

ABSTRACT

GHOSAL, RAHUL. Hypothesis Testing and Variable Selection in Functional Concurrent Regression Model. (Under the direction of Arnab Maity.)

Functional regression is an active area of research in functional data analysis with a vast literature to address problems where the response variable, the predictors, or both are functions over some continuous index such as time. In this dissertation, we develop hypothesis testing and variable selection methods for functional concurrent regression models. Chapter 1 gives a general introduction to functional concurrent regression models and a brief outline of this thesis.

Functional linear concurrent model has been the most frequently used functional concurrent regression model with several applications in longitudinal and spatiotemporal studies. Although there exist multiple methods for estimation in such model, methods for testing the global effect of predictors are relatively few. A novel method for testing the null hypothesis of no effect of a covariate on the response is proposed in Chapter 2. Asymptotic distribution of the test statistic is established under null using standard assumptions. Numerical simulations are used to illustrate the performance of the proposed testing method in terms of type I error rate and power. Applications of the proposed testing method are demonstrated on real data.

In Chapter 3, a method for variable selection is developed for the functional linear concurrent regression model. The work is motivated by a fisheries footprint study, with a goal to select the relevant predictors influencing fisheries footprint over time as well as to estimate their dynamic effects. We extend the classically used scalar on scalar variable selection methods like LASSO, SCAD and MCP for this purpose. Our proposed penalty on the coefficient functions penalizes both their roughness and departure from sparsity. Numerical simulations are provided to illustrate satisfactory selection performance and estimation accuracy of the proposed method. Finally, the proposed method is applied to real data for selection of influential time varying covariates and estimating their concurrent effects.

In Chapter 4, we extend the proposed method for performing variable selection in non-

parametric functional concurrent regression. Nonparametric functional regression models are useful for capturing complex relationships present between response and covariates, especially if the effects are non-linear. We model the nonparametric functions in terms of bi-variate basis expansion using a tensor product of univariate basis functions and impose group penalty on the basis coefficients to perform variable selection. Numerical analysis using simulations study illustrates satisfactory performance of the proposed method in terms of selection accuracy and out of sample prediction accuracy. Real data applications are provided on a dietary calcium absorption study and a bike sharing data.

© Copyright 2019 by Rahul Ghosal

All Rights Reserved

Hypothesis Testing and Variable Selection in Functional Concurrent Regression Model

by
Rahul Ghosal

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2019

APPROVED BY:

Luo Xiao

Eric Chi

Stefano B Longo

Arnab Maity
Chair of Advisory Committee

DEDICATION

To Ma, Baba, and Enakshi for being there for me, always.

BIOGRAPHY

Rahul Ghosal was born on November 12, 1992 in Uttarpara, a small suburb in West Bengal, India. He received his schooling from three different institutions; Uttarpara Children's Own Home (till 4th standard), Baranagore Ramakrishna Mission Ashrama High School (till 10th standard) and Uttarpara Govt. High School (till 12th standard), each of which played an important role in his education both academic and outside. He got his Bachelor's (B.Stat.) and Master's (M.Stat.) degree in Statistics from Indian Statistical Institute, Kolkata. Rahul joined the Department of Statistics at North Carolina State University as a graduate student in 2016.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Arnab Maity for his constant support and guidance throughout my PhD program at NC State and believing in me. Be it my research or mentoring me, he was always just one door knock away. I would like to thank Dr. Luo Xiao and Dr. Eric Chi for being being part of my advisory committee and giving me guidance me both in my research and my career. I would also like to thank Dr. Stefano B Longo for being in my advisory committee and helping me out during our collaboration. I would like to thank Dr. Sujit Ghosh for working with me and mentoring me in realizing my future career goals. I would like to thank Dr. Donald Martin and Dr. Charles Smith for their ever smiling and encouraging presence whenever I met them. I am also very grateful to professors at NC State, Drs. Dennis Boos, Jacqueline Hughes-Oliver, Subhashis Ghoshal, Soumendra Nath Lahiri, Marie Davidian, Donna Barton, Paul Savariappan for their teaching and the example they set. I would like to thank Terry Byron for helping me out with any computing difficulties I faced during my research. I would specially like to thank Alison and Lanakila for being department moms and taking care of us, without them nothing would work. I would like to thank Dr. Debasis Sengupta at Indian Statistical Institute for his guidance and being an inspiration to look up to. I would like to thank Dr. Sedigheh Mirzaei Salehabadi and Timothy Clark for being amazing collaborators to work with. I would like to mention and thank my Mathematics and Statistics teacher at Uttarpara Govt. High School, Mr. Abhijit Sengupta who inspired me and made me fall in love with Mathematics and Statistics.

