses, we consider the Alzheimer's Disease Assessment Scale (ADAS) sub-scores and here $p = 11$. For most studies, a total score by combining the 11 different scores is considered usually. In this data analysis, we aim to look at the effect of the predictors on each of the sub scores. This is of great importance since although these sub-scores are correlated but each one represent a unique part of the assessment. The predictors considered here include gender, age, and "top" 10 brain connectivities. For computation of the brain connectivities, more details can be found in Section 2.6.2. The "top" 10 candidates were

found by ordering the p-values comparing the AD group and CN group, which means only the top 10 that differentiate AD the most were considered.

All variables in $\mathbf{X}$ and $\mathbf{Y}$ were standardized to have sample mean 0 and sample variance 1. We considered 4 different methods: the ordinary least square, the original predictor envelope, the scaled predictor envelope and the proposed copula predictor envelope. We examined the average of the prediction errors from 50 five-fold cross validations with random splits as the evaluation metric. At $u = 1$, the ordinary least square estimator has prediction error as large as 5.68. The original and scaled predictor envelope estimators reports prediction error as large as 4.58 and 3.98, respectively. The proposed copula predictor envelope model has prediction error of 2.38. This is about 40.49% of reduction of prediction error compared to the ordinary least squares, and another 40.20% reduction compared to scaled predictor envelope estimator. Using cross validation, the selected u for the proposed copula envelope is 5 with the prediction error as 1.84.

3.7   Discussion

Dimension reduction in the predictor space is one of the most important tasks in statistics. This problem has been addressed by many traditional methods including reduced rank regression, principal component regression, ridge regression and other type of penalized regression. Most of these aforementioned methods are invariant or equivariant to a scale transformation of the predictors. The sparse predictor envelope proposed by Cook and Su (2016)[13] is the first scale-invariant method.

In this chapter, we have developed the copula predictor envelope model, which further generalizes the scaled predictor envelope model. The main idea behind this work is the use of rank-based covariance estimators, which are invariant to any monotone transformations. The block structure for Λ introduced in [13] is kept in this work, although one would argue in practice that it is difficult to assume multiple predictors to have similar variances. However, as the theoretic properties have shown, the proposed copula predictor envelope model is still useful when the number of responses is larger than one. It is worthy pointing out that the predictor envelope based methods depends on the collinearity rather than the mitigation through regularization.

For future research, the proposed predictor envelope model can be further applied to other special models, such as the latent factor model, where one can treat the latent effect as a linear transformation of the observed covariates. Also, censored predictors can also be considered to further broaden the applicability of the work. In the context of censored data, for the unknown transformation we can consider here is a piecewise linear function. The difficulties lie in there includes the discontinuity of the transformation function at points where the cutoff of the data happens. The work in this chapter is confined to the cases where $n > p$. A scaled version of the work which allows the cases when $n < p$ is highly important.

Chapter 4

Envelope-based High-dimensional Multivariate Test for Mean Vector

4.1 Introduction

Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be independent and identically distributed (i.i.d.) copies of the random vector $Y \in \mathbb{R}^r$ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma} > 0$. In this chapter, we consider the testing problem for the following hypothesis:

$$H_0 : \boldsymbol{\mu} = \mathbf{0}, \quad H_A : \boldsymbol{\mu} \neq \mathbf{0}.$$

The Hotelling T^2 test has been widely used in multivariate analysis since its proposal due to its many nice properties: it is *uniformly the most powerful* of the affine invariant tests. However, its performance deteriorates quickly or even not well-defined when the number of features r is comparable or even larger than the sample size n. There has been many work proposed in the literature to address this issue ([44] [23] [6] [27]). In this chapter, we propose a novel envelope-based high-dimensional multivariate test.

Envelope [11], as a descendant of sufficient dimension reduction, has wit-

nessed its success in regression analysis since its proposal due to its ability to achieve massive efficiency gains in parameter estimation. More details of envelopes can be found in Cook's monograph [7]. Most of current research in envelopes focus on regression models; however, envelopes can also be used in the estimation of a multivariate mean. For example, Su and Cook (2013)[40] proposed envelope models that accommodate heteroscedastic error structure in the framework of estimating multivariate means for different populations.

To describe the use of envelopes for the estimation of a multivariate mean $\boldsymbol{\mu}$, let $\mathcal{S} \in \mathbb{R}^p$ denote the smallest subspace with the properties

$$\text{(a) } \boldsymbol{\mu} \in \mathcal{S} \text{ and (b) } \mathbf{P}_{\mathcal{S}}\mathbf{Y} \perp\!\!\!\perp \mathbf{Q}_{\mathcal{S}}\mathbf{Y}. \tag{4.1}$$

Conditions (a) and (b) indicate that marginal information on $\boldsymbol{\mu}$ is available from $\mathbf{P}_{\mathcal{S}}\mathbf{Y}$, which $\mathbf{Q}_{\mathcal{S}}\mathbf{Y}$ supplies no marginal information about $\boldsymbol{\mu}$, which is equivalent to say that $\mathcal{S}$ is a reducing subspace of $\boldsymbol{\Sigma}$. In short, we are led to the $\boldsymbol{\Sigma}$-envelope of $\mathcal{M} := \text{span}(\boldsymbol{\mu}), \varepsilon_{\boldsymbol{\Sigma}}(\mathcal{M})$. Then $\mathbf{P}_{\varepsilon}\mathbf{Y}$ contains all of the material information on $\boldsymbol{\mu}$ with *material* variation $\mathbf{P}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{P}_{\varepsilon}$, and $\mathbf{Q}_{\varepsilon}\mathbf{Y}$ contains all of the *immaterial* information on $\boldsymbol{\mu}$ with material variation $\mathbf{Q}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{Q}_{\varepsilon}$.

To gain intuition about the potential gain using the envelope-based mean, we suppose the envelope $\varepsilon_{\boldsymbol{\Sigma}}(\mathcal{M})$ is known, and then the maximum likelihood estimator of $\boldsymbol{\mu}$ is just $\hat{\boldsymbol{\mu}} = \mathbf{P}_{\varepsilon}\bar{\mathbf{Y}}$, which has variance $n^{-1}\mathbf{P}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{P}_{\varepsilon}$. Since $\boldsymbol{\Sigma} = \mathbf{P}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{P}_{\varepsilon} + \mathbf{Q}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{Q}_{\varepsilon}$, we have

$$\text{var}(\bar{\mathbf{Y}}) - \text{var}(\hat{\boldsymbol{\mu}}) = n^{-1}\mathbf{Q}_{\varepsilon}\boldsymbol{\Sigma}\mathbf{Q}_{\varepsilon}, \tag{4.2}$$

so the potential gain depends on the sample size and the immaterial variation. When p is large, it is very likely for $\mathbf{Y}$ to contain immaterial information, which means $\boldsymbol{\mu}$ is very likely to fall in an eigenspace of $\boldsymbol{\Sigma}$.

Here we are only consider the scenario when $n > r$, and when $n \leq r$, sparse envelope models can be considered by assuming some of the elements in $\boldsymbol{\mu}$ or $\boldsymbol{\Sigma}$ to be 0.

4.2   Envelope-based Hotelling T^2 Test

Let $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times u}$ be a semi-orthogonal basis matrix for the envelope $\boldsymbol{\varepsilon_\Sigma}(\mathcal{M})$, where $u = \dim(\boldsymbol{\varepsilon_\Sigma}(\mathcal{M}))$, and let $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma_0})$ be an orthogonal matrix so that $\boldsymbol{\Gamma_0}$ is a basis matrix for the orthogonal complement of $\boldsymbol{\varepsilon_\Sigma}(\mathcal{M})$. Since $\boldsymbol{\mu} \in \boldsymbol{\varepsilon_\Sigma}(\mathcal{M})$ by construction, we can write $\boldsymbol{\mu} = \boldsymbol{\Gamma\eta}$ for coordinator vector $\boldsymbol{\eta} \in \mathbb{R}^{u \times 1}$. Let $\boldsymbol{\Omega} = \boldsymbol{\Gamma^\mathsf{T} \Sigma \Gamma} > 0$, and $\boldsymbol{\Omega_0} = \boldsymbol{\Gamma_0^\mathsf{T} \Sigma_0 \Gamma_0} > 0$, then the envelope model for the multivariate mean can be summarized as

$$\mathbf{Y} \sim N(\boldsymbol{\Gamma\eta}, \boldsymbol{\Gamma\Omega\Gamma^\mathsf{T}} + \boldsymbol{\Gamma_0 \Omega_0 \Gamma_0^\mathsf{T}}). \tag{4.3}$$

The number of real parameters in model (4.3) is

$$N_u = u(r - u) + u + u(u + 1)/2 + (r - u)(r - u + 1)/2$$

$$= r(r + 1)/2 + u.$$

The first count $u(r - u)$ is the number of parameters needed to determine the envelope $\boldsymbol{\varepsilon_\Sigma}(\mathcal{M})$.