My time at Raleigh has been a memorable experience thanks to my friends here, who have made this feel a home away from home. I have spent some wonderful time and made some amazing memories with Suman, Indrabati, Sahoooda, Golda, Arkopalda, Hazrada, Sohinidi, Moumitadi, Chak, Salilda, Dhrubajyoti (DJG), Priyam (Ala) da, Tuhin, Sukanya, Rahul, Sayak, Abhishekda over the years.

I want to thank Enakshi, who has been a constant support in my highs and lows and has

always been there with me in everything. I want to thank her and my two best friends Debarun and Rajit for still keeping the child inside me alive and just being crazy with me in general. I want to thank my friends back home, for just being my friends, the Danesh Eleven (maybe more than eleven); you know who you are, I love you guys and miss you. Lokkhichanas I miss you too.

I want to thank all my relatives and cousins for their love, motivation and support in everything I do. I want to thank my Kaka, Kakima, Pisimoni for their blessings and bhai Arka, who has grown up in front of my eyes and is one of my dearest friend now. I want to mention my grandfather late Shri Gourchandra Ghosal, under whom my education started. I want to remember my grandmother late Smt. Chayarani Ghosal, she was an important part of my growing up. I really miss her.

Finally, I want to say I owe all of my achievements to my parents, Mrs. Mita Ghosal and Mr. Gautam Ghosal. Their support, hard work and guidance have made me who I am today. I really hope I can become as good a human being as they are. I learnt it is important to say to your parents how much you love them. Ma, Baba I love you.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
Chapter 1 Introduction	1
Chapter 2 A Score Based Test for Functional Linear Concurrent Regression .	5
2.1 Introduction	5
2.2 Methodology	8
2.2.1 Modeling framework	8
2.2.2 Equivalent random effects model	9
2.2.3 Testing method	11
2.2.4 Estimation of Covariance Matrix	14
2.2.5 Extension to Sparse and Noisy Covariate	15
2.2.6 Extension to multiple covariates	17
2.3 Simulation Study	18
2.3.1 Study design	18
2.3.2 Simulation Results	19
2.4 Real Data Applications	24
2.4.1 Gait Data	25
2.4.2 Calcium Absorption Data	27
2.5 Discussion and Future Work	29
Chapter 3 Variable Selection in Functional Linear Concurrent Regression . . .	31
3.1 Introduction	31
3.2 Methodology	33
3.2.1 Modeling Framework and Variable Selection Method	33
3.2.2 Incorporating Covariance Structure into variable selection	37
3.2.3 Extension to Sparse data and Noisy Covariates	39
3.3 Simulation Study	40
3.3.1 Simulation Set Up	40
3.3.2 Simulation Results	41
3.4 Real Data Applications	46
3.4.1 Study of Dietary Calcium Absorption	46
3.4.2 Study of Fisheries Footprint	50
3.5 Discussion	57
Chapter 4 Variable Selection in Nonparametric Functional Concurrent Re-	
gression	59
4.1 Introduction	59
4.2 Methodology	61
4.2.1 Modeling Framework and Variable Selection Method for NPFCM	61

4.2.2	Extension to Sparse and Noisy Data	65
4.3	Simulation Study	66
4.3.1	Simulation Set Up	66
4.3.2	Simulation Results	67
4.4	Real Data Applications	71
4.4.1	Study of Dietary Calcium Absorption	71
4.4.2	Study of Bike Sharing Data	73
4.5	Discussion	77
Chapter 5 Conclusion		79
BIBLIOGRAPHY		81
APPENDIX		86
Appendix A	Proofs and Additional Results	87
A.1	Proof of Theorem 1	87
A.2	Proof of Theorem 2	88
A.3	Asymptotic distribution of p-value under null	91