Now let us consider the maximum likelihood estimator (MLE) of μ. After some simplification, the log-likelihood can be written as (see Chapter 5 in [7]):

$$L_u(\boldsymbol{\Gamma}) = -(n/2) \log |\boldsymbol{\Gamma^\mathsf{T} S_Y \Gamma}| - (n/2) \log |\boldsymbol{\Gamma^\mathsf{T} T_Y^{-1} \Gamma}| + c, \tag{4.4}$$

where $\mathbf{S_Y}$ denotes the sample covariance of $\mathbf{Y}$ and $\mathbf{T_Y} = n^{-1} \sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i^\mathsf{T}$ denote the matrix of raw second moments of $\mathbf{Y}$, and $c = -(n/2) \log |\mathbf{T_Y}| -$

$nr/2$. For a given dimension u, the MLE of model (4.3) can be now to written as

$$\widehat{\varepsilon}_{\mathbf{\Sigma}(\mathcal{M})} = \operatorname{span}\{\arg\max L_u(\mathbf{\Gamma})\}$$

$$\widehat{\boldsymbol{\eta}} = \widehat{\mathbf{\Gamma}}^{\mathsf{T}}\bar{\mathbf{Y}}$$

$$\widehat{\mathbf{\Omega}} = \widehat{\mathbf{\Gamma}}^{\mathsf{T}}\mathbf{S_Y}\widehat{\mathbf{\Gamma}}$$

$$\widehat{\mathbf{\Omega}}_0 = \widehat{\mathbf{\Gamma}}_0^{\mathsf{T}}\mathbf{T_Y}\widehat{\mathbf{\Gamma}}_0$$

$$\widehat{\boldsymbol{\mu}} = \mathbf{P}_{\widehat{\mathbf{\Gamma}}}\bar{\mathbf{Y}}$$

$$\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Omega}}\widehat{\mathbf{\Gamma}}^{\mathsf{T}} + \widehat{\mathbf{\Gamma}}_0\widehat{\mathbf{\Omega}}_0\widehat{\mathbf{\Gamma}}_0^{\mathsf{T}},$$

where the maximum is over the set of all semi-orthogonal matrices $\mathbf{\Gamma} \in \mathbb{R}^{r\times u}$. From the standard likelihood theory, $\sqrt{n}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu})$ is asymptotically normal with mean $\mathbf{0}$ and variance $\operatorname{avar}(\sqrt{n}\widehat{\boldsymbol{\mu}})$, which has the form

$$\mathbf{\Sigma}_{\mathrm{env}} = \operatorname{avar}(\sqrt{n}\widehat{\boldsymbol{\mu}})$$

$$= \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^{\mathsf{T}} + (\boldsymbol{\eta}^{\mathsf{T}} \otimes \mathbf{\Gamma}_0)\mathbf{V}^{\dagger}(\boldsymbol{\eta} \otimes \mathbf{\Gamma}_0^{\mathsf{T}}) \qquad (4.5)$$

$$\leq \operatorname{avar}(\sqrt{n}\bar{\mathbf{Y}}) = \mathbf{\Sigma}.$$

After obtaining the envelope-based estimator of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$, the envelope-based Hotelling's T^2 statistics for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ can be written as

$$t_{\mathrm{env}}^2 = n(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0)^{\mathsf{T}}\widehat{\mathbf{\Sigma}}_{\mathrm{env}}^{-1}(\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0). \qquad (4.6)$$

Also, we have

$$\frac{n-r}{r(n-1)}t_{\mathrm{env}}^2 \sim F_{r,n-r}.$$

Then, the quantity on the left hand side can be used to evaluate the p-value using the F-distribution.

4.3   Likelihood ratio test

Consider the following 4 different models:

1. Model under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$;

2. Model under $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$;

3. Model under $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu_0}$ with $\boldsymbol{\mu_0} = \boldsymbol{\Gamma}\boldsymbol{\eta}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathsf{T}} + \boldsymbol{\Gamma_0}\boldsymbol{\Omega_0}\boldsymbol{\Gamma_0^{\mathsf{T}}}$;

4. Model under $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu_0}$ with $\boldsymbol{\mu} = \boldsymbol{\Gamma}\boldsymbol{\eta}$ and $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathsf{T}} + \boldsymbol{\Gamma_0}\boldsymbol{\Omega_0}\boldsymbol{\Gamma_0^{\mathsf{T}}}$.

The likelihood ratio test requires that the two models (under full and reduced model) be nested, which means that the more complex model can be transformed into the simpler model by imposing constraints on the former's parameters.

Model 4 and Model 1 are not nested. In fact, model 4 is nested within model 2 (imposing the constraints on $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$), model 1 is nested within model 2. These relationships can be summarized in the following diagram:
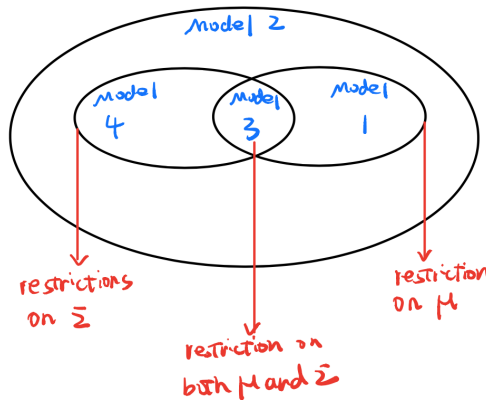
Traditionally, the hypothesis testings usually compare Model 1 and Model 2. Previously we have compared Model 4 and Model 1, however, they appear

to be not nested. So here we consider comparing Model 4 and Model 3. Specifically, we need to derive the MLE under these two models. Section 2 covered the scenario under Model 4, so here we derive the MLE under Model 3.

Under Model 3, the log-likelihood can be written as

$$
L_u(\boldsymbol{\Gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega_0}) = -\frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{n}{2}\log|\boldsymbol{\Omega_0}|
$$
$$
\qquad - \frac{1}{2}\sum_i (\mathbf{Y}_i - \boldsymbol{\Gamma\eta})^{\mathsf{T}}(\boldsymbol{\Gamma\Omega}^{-1}\boldsymbol{\Gamma}^{\mathsf{T}} + \boldsymbol{\Gamma_0\Omega_0^{-1}\Gamma_0^{\mathsf{T}}})(\mathbf{Y}_i - \boldsymbol{\Gamma\eta}) \qquad (4.7)
$$
$$
\boldsymbol{\Gamma\eta} = \boldsymbol{\mu}_0.
$$

The u linear constraints $\boldsymbol{\Gamma\eta} = \boldsymbol{\mu}_0$ here can be viewed as solving $\boldsymbol{\eta}$ given $\boldsymbol{\Gamma}$ and $\boldsymbol{\mu}_0$. Denote $L_{u,\lambda}$ as the penalized version of L_u, which has the following form:

$$
L_{u,\lambda}(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega_0}) = L_u(\boldsymbol{\Gamma}, \boldsymbol{\Omega}, \boldsymbol{\Omega_0}) + \lambda||\boldsymbol{\Gamma\eta} - \boldsymbol{\mu}_0||_2. \qquad (4.8)
$$

The penalization here is equivalent to a regression problem and the least square solution $\hat{\boldsymbol{\eta}} = (\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^{\mathsf{T}}\boldsymbol{\mu}_0$ minimizes the $l_2 - norm$. Substitute $\boldsymbol{\eta} = \hat{\boldsymbol{\eta}}$ into equation (4.7) and after some simplification, we have

$$
L_u(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega_0}) = -\frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\Gamma}^T\mathbf{Y}_i - \hat{\boldsymbol{\eta}})^T\boldsymbol{\Omega}^{-1}(\boldsymbol{\Gamma}^T\mathbf{Y}_i - \hat{\boldsymbol{\eta}})
$$
$$
\qquad - \frac{n}{2}\log|\boldsymbol{\Omega_0}| - \frac{1}{2}\sum_{i=1}^{n}\mathbf{Y}_i^T\boldsymbol{\Gamma_0\Omega_0^{-1}\Gamma_0^T}\mathbf{Y}_i. \qquad (4.9)
$$

Since $\boldsymbol{\Gamma}$ is semi-orthogonal and $\hat{\boldsymbol{\eta}} = (\boldsymbol{\Gamma}^T\boldsymbol{\Gamma})^{-1}\boldsymbol{\Gamma}^T\mu_0 = \boldsymbol{\Gamma}^T\mu_0$, we have

$$
L_u(\boldsymbol{\Gamma}, \boldsymbol{\eta}, \boldsymbol{\Omega}, \boldsymbol{\Omega_0}) = -\frac{n}{2}\log|\boldsymbol{\Omega}| - \frac{1}{2}\sum_{i=1}^{n}[\boldsymbol{\Gamma}^T(\mathbf{Y}_i - \mu_0)]^T\boldsymbol{\Omega}^{-1}[\boldsymbol{\Gamma}^T(\mathbf{Y}_i - \mu_0)]
$$
$$
\qquad - \frac{n}{2}\log|\boldsymbol{\Omega_0}| - \frac{1}{2}\sum_{i=1}^{n}\mathbf{Y}_i^T\boldsymbol{\Gamma_0\Omega_0^{-1}\Gamma_0^T}\mathbf{Y}_i.
$$
$$
\qquad (4.10)
$$

Taking derivative w.r.t. $\mathbf{\Omega}^{-1}$ and $\mathbf{\Omega}_0^{-1}$, and using S_{μ_0} to denote $\frac{1}{n}\sum_i(\mathbf{Y}_i - \mathbf{\mu}_0)^\intercal(\mathbf{Y}_i - \mathbf{\mu}_0)$ and $T_Y = \frac{1}{n}\sum_i \mathbf{Y}_i \mathbf{Y}_i^\intercal$, we have

$$
\begin{aligned}
\frac{\partial L_u}{\partial \mathbf{\Omega}^{-1}} &= \frac{n}{2}\mathbf{\Omega} - \frac{n}{2}\mathbf{\Gamma}^\intercal S_{\mu_0}\mathbf{\Gamma}; \\
\frac{\partial L_u}{\partial \mathbf{\Omega}_0^{-1}} &= \frac{n}{2}\mathbf{\Omega}_0 - \frac{n}{2}\mathbf{\Gamma}_0^\intercal T_Y \mathbf{\Gamma}_0.
\end{aligned}
\tag{4.11}
$$

So for fixed $\mathbf{\Gamma}$, $L_u(\mathbf{\Gamma}, \mathbf{\Omega}, \mathbf{\Omega_0})$ is maximized at $\mathbf{\eta} = \mathbf{\Gamma}^\intercal \mathbf{\mu}_0$, $\mathbf{\Omega} = \mathbf{\Gamma}^\intercal S_{\mu_0}\mathbf{\Gamma}$, $\mathbf{\Omega}_0 = \mathbf{\Gamma}_0^\intercal T_Y \mathbf{\Gamma}_0$. Substituting these relationships into equation (4.7) and simplifying leads to a partially maximized log-likelihood

$$
\begin{aligned}
L_u(\mathbf{\Gamma}) &= -\frac{n}{2}\log|\mathbf{\Gamma}^\intercal S_{\mu_0}\mathbf{\Gamma}| - \frac{n}{2}\log|\mathbf{\Gamma}_0^\intercal T_Y \mathbf{\Gamma}_0| - \frac{nr}{2} \\
&= -\frac{n}{2}\log|\mathbf{\Gamma}^\intercal S_{\mu_0}\mathbf{\Gamma}| - \frac{n}{2}\log|\mathbf{\Gamma}^\intercal T_Y^{-1}\mathbf{\Gamma}| + c,
\end{aligned}
\tag{4.12}
$$

where $c = -n/2\log|S_{\mu_0}| - nr/2$. For a given dimension u, the MLE under $H_0 : \mathbf{\mu} = \mathbf{\mu_0}$ can be summarized as

$$
\widehat{\mathbf{\varepsilon}}_{\mathbf{\Sigma}(\mathcal{M})} = \operatorname{span}\{\arg\max L_u(\mathbf{\Gamma})\}
$$

$$
\widehat{\mathbf{\eta}} = \widehat{\mathbf{\Gamma}}^\intercal \mathbf{\mu}_0
$$

$$
\hat{\mathbf{\mu}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\eta}}
$$

$$
\widehat{\mathbf{\Omega}} = \widehat{\mathbf{\Gamma}}^\intercal S_{\mu_0}\widehat{\mathbf{\Gamma}}
$$

$$
\widehat{\mathbf{\Omega}}_0 = \widehat{\mathbf{\Gamma}}_0^\intercal T_Y \widehat{\mathbf{\Gamma}}_0
$$

$$
\widehat{\mathbf{\Sigma}} = \widehat{\mathbf{\Gamma}}\widehat{\mathbf{\Omega}}\widehat{\mathbf{\Gamma}}^\intercal + \widehat{\mathbf{\Gamma}}_0\widehat{\mathbf{\Omega}}_0\widehat{\mathbf{\Gamma}}_0^\intercal,
$$

The LRT test statistics can be written as

$$
\begin{aligned}
\Lambda &= 2(L_u - L_{u_0}) \\
&= n\log|\widehat{\mathbf{\Sigma}}_{\mu_0}| + \sum_i(\mathbf{Y}_i - \widehat{\mathbf{\mu}}_0)^\intercal \widehat{\mathbf{\Sigma}}_{\mu_0}^{-1}(\mathbf{Y}_i - \widehat{\mathbf{\mu}}_0) - n\log|\widehat{\mathbf{\Sigma}}| \\
&\quad - \sum_i(\mathbf{Y}_i - \hat{\mathbf{\mu}})^\intercal \widehat{\mathbf{\Sigma}}^{-1}(\mathbf{Y}_i - \hat{\mathbf{\mu}}).
\end{aligned}
\tag{4.13}
$$

The Wilk's theorem states that for large sample sizes, Λ follows a chi-squared distribution with v degrees of freedom, where v is the number of free parameters under H_A not in H_0. In this case, it is equal to u.

4.4 Simulations and Data Analysis

4.4.1 Simulations

We first examine how the test statistics look like under H_0 and H_A. The following histograms are based on 1000 replications. The red solid curve is the null distribution $\chi^2(u)$, and the parameters of the experiment are shown in the bottom of each figure.

Here we compare the finite-sample performances of the proposed envelope-based test with the Hotelling's T^2 test. Based on (4.2), we can see that the envelope-based estimator $\widehat{\boldsymbol{\mu}}$ is more efficient than the sample mean $\bar{\mathbf{Y}}$ when $\boldsymbol{\mu}$ fall in an eigenspace of $\boldsymbol{\Sigma}$. We use the R package Renvlp [29] to compute the envelopes in all simulations.

First we consider a special case when $\boldsymbol{\mu}$ fall in the last eigenspace of $\boldsymbol{\Sigma}$ (the eigenvector of the smallest eigenvalue), the envelope should provide *massive* estimation gains in this case since the immaterial part in this case is huge. We randomly generate a $\boldsymbol{\Sigma}$ using $\boldsymbol{\Sigma} = V^\mathsf{T}V$, where $V_{ij} \sim N(0,1)$, so here $\boldsymbol{\Sigma}$ is randomly generated, and it has no specific structure. For $\boldsymbol{\mu}$, the three choices are: (1) $\boldsymbol{\mu}_0 = v_{\min}(\boldsymbol{\Sigma})$, where $v_{\min}(\boldsymbol{\Sigma})$ represents the eigenvector associated with the smallest eigenvalue of $\boldsymbol{\Sigma}$; (2) $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_0 + \max|\boldsymbol{\mu}_0|$, where $\max|\boldsymbol{\mu}_0|$ denotes the element in $\boldsymbol{\mu}_0$ with the largest absolute value. This situation represents a pure location shift of $\boldsymbol{\mu}$; and (3) $\boldsymbol{\mu}_2 = \boldsymbol{\mu}_0 + N(0, \max|\boldsymbol{\mu}_0|)$, here
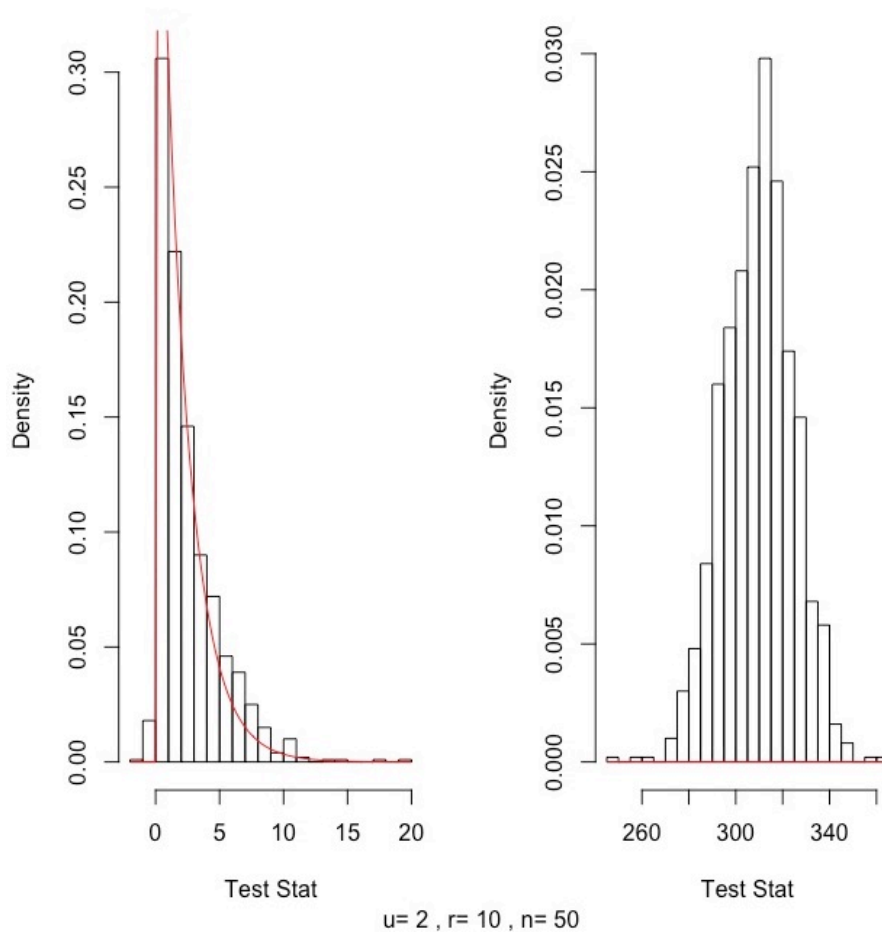
Figure 4.1: The histogram of the test statistics under null hypothesis H_0 (left) and alternative hypothesis H_A (right) with $u = 2, r = 10$ and $n = 50$. The red curve here is the density function of $F_{r,n-r}$.