LIST OF TABLES

Table 2.1	Estimated type I error along with their standard error at $\alpha = 5\%$ level. . .	20
Table 3.1	Comparison of selection percentages (%) of different variables and average model size, without pre whitening.	43
Table 3.2	Comparison of selection percentages (%) of different variables and average model size, with pre whitening.	44
Table 3.3	Comparison of MC absolute bias and mean square error.	45
Table 3.4	Selection Percentages (%) of variables in Calcium absorption Study.	48
Table 3.5	List of covariates in the Fisheries Footprint Study.	52
Table 4.1	Comparison of selection percentages (%) of different variables and mean model size, Scenario A.	68
Table 4.2	Quantiles of out of sample R_{Med}^2 based on MC simulation, Scenario A. . .	69
Table 4.3	Comparison of selection percentages (%) of different variables, Scenario B. . .	70
Table 4.4	Quantiles of out of sample R_{Med}^2 based on MC simulation, Scenario B. . .	71
Table 4.5	Selection Percentages (%) of variables in Calcium absorption Study using NPFCM.	73
Table 4.6	Selection Percentages (%) of variables in Bike Sharing Study.	75

LIST OF FIGURES

Figure 2.1	Results of simulation study as described in Section 2.3, Scenario A. Displayed are the power curves for our proposed procedure (solid line) and the bootstrap-F test based method (dashed line) for dense (left column) and sparse (right column) sampling designs with sample sizes $n = 100$ (top row) and $n = 300$ (bottom row).	21
Figure 2.2	Power curve for simulation Scenario A, additional simulation study, $n=39$, $m=20$	22
Figure 2.3	Results of simulation study as described in Section 2.3, Scenario B. Displayed are the power curves for our proposed procedure with sample sizes $n = 100$ (dashed line) and $n = 300$ (solid line) for dense (left column) and sparse (right column) sampling designs.	23
Figure 2.4	Effect of number of basis on the power of the test, simulation Scenario A, $n=300$	24
Figure 2.5	Measurement of hip angles and knee angles in the gait study.	25
Figure 2.6	Results from simulation study mimicking gait data. Displayed is the power curve of our test.	27
Figure 2.7	Observed calcium absorption and calcium intake in the calcium absorption study.	28
Figure 3.1	MC estimates and point wise confidence intervals of the coefficient functions ($n=200$).	42
Figure 3.2	Calcium absorption and covariate profiles of patients along their ages.	47
Figure 3.3	Bootstrap estimate and point wise confidence interval of the the effect of calcium intake (left panel) and percentage observation available in different age groups (right panel).	49
Figure 3.4	Fisheries footprint of the nations over year 1970-2009.	52
Figure 3.5	Mean fisheries footprint along with their 95% confidence interval.	53
Figure 3.6	Estimate of the linear concurrent effect of GDP per capita on fisheries footprint (SCAD).	54
Figure 3.7	Estimate of the linear concurrent effect of urban population on fisheries footprint (SCAD).	55
Figure 3.8	Profile of fisheries footprint, GDP per capita and urban population of three representative countries.	56
Figure 4.1	Estimate of non-parametric effect of calcium intake on calcium absorption from proposed variable selection method in NPFCM.	73
Figure 4.2	Hourly casual Bike rentals and weather variables profile in bike sharing data.	74
Figure 4.3	Estimated non-parametric concurrent effect of the weather variables. Displayed are the effects of temperature (top left), humidity (top right) and windspeed (bottom row).	76

Chapter 1

Introduction

Function on function regression is an active area of research in functional data analysis which focuses on regression problems where both the response variable and the covariates are functions over some continuous domain such as time. Functional concurrent regression model (Kim et al., 2018) is a special class of function on function regression models, where the predictors influence the response concurrently, i.e., the value of the response at a specific time point depends on the predictor values at that time point only, and not on the past. Functional linear concurrent regression model (Ramsay and Silverman, 2005) is a further subclass of functional concurrent regression models, where the dependence of the response on the predictors is assumed to be linear and is modeled via smooth univariate regression functions. This is similar to the varying coefficient model (Hastie and Tibshirani, 1993) and has been widely used in functional data analysis. Several applications and extensions of these models can be found in various branches of longitudinal data analysis and also in spatiotemporal studies. The domain of functional regression modeling is diverse with data coming from finance, agriculture, biomedical studies (fMRI, EEG), spectral measurements, etc. Often these data require dynamic modeling of the response curve based on other available predictor curves and functional concurrent regression models are particularly suitable for this purpose. Functional linear concurrent model is the most commonly used function on function regression model due to its interpretability. There is a vast