Figure 4.2: The histogram of the test statistics under null hypothesis H_0 (left) and alternative hypothesis H_A (right) with $u = 2, r = 10$ and $n = 20$. The red curve here is the density function of $F_{r,n-r}$.
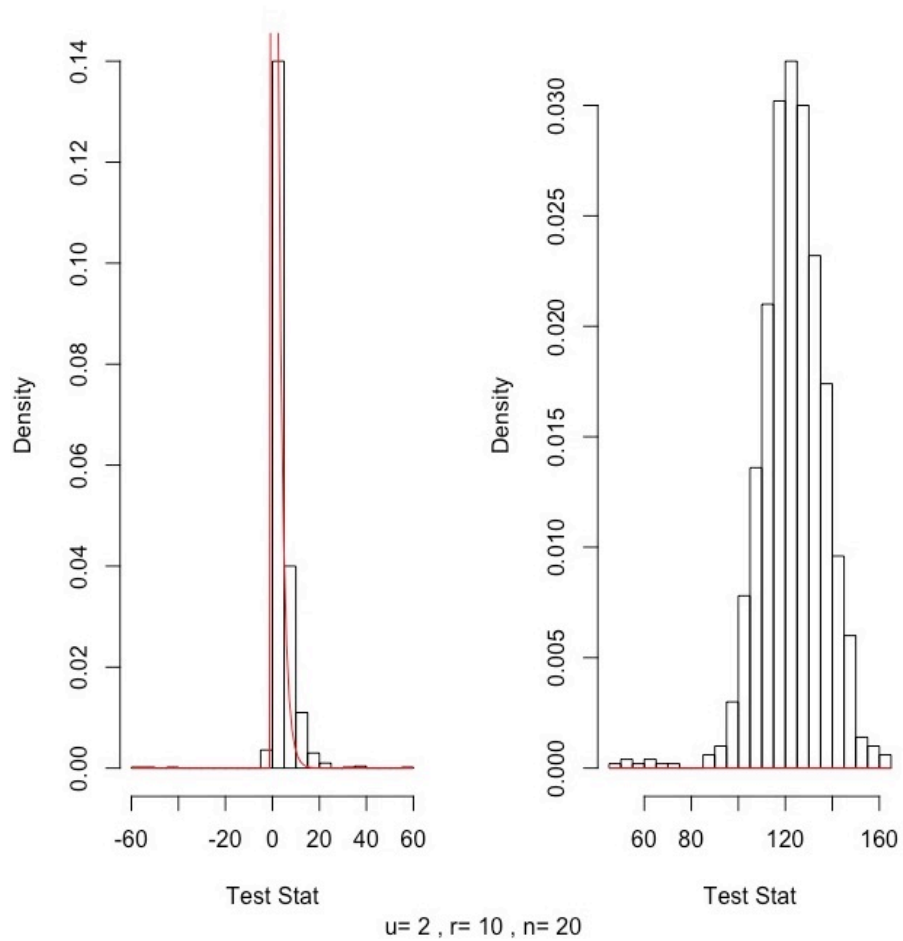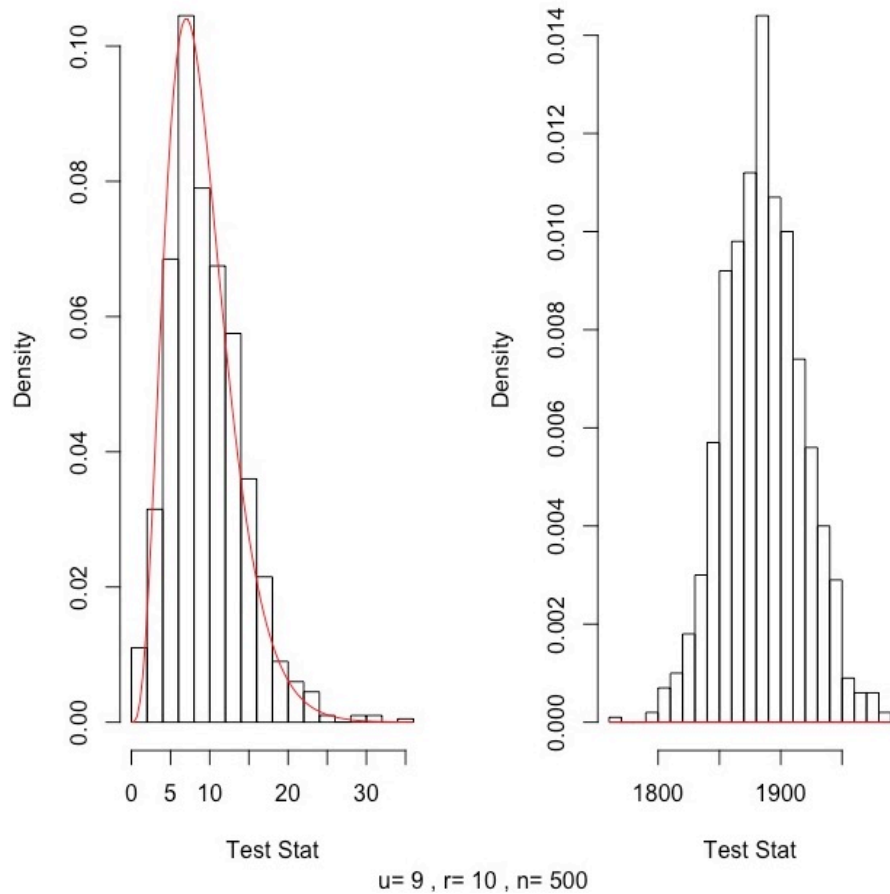
Figure 4.3: The histogram of the test statistics under null hypothesis H_0 (left) and alternative hypothesis H_A (right) with $u = 9, r = 10$ and $n = 500$. The red curve here is the density function of $F_{r,n-r}$.

we are mimicking a more stochastic change.

The last paragraph described a situation when we have the dimension of the envelope $u = 1$. Similarly, we also consider a more realistic situation when $\boldsymbol{\mu}$ can be expanded using a few eigenvectors of $\boldsymbol{\Sigma}$. Specifically, we consider $u = 5$.

Before we examine the performance of the proposed test, we first verify that the null distribution of the proposed envelope-based test statistics follows the reference distribution $F_{r,n-r}$. Figures 4.4 and 4.5 show the histogram of the test statistics for the Hotelling T^2 and the proposed envelope-based test when H_0 is true for small and large u, respectively. The red curve is the density function of the reference distribution $F_{r,n-r}$, and we can see that they match each other quite well.

Figure 4.4: The histogram of the test statistics for the Hotelling T^2 and proposed envelope test (**small** u) when H_0 is true. he red curve is the density function of $F_{r,n-r}$. The parameters of this experiment are specified on the bottom.
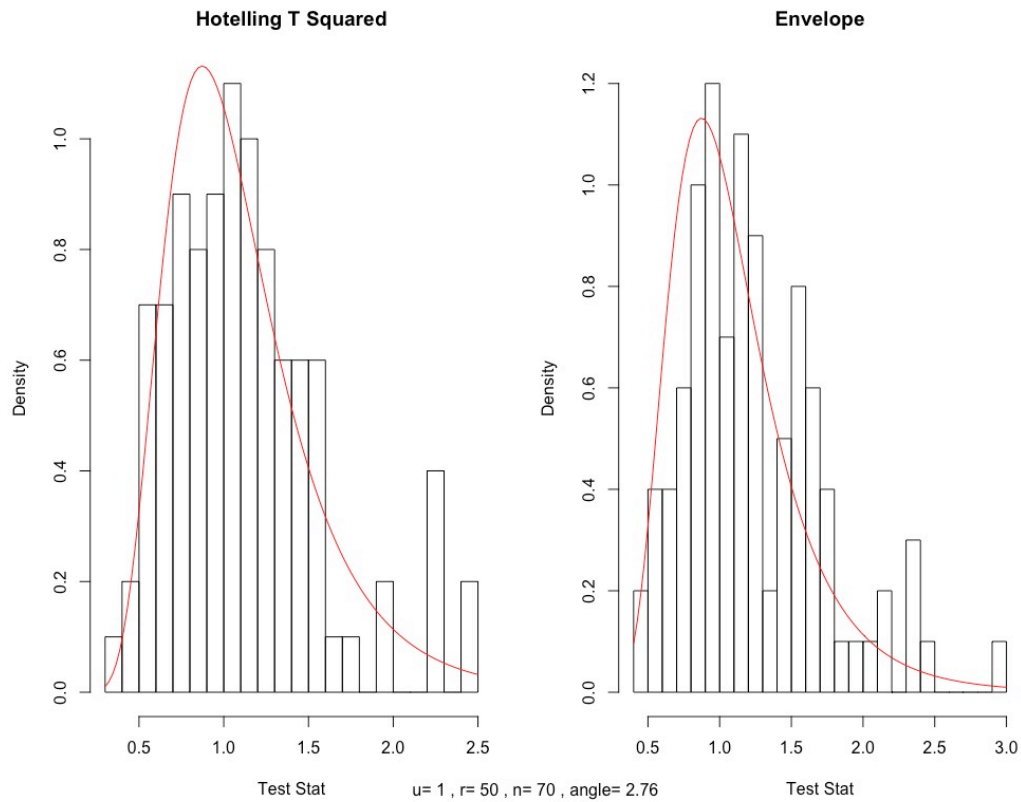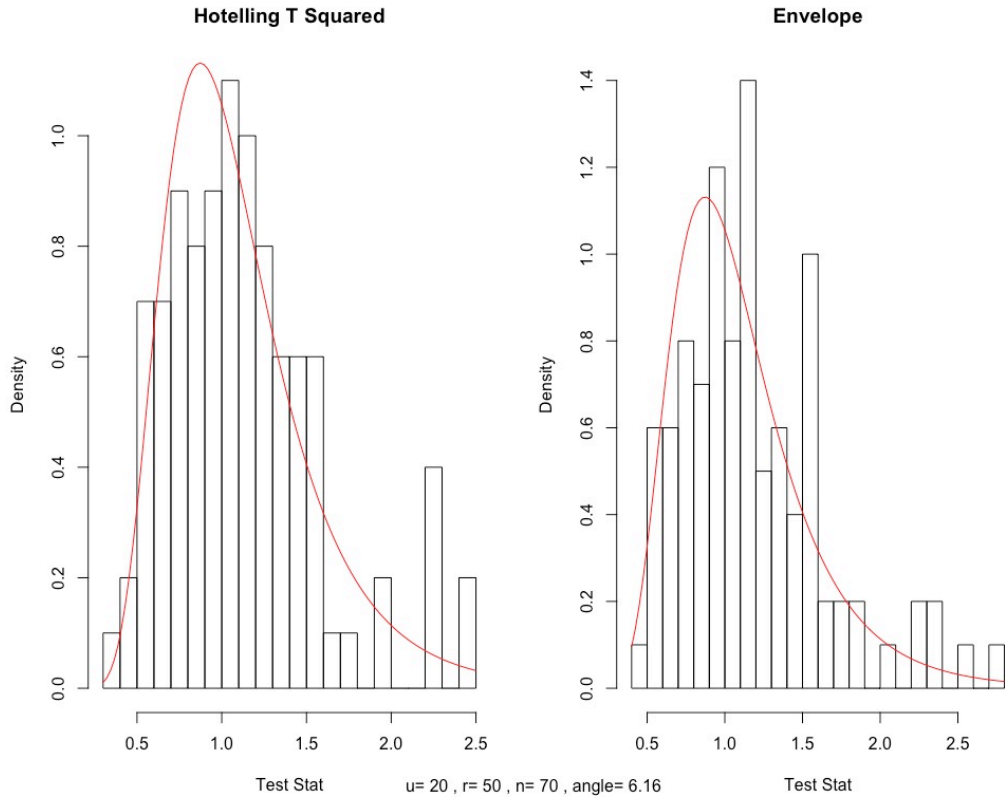
Figure 4.5: The histogram of the test statistics for the Hotelling T^2 and proposed envelope test (**large** u) when H_0 is true. he red curve is the density function of $F_{r,n-r}$. The parameters of this experiment are specified on the bottom.

For each simulation run, a multivariate normal distribution is simulated using one of the three $\boldsymbol{\Sigma}$ and the resulting $\boldsymbol{\mu}_0$ mentioned above. Three hypotheses were conducted with $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$, $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_2$, the first null hypothesis is used to access the *type I error* of the test; whereas, the other two tests can be used to access the *type II error* or *power* of the test. All simulation results are based on 500 repetitions. Here we present results for $r = 50$, and $n = 55, 60, 75, 250$. For the dimension of envelope u, we consider

Table 4.1: (a) the F norm between the true mean $\boldsymbol{\mu}$ and the estimated $\widehat{\boldsymbol{\mu}}$; (b) the angle between the estimated $\widehat{\boldsymbol{\Gamma}}$ and the true $\boldsymbol{\Gamma}$, which is the last eigenvector of $\boldsymbol{\Sigma}$.

		$u = 1$			
		$n = 55$	$n = 60$	$n = 75$	$n = 250$
	Hotelling's T^2	6.67	6.13	4.88	3.14
	Env $(u = 1)$	0.54	0.47	0.29	0.11
(a): $||\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0||_{\mathrm{F}}$	Env $(u = 3)$	0.80	0.63	0.41	0.20
	Env $(u = 8)$	1.46	1.57	1.68	1.46
	Env $(u = 1)$	4.21	3.68	2.29	0.88
(b): angle	Env $(u = 3)$	4.10	4.69	2.75	1.08
	Env $(u = 8)$	3.03	3.85	1.46	0.85
		$u = 5$			
		$n = 55$	$n = 65$	$n = 100$	$n = 250$
	Hotelling's T^2	6.67	0.85	0.68	0.43
	Env $(u = 5)$	4.91	0.45	0.29	0.17
(a): $||\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}_0||_{\mathrm{F}}$	Env $(u = 6)$	4.58	0.49	0.36	0.23
	Env $(u = 12)$	5.78	0.51	0.55	0.38
	Env $(u = 5)$	6.34	6.17	5.43	2.14
(b): angle	Env $(u = 6)$	5.62	5.87	5.07	1.58
	Env $(u = 12)$	4.63	4.46	4.43	1.04

the correct u, as well as a few u that are larger than the true one.

Table (4.1) provides a summary of the estimation process and Table (4.2) shows the performance in hypothesis testing. Based on both tables, we have the following main findings:

1. When the envelope dimension u is small, the proposed envelope test shows great power improvement compared with the Hotelling's T^2 test. The proposed test performs reasonably well when the sample size n is relatively small $(n = 55, \ r = 50)$

2. The advantage of the proposed test drops when u increases; this can be expected because the variance gain is smaller.

Table 4.2: Percentage of H_0 rejected using $\alpha = 0.05$.

		$n = 55$	$n = 60$	$n = 75$	$n = 250$
		$u = 1$			
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$	Hotelling's T^2	0.04	0.10	0.05	0.04
	Env $(u = 1)$	0.05	0.09	0.05	0.05
	Env $(u = 3)$	0.04	0.11	0.06	0.05
	Env $(u = 8)$	0.05	0.12	0.04	0.05
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$	Hotelling's T^2	0.08	0.24	0.26	1.00
	Env $(u = 1)$	0.98	1.00	1.00	1.00
	Env $(u = 3)$	0.98	1.00	1.00	1.00
	Env $(u = 8)$	0.58	0.78	0.81	1.00
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_2$	Hotelling's T^2	0.03	0.13	0.20	0.98
	Env $(u = 1)$	0.93	1.00	1.00	1.00
	Env $(u = 3)$	0.98	0.99	0.98	1.00
	Env $(u = 8)$	0.47	0.54	0.68	1.00
		$u = 5$			
		$n = 55$	$n = 60$	$n = 75$	$n = 250$
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$	Hotelling's T^2	0.05	0.08	0.06	0.04
	Env $(u = 5)$	0.06	0.05	0.06	0.06
	Env $(u = 6)$	0.05	0.06	0.07	0.05
	Env $(u = 10)$	0.06	0.06	0.05	0.07
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_1$	Hotelling's T^2	0.04	0.11	0.06	0.14
	Env $(u = 5)$	0.15	0.20	0.17	0.50
	Env $(u = 6)$	0.14	0.27	0.18	0.48
	Env $(u = 10)$	0.16	0.20	0.15	0.39
$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_2$	Hotelling's T^2	0.03	0.06	0.05	0.07
	Env $(u = 1)$	0.06	0.18	0.20	0.48
	Env $(u = 3)$	0.07	0.15	0.16	0.45
	Env $(u = 8)$	0.05	0.19	0.15	0.40

3. The performance of the test is tightly related to the MSE of $\widehat{\mu}$. A reason why the envelope-based test is more powerful is because the variance of $\widehat{\mu}$ is smaller than the variance of sample mean $\bar{x}$ (imagine when $r = 1$, we have two normal density functions with the same location, one with smaller variance and the other one with larger variance).

4.4.2 Data Analysis

In Section 3.6.2, we studied the ADNI data. One of the main goals in the ADNI study is to find phenotype related to Alzheimer's Disease (AD). In Chapter 3, we looked at the predictive power of different demographic, clinical and neurological variables. Specifically, the high-dimensional brain connectivity computed from the fMRI data was studied carefully to see whether the brain connectivity of patients with AD have different patterns. Instead of using the brain connectivity between different regions as predictors to predict the AD status, another approach to deal with it is to compare the brain connectivity matrix (or vector) between AD and normal patients.

In this specific application setting, the phenotype (response) we are considering here is the brain connectivity matrix. The brain connectivity matrix is computed by considering the similarities of the brain activities between different voxels (or regions). The dimension of the matrix depends on the parcellation of the brains; for example, in Chapter 3, we applied the AAL template, which has 116 ROIs. For each element in the matrix, different measures can be used to calculate the closeness of the time series between different regions. Typical measurements used include correlation, partial correlation and sparse correlation. Figure (4.6) visualizes the binary connectivity
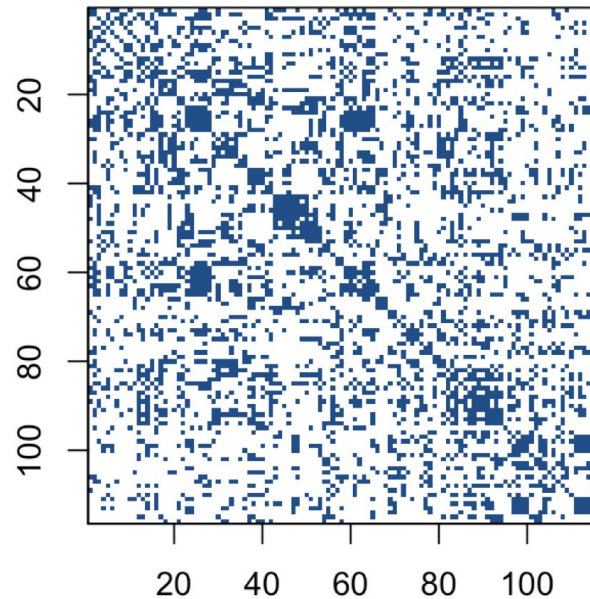
Figure 4.6: The connectivity matrix computed using correlation based on original BOLD signals with threshold 0.4 (Control ID 002-4225).

matrix using threshold 0.4 for a control subject, and Figure (4.7) for an AD subject. Different patterns can be found in certain "neighborhood" of these two matrices.

Since it is difficult to test whether the distributions of the matrices are the same or not, typically researchers focus on testing each individual element in the connectivity matrix, and it then becomes a multiple testing problem. However, this type of analysis fails to consider the connection and pattern among these brain regions. Another type of analysis uses some network-based
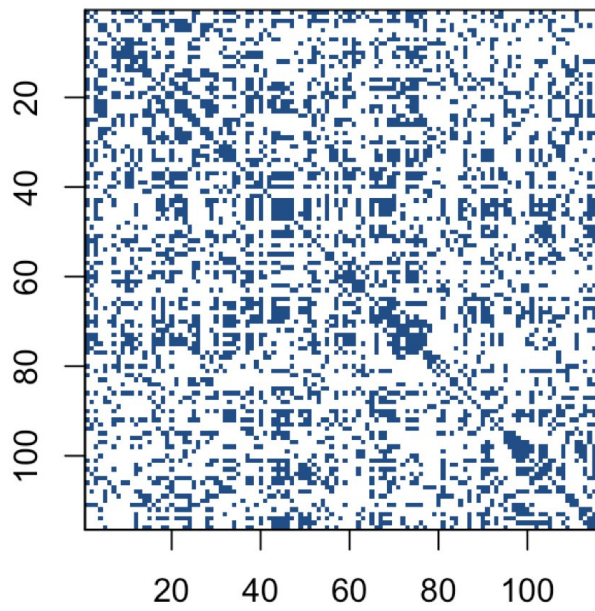
Figure 4.7: The connectivity matrix computed using correlation based on original BOLD signals with threshold 0.4 (AD ID 019-4477).

summary statistics calculated from the matrix, such as the number of small "neighborhoods" and the number of connected brain regions. In this analysis, we stack the brain connectivity matrix into a vector and do the hypothesis testing on the entire vector. Since the connectivity matrix is a symmetric matrix, we only take the lower half of the matrix and stack the columns of the matrix into a vector. Furthermore, since there is a small number of subjects available in the ADNI study, we first only consider selected the cerebellum region, which results in 14 ROIs. These 14 ROIs produce $14 \times 13/2 = 91$ brain connectivity measurements. For the control group, we have 33 subjects, and for the AD group, we have 24 subjects. In this case, the dimension of the response (r) is larger than the sample size (n), and the traditional Hotelling's T^2 test no longer applies. We apply the proposed envelope-based Hotelling's T^2 test. The tuning parameter u was chosen using AIC and we selected $u = 5$, which suggest that 91 closely related brain connectivity measures can be explained using an envelope structure with dimension 5. The reported p-value is 0.00345. This means that there are significant differences between the control group and the AD group using $\alpha = 0.05$. These results highlight the usefulness of the proposed envelope-based test statistics, especially in the scenario when the number of responses is larger than the number of observations, which is more and more common in this era of big data.

Chapter 5

Conclusion

The evolution of data acquisition technologies and computing power has allowed researchers nowadays to collect and store data with high dimensionality and complex structure much more efficiently. Examples can be found in gene expression microarray data, single nucleotide polymorphism (SNP) data, magnetic resonance imaging (MRI) data, high-frequency financial data, and others. Estimation and testing are unarguably the most fundamental tasks in statistical inference. As a rule of the thumb, with the increase of the dimensionality of the data, the amount of data needed to achieve certain level of statistical accuracy increases greatly. However, in practice, it is often too expensive or even impossible to collect the amount of data needed, for example, in the setting of clinical trials and neuroimaging experiments. The blessings come from high dimensional data is that many of the variables collected might be correlated with each other, and some of them are not relevant to the tasks. Thus it is of great importance to learn the inner structure of the data and find an efficient representation of them. Broadly speaking, dimension reduction represents the action of replacing data with a lower dimensional function of the data. In

this thesis, I studied a specific type of dimension reduction methods called envelopes.

In this thesis, I focused on a few issues that have not been properly addressed in the current literature. First, most of current envelope models are based on the multivariate linear regression model. The linearity assumption in there can be very strict and essentially uncheckable in practice. For example, it is very common to log transform a highly skewed variable such as white cell counts or viral loads to improve the linearity between responses and covariates. Other transformations such as Box-Cox transformation, Fisher's z transformation and variance stabilization transformation have been frequently used to improve a linear fit. The envelope models available in the current literature are not invariant or equivariant to these transformations, and even the dimension of the envelope will change. We have observed in practice that such transformations often resulting in envelope based estimators reduce back to the ordinary least square estimators, which means no efficiency gains can be achieved. In order to address this issue, I proposed to combine a Gaussian copula regression model and envelopes. The essential tool I used was the rank-based correlation estimators, which are invariant to continuous monotone transformations. Specifically, I focused on both response envelope and predictor envelope models. In Chapter 2, I studied the response envelope model and proposed a novel Gaussian copula response envelope model, which is equivariant to a large class of popular transformation of the responses. In Chapter 3, I focused on predictor envelope model, which aimed to achieve dimension reduction in the predictor space. Even though the goal is the same, the techniques used in Chapter 3 were very different from those used in Chapter 2.

The proposed Gaussian copula envelope models have the potential for bringing huge efficiency gains when the original envelope model fails to do so.

Most of the current literature on envelopes focus on regression models. In the area of hypothesis testing in high-dimensional data, envelope models actually can bring in huge efficiency gains. In Chapter 4, I explored the use of envelope models in the domain of hypothesis testing. Specifically, I focused on an envelope-based Hotelling's T^2 test and a likelihood ratio test. The proposed tests are more efficient when the high-dimensional mean vector can be represented using a lower dimensional envelope structure in the covariance matrix.

References

[1] Paul S Aisen et al. "Clinical Core of the Alzheimer's Disease Neuroimaging Initiative: progress and plans." In: *Alzheimer's and Dementia* 6.3 (2010), pp. 239–246.

[2] *Alzheimer's Disease Neuroimaging Initiative*. `http://adni.loni.usc.edu/`.

[3] T Tony Cai, Jing Ma, Linjun Zhang, et al. "CHIME: Clustering of high-dimensional Gaussian mixtures with EM algorithm and its optimality." In: *The Annals of Statistics* 47.3 (2019), pp. 1234–1267.

[4] T Tony Cai and Linjun Zhang. "High-dimensional Gaussian copula regression: Adaptive estimation and statistical inference." In: *Statistica Sinica* (2018), pp. 963–993.

[5] Raymond J Carroll and David Ruppert. *Transformation and weighting in regression*. Vol. 30. CRC Press, 1988.

[6] Anirvan Chakraborty, Probal Chaudhuri, et al. "Tests for high-dimensional data based on means, spatial signs and spatial ranks." In: *The Annals of Statistics* 45.2 (2017), pp. 771–799.

[7] R Dennis Cook. *An introduction to envelopes: dimension reduction for efficient estimation in multivariate statistics*. Vol. 401. John Wiley & Sons, 2018.

[8] R Dennis Cook, Liliana Forzani, and Zhihua Su. "A note on fast envelope estimation." In: *Journal of Multivariate Analysis* 150 (2016), pp. 42–54.

[9] R Dennis Cook, IS Helland, and Z Su. "Envelopes and partial least squares regression." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75.5 (2013), pp. 851–877.

[10] R Dennis Cook, Bing Li, and Francesca Chiaromonte. "Dimension reduction in regression without matrix inversion." In: *Biometrika* 94.3 (2007), pp. 569–584.

[11] R Dennis Cook, Bing Li, and Francesca Chiaromonte. "Envelope models for parsimonious and efficient multivariate linear regression." In: *Statistica Sinica* (2010), pp. 927–960.

[12] R Dennis Cook and Zhihua Su. "Scaled envelopes: scale-invariant and efficient estimation in multivariate linear regression." In: *Biometrika* 100.4 (2013), pp. 939–954.

[13] R Dennis Cook and Zhihua Su. "Scaled predictor envelopes and partial least-squares regression." In: *Technometrics* 58.2 (2016), pp. 155–165.

[14] R Dennis Cook and Xin Zhang. "Algorithms for envelope estimation." In: *Journal of Computational and Graphical Statistics* 25.1 (2016), pp. 284–300.

[15] R Dennis Cook and Xin Zhang. "Fast envelope algorithms." In: *Statistica Sinica* 28.3 (2018), pp. 1179–1197.

[16] Sijmen De Jong. "SIMPLS: an alternative approach to partial least squares regression." In: *Chemometrics and intelligent laboratory systems* 18.3 (1993), pp. 251–263.

[17] Shanshan Ding et al. "Envelope quantile regression." In: *Statistica Sinica* (2019), To appear.

[18] Jianqing Fan, Yang Feng, and Rui Song. "Nonparametric independence screening in sparse ultra-high-dimensional additive models." In: *Journal of the American Statistical Association* 106.494 (2011), pp. 544–557.

[19] Jianqing Fan and Runze Li. "Variable selection via nonconcave penalized likelihood and its oracle properties." In: *Journal of the American Statistical Association* 96.456 (2001), pp. 1348–1360.

[20] Jianqing Fan and Jinchi Lv. "Sure independence screening for ultrahigh dimensional feature space." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.5 (2008), pp. 849–911.

[21] Jianqing Fan, Richard Samworth, and Yichao Wu. "Ultrahigh Dimensional Feature Selection: beyond The Linear Model." In: *Journal of Machine Learning Research* 10 (Dec. 2009), pp. 2013–2038. ISSN: 1532-4435. URL: http://dl.acm.org/citation.cfm?id=1577069.1755853.

[22] Jianqing Fan, Rui Song, et al. "Sure independence screening in generalized linear models with NP-dimensionality." In: *The Annals of Statistics* 38.6 (2010), pp. 3567–3604.

[23] Long Feng, Changliang Zou, and Zhaojun Wang. "Multivariate-sign-based high-dimensional tests for the two-sample location problem." In: *Journal of the American Statistical Association* 111.514 (2016), pp. 721–735.

[24] Xuming He, Lan Wang, and Hyokyoung Grace Hong. "Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data." In: *The Annals of Statistics* 41.1 (2013), pp. 342–369.

[25] Yong He et al. "Robust feature screening for elliptical copula regression model." In: *Journal of Multivariate Analysis* 173 (2019), pp. 568–582.

[26] Clifford R Jack Jr et al. "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods." In: *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* 27.4 (2008), pp. 685–691.

[27] Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. "Robust multivariate nonparametric tests via projection-pursuit." In: *arXiv preprint arXiv:1803.00715* (2018).

[28] Kuang-Yao Lee, Bing Li, Francesca Chiaromonte, et al. "A general theory for nonlinear sufficient dimension reduction: Formulation and estimation." In: *The Annals of Statistics* 41.1 (2013), pp. 221–249.

[29] Minji Lee and Zhihua Su. "Renvlp: Computing Envelope Estimators." In: ().

[30] Minji Lee and Zhihua Su. "A Review of Envelope Models." In: *International Statistical Review* (2019).

[31] Bing Li, Jun Song, et al. "Nonlinear sufficient dimension reduction for functional data." In: *The Annals of Statistics* 45.3 (2017), pp. 1059–1095.

[32] Gaorong Li et al. "Robust rank correlation based screening." In: *The Annals of Statistics* (2012), pp. 1846–1877.

[33] Runze Li, Wei Zhong, and Liping Zhu. "Feature screening via distance correlation learning." In: *Journal of the American Statistical Association* 107.499 (2012), pp. 1129–1139.

[34] Jingyuan Liu, Wei Zhong, and Runze Li. "A selective overview of feature screening for ultrahigh-dimensional data." In: *Science China Mathematics* 58.10 (2015), pp. 1–22.

[35] *MRI SCANNER PROTOCOLS*. http://adni.loni.usc.edu/methods/documents/mri-protocols/.

[36] Yeonhee Park, Zhihua Su, and Hongtu Zhu. "Groupwise envelope models for imaging genetic analysis." In: *Biometrics* 73.4 (2017), pp. 1243–1253.

[37] Ernesto J Sanz-Arigita et al. "Loss of 'small-world' networks in Alzheimer's disease: graph analysis of fMRI resting-state functional connectivity." In: *PloS one* 5.11 (2010), e13788.

[38] Alexander Shapiro. "Asymptotic theory of overparameterized structural models." In: *Journal of the American Statistical Association* 81.393 (1986), pp. 142–149.

[39] Zhihua Su and R Dennis Cook. "Partial envelopes for efficient estimation in multivariate linear regression." In: *Biometrika* 98.1 (2011), pp. 133–146.

[40] Zhihua Su and R Dennis Cook. "Estimation of multivariate means with heteroscedastic errors using envelope models." In: *Statistica Sinica* 23.1 (2013), pp. 213–230.

[41] Robert Tibshirani. "Regression shrinkage and selection via the lasso." In: *Journal of the Royal Statistical Society. Series B (Methodological)* 1 (1996), pp. 267–288.

[42] Read D Tuddenham. "Physical growth of California boys and girls from birth to eighteen years." In: *University of California publications in child development* 1 (1954), pp. 183–364.

[43] Nathalie Tzourio-Mazoyer et al. "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain." In: *Neuroimage* 15.1 (2002), pp. 273–289.

[44] Lan Wang, Bo Peng, and Runze Li. "A high-dimensional nonparametric multivariate test for mean vector." In: *Journal of the American Statistical Association* 110.512 (2015), pp. 1658–1669.

[45] Herman Wold. "Estimation of principal components and related models by iterative least squares." In: *Multivariate analysis* (1966), pp. 391–420.

[46] Cun-Hui Zhang. "Nearly unbiased variable selection under minimax concave penalty." In: *The Annals of Statistics* 38.2 (2010), pp. 894–942.

[47] Yue Zhao, Christian Genest, et al. "Inference for elliptical copula multivariate response regression models." In: *Electronic Journal of Statistics* 13.1 (2019), pp. 911–984.

[48] Li-Ping Zhu et al. "Model-free feature screening for ultrahigh-dimensional data." In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1464–1475.

[49] Hui Zou. "The adaptive lasso and its oracle properties." In: *Journal of the American statistical association* 101.476 (2006), pp. 1418–1429.

[50] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net." In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005), pp. 301–320.

Appendix

Proof of Theorem 2.5.1. In order to use Proposition 3.1 in Shapiro (1986) [38], we first match our notations to the ones in Shapiro's. The θ_s in Shapiro's context is our $\phi = \{\lambda^T, \text{vec}(\mathbf{\Gamma})^T, \text{vech}(\mathbf{\Omega})^T, \text{vech}(\mathbf{\Omega}_0)^T\}^T$. Shapiro's $\hat{x}_s$ corresponds to our $\{\text{vec}(\widehat{\boldsymbol{\beta}})^T, \text{vech}(\widehat{\mathbf{\Sigma}}^T)\}^T$, and Shapiro's ξ_s is $\{\text{vec}(\boldsymbol{\beta})^T, \text{vech}(\mathbf{\Sigma})^T\}^T$ in our context. The discrepancy function F_s in Shapiro's is our log likelihood function, after omitting a constant factor

$$
\begin{aligned}
F_s = L_1/n &= -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\text{tr}\{(U - F\boldsymbol{\beta}^T)\mathbf{\Sigma}^{-1}(U - F\boldsymbol{\beta}^T)^T/n\} \\
&= -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}\text{tr}[\mathbf{\Sigma}^{-1}\{n\tilde{\mathbf{\Sigma}}_{res} + (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})(F^T F/n)(\widehat{\boldsymbol{\beta}}^T - \boldsymbol{\beta}^T)\}].
\end{aligned}
$$

As F_s is constructed under normal likelihood function, it satisfies the conditions 1-4 in Shapiro's. Shapiro's Δ_s is the gradient matrix $\partial\xi_s/\partial\theta_s$, which is equivalent to H in our context. Consider $e = U - F\boldsymbol{\beta}^T$, Shapiro's $V_s = \text{bdig}\{(F^T F/n) \otimes \mathbf{\Sigma}^{-1}, E_r^T(\mathbf{\Sigma}^{-1} \otimes \mathbf{\Sigma}^{-1})E_r/2\}$ is half of the Hessian matrix $\partial^2 F_s/\partial\xi_s\partial\xi_s^T$ evaluated at (ξ_s, ξ_s). Since $\sum_{i=1}^n X_i X_i^T/n > 0$, V_s is full rank and $\text{rank}(\Delta_s^T V_s \Delta_s) = \text{rank}(\Delta_s)$. Therefore, all conditions in Proposition 3.1 in Shapiro (1986) are satisfied and the maximizers $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{\Sigma}}$ are uniquely defined.

Proof of Theorem 2.5.2 As we have over-parameterization in $\mathbf{\Gamma}$, we use Proposition 4.1 in Shapiro (1986) [38]. The conditions of Proposition 4.1 are

the same as those in Proposition 3.1 of Shapiro (1986) [38], except with an additional assumption that $n^{1/2}(\hat{x}_s - \xi_s)$ is asymptotically normal. Since the condition on p_{ii} guarantees that the asymptotic distribution of $n^{1/2}\{(\text{vec}(\widehat{\boldsymbol{\beta}})^T, \text{vech}(\widehat{\boldsymbol{\Sigma}}_{res})^T)^T - (\text{vec}(\boldsymbol{\beta})^T, \text{vech}(\boldsymbol{\Sigma})^T)^T)\}$ is multivariate normal, so this additional assumption is also satisfied. Following Shapiro's notation, the asymptotic variance has the form

$$\Delta_s(\Delta_s^T V_s \Delta_s)^\dagger \Delta_s^T V_s \Gamma_s V_s \Delta_s (\Delta_s^T V_s \Delta_s)^\dagger \Delta_s^T,$$

where Shapiro's Δ_s is the asymptotic variance of $\{(\text{vec}(\widehat{\beta})^T, \text{vech}(\widehat{\boldsymbol{\Sigma}}_{res})^T)^T\}$. With the additional assumption of normality, Shapiro's $\Gamma_s = V_s^{-1}$. Therefore the asymptotic covariance matrix has the form $\Delta_s(\Delta_s^T V_s \Delta_s)^\dagger \Delta_s^T$, which is $V = H(H^H J H)^\dagger H^T$ in our notation. Since V is invariant under full rank linear transformations of the columns of H, we next transform the columns of H by the non-singular matrix

$$T = \begin{pmatrix} I_{r-1} & 0 \\ -(H_2^T J H_2)^\dagger H_2^T J H_1 & I_{r(r+1)/2} \end{pmatrix}.$$

Then $HT = (Q_{H_2(J)} H_1, H_2)$ and

$$T^T H^T J H T = \text{bdig}(H_1^T Q_{H_2(J)}^T J Q_{H_2(J)} H_1, G_o^T J_o G_o).$$

We then have

$$V = HT(T^T H^T J H T)^\dagger T^T H^T = J^{-1/2} P J^{-1/2} + D_\lambda G_o (G_o^T J_o G_o)^\dagger G_o^T D_\lambda^T,$$

where P is the projection onto the span of $J^{1/2} Q_{H_2(J)} H_1$. The second term on the right of the last expression is the same as V_2 stated in Proposition 4.1 in Shapiro's.

Proof of Theorem 3.5.1 Similar to the proof in Theorem 2.5.2, due to the over-parameterization in Γ, we apply Proposition 4.1 in Shapiro (1986)[38]. The conditions of Proposition 4.1 are the same as those in Proposition 3.1 of Shapiro (1986) [38], except with an additional assumption that $n^{1/2}(\hat{h} - h)$ is asymptotically normal. This requires that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta})$ converge in distribution to a multivariate normal distribution. Since $\hat{\boldsymbol{\beta}}_{ols} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$, and $\mathbf{X}^T\mathbf{X}/n$ converges in probability to $\boldsymbol{\Sigma}_{\mathbf{X}}$. Since $(\mathbf{Y}, \mathbf{X})$ has finite the fourth moment, the sequence $\sqrt{n}(\mathbf{X}^T\mathbf{Y}/n - \boldsymbol{\Sigma}_{\mathbf{XY}})$ converges in distribution. By Slutsky's theorem, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution to a multivariate normal distribution. Similarly we have $\sqrt{n}(\hat{h} - h)$ is multivariate normal. Let $\hat{h}$ be the copula predictor envelope estimator of h, and then $\hat{h}$ is a consistent estimator of h, and $\sqrt{n}(\hat{h} - h)$ is asymptotically normally distributed. As $\sqrt{n}[\{\text{vec}^T(\widehat{\boldsymbol{\beta}}_{cpe}), \text{vech}^T(\widehat{\Sigma}_{\mathbf{X},cpe})\}^T - \{\text{vec}^T(\boldsymbol{\beta}), \text{vech}^T(\Sigma_{\mathbf{X}})\}^T]$ is asymptotically normally distributed, which proves Theorem 3.5.1.

Proof of Theorem 3.5.2 According to Proposition 4.1 in Shapiro (1986), assuming normality, the asymptotic variance of the proposed copula predictor envelope estimator of h has the form $H^*(H^{*T}J^*H^*)^{\dagger}H^{*T}$. As J^* and H^* both have block diagonal structure, the asymptotic variance of $\{\text{vec}^T(\widehat{\boldsymbol{\beta}}), \text{vech}(\widehat{\Sigma}_{\mathbf{X}})\}^T$ is $H(H^TJH)^{\dagger}H^T$, where

$$J = \begin{pmatrix} \boldsymbol{\Sigma}_{\mathbf{Y}|\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}} & 0 \\ 0 & 0.5E_p^T(\boldsymbol{\Sigma}_{\mathbf{X}}^{-1} \otimes \boldsymbol{\Sigma}_{\mathbf{X}}^{-1})E_p \end{pmatrix},$$

and H^T has the form

$$\begin{pmatrix} -(\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\Lambda^{-1} \otimes \Lambda^{-1})L & 0 \\ I_r \otimes \Lambda^{-1}\boldsymbol{\Gamma} & 0 \\ \boldsymbol{\eta}^T \otimes \boldsymbol{\Gamma}^{-1} & 2C_p(\Lambda\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes \Lambda - \Lambda\boldsymbol{\Gamma} \otimes \Lambda\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^T) \\ 0 & C_p(\Lambda\boldsymbol{\Gamma} \otimes \Lambda\boldsymbol{\Gamma})E_u \\ 0 & C_p(\Lambda\boldsymbol{\Gamma}_0 \otimes \Lambda\boldsymbol{\Gamma}_0)E_{p-u} \end{pmatrix}.$$

Let $H = (H_1, H_2)$, where H_1 is the first column of H. Define $H_2 = DH_0$, where

$$D = \begin{pmatrix} I_r \otimes \Lambda^{-1} & 0 \\ 0 & C_p(\Lambda \otimes \Lambda)E_p \end{pmatrix},$$

and

$$H^T = \begin{pmatrix} I_r \otimes \boldsymbol{\Gamma} & 0 \\ \boldsymbol{\eta}^T \otimes I_p & 2C_p(\boldsymbol{\Gamma}\boldsymbol{\Omega} \otimes I_p - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0 \boldsymbol{\Gamma}_0^T) \\ 0 & C_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})E_u \\ 0 & C_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)E_{p-u} \end{pmatrix},$$

then we have $H_2(H_2^T J H_2)^\dagger = DH_o(H_o^T J_o H_o)^\dagger H_o^T D^T$, where $J_o = D^T JD$. Let

$$T = \begin{pmatrix} I_{p-1} & 0 \\ -(H_2^T J H_2)^\dagger H_2^T J H_1 & I_{p(p+1)/2} \end{pmatrix},$$

and then we have $HT = (H_1 - P_{H_2(J)}H_1, H_2) = (Q_{H_2(J)}H_1, H2)$. Then we have

$$T^T H^T J H T = \begin{pmatrix} H_1^T Q_{H_2(J)}^T J Q_{H_2(J)} H_1 & 0 \\ 0 & H_2^T J H_2 \end{pmatrix}.$$

Since $(T^T H^T J H T)^\dagger = \text{bdiag}((H_1^T Q_{H_2(J)}^T J Q_{H_2(J)} H_1)^\dagger, (H_2^T J H_2)^\dagger)$, we have

$$H(H^T J H)^\dagger H^T = Q_{H_2(J)}H_1(H_1^T Q_{H_2(J)}^T J Q_{H_2(J)} H_1)^\dagger H_1^T Q_{H_2(J)}^T$$

$$+ DH_o(H_o^T J_o H_o)^\dagger H_o^T D^T$$

$$\equiv A + B.$$

In B, the upper left $pr \times pr$ block of $H_o(H_o^T J_o H_o)^\dagger H_o^T$ is equal to the asymptotic variance of $\text{vec}(\widehat{\boldsymbol{\beta}}_o)$, which does not depend on scaling Λ, hence the upper left $pr \times pr$ block of B is the asymptotic variance of $(I_p \otimes \Lambda^{-1})\text{vec}(\widehat{\boldsymbol{\beta}}_o)$, which is the cost of estimating $\boldsymbol{\beta}$ when Λ is known.

In A, $Q_{H_2(J)}^T = DQ_{H_o(J_o)}D^{-1}$. Also

$$Q_{H_2(J)}^T J Q_{H_2(J)} = D^{-T} Q_{H_o(J_o)}^T J_o Q_{H_o(J_o)} D^{-1}.$$

Now we simplify $D^{-1}H_1$,

$$D^{-1}H_2 = \text{bdiag}(I_r \otimes \Lambda, C_p(\Lambda^{-1} \otimes \Lambda^{-1})E_p) \begin{pmatrix} -(\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\Lambda^{-1} \otimes \Lambda^{-1}) \\ 2C_p(\Lambda\boldsymbol{\Sigma}_o \otimes I_p) \end{pmatrix} L$$

$$= \begin{pmatrix} -\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T\Lambda_p^{-1} \\ 2C_p(\boldsymbol{\Sigma}_o \otimes \Lambda^{-1}) \end{pmatrix} L$$

$$= \text{bdiag}(-\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T \otimes I_p, 2C_p(\boldsymbol{\Sigma}_o \otimes I_p)) \begin{pmatrix} \Lambda^{-1} \otimes I_p \\ I_p \otimes \Lambda^{-1} \end{pmatrix} L$$

$$= \text{bdiag}(-\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T \otimes I_p, 2C_p(\boldsymbol{\Sigma}_o \otimes I_p))(1_2 \otimes L)\Lambda_1^{-1}$$

$$\equiv K\lambda_1^{-1},$$

where $\Lambda_1^{-1} = \text{diag}\{\lambda_1^{-1}, \ldots, \lambda_{q-1}^{-1}\}$, $1_2 = (1,1)^T$, and $K = \text{bdiag}(-\boldsymbol{\eta}^T\boldsymbol{\Gamma}^T \otimes I_p, 2C_p(\boldsymbol{\Sigma}_o \otimes I_p))(1_2 \otimes L)$. Since K does not depend on Λ, then $A = DG(G^T J_o G)^{\dagger T} D^T$. Finally we have the asymptotic variance of $\{\text{vec}^T(\widehat{\boldsymbol{\beta}}), \text{vech}(\widehat{\boldsymbol{\Sigma}}_{\mathbf{X}})\}^T$ has the form $A + B$. This completes the proof of Theorem 3.5.2.