literature in functional linear concurrent regression model regarding estimation of the regression functions; using kernel-local polynomial smoothing (Wu et al., 1998; Hoover et al., 1998; Fan and Zhang, 1999; Kauermann and Tutz, 1999), polynomial spline (Huang et al., 2002, 2004), smoothing splines (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001; Eubank et al., 2004) among many others. Estimation of the regression functions is important as they can reveal the time varying effect the predictors have on the response.

Extending the functional linear concurrent model (FLCM) is the nonparametric or general functional concurrent model which overcomes the assumption of linear dependence between response and covariates in FLCM; which may not be true in many real data applications, by specifying a nonparametric relationship, which is more general and flexible for modeling purposes. Here also, there exists multiple methods for estimation such as smoothing splines (Kim et al., 2018), Gaussian process regression (Shi et al., 2005), local kernel smoothing techniques (Jiang et al., 2011), etc. For non-Gaussian functional responses, an extension of the nonparametric functional concurrent model is the generalized functional concurrent model of Wang and Shi (2014).

While several methods for estimation and prediction are available in the literature for the functional linear concurrent model (FLCM) or the nonparametric functional concurrent model (NPFCM), methods for testing of global effects of covariates and methods for variable selection are relatively few, specially in the context of function on function regression models. In this dissertation, therefore, the focus has been to develop methods for hypothesis testing and variable selection in the context of functional concurrent regression models.

In Chapter 2, we propose a novel method for testing the null hypothesis of no effect of a covariate on the response in the context of functional linear concurrent regression. Although there have been several methods developed for estimation in the context of functional linear concurrent regression, methods for testing global effects are relatively few. We show that the testing problem can be reduced to testing for zero variance components for a set of random effects. For this purpose, we use a one-sided score test approach, which is an extension of the classical score

test. We establish asymptotic distribution of our test statistic under null and provide theoretical justification as to why our testing procedure has the right levels (asymptotically) under null using standard assumptions. Using numerical simulations, we show that our testing method has the desired type I error rate as well as a higher power than a standard bootstrapped F test used in the literature. Our model and testing procedure are shown to give good performances even when the data is sparsely observed, and the covariate is contaminated with noise. We also demonstrate application of our method on real data.

In Chapter 3, we develop a method for variable selection in functional linear concurrent regression. Our research is motivated by a fisheries footprint study where one of the goals is to identify important time varying socio-structural drivers influencing patterns of seafood consumption and hence fisheries footprint over time. We develop a variable selection method in functional linear concurrent regression extending the classically used scalar on scalar variable selection methods like LASSO, SCAD and MCP. We show in functional linear concurrent regression the variable selection problem can be addressed as a group LASSO, and their natural extension; group SCAD or a group MCP problem. Through simulations, we illustrate our proposed method, particularly with group SCAD or group MCP penalty can pick out the true underlying variables with high accuracy and has minimal false positive and false negative rate. Finally, we apply our method in a study of dietary calcium absorption and fisheries footprint, for selection of influential time varying covariates.

In Chapter 4, we extend our method for performing variable selection in nonparametric functional concurrent regression. The assumption of linearity in the functional linear concurrent regression model is too restrictive in most real world applications. Nonparametric functional concurrent regression model, on the other hand, is much more flexible and can capture complex relationships present between response and covariates. We model the nonparametric functions in terms of bi-variate basis expansion using a tensor product of univariate basis functions and impose group penalty on the basis coefficients to perform variable selection. Numerical analysis using simulations show satisfactory performance of the proposed method in terms of

selection accuracy and out of sample prediction performance. Real data applications on a dietary calcium absorption study and a bike sharing data demonstrate the proposed method can be used successfully for variable selection while simultaneously estimating nonparametric concurrent effect of the predictors which helps to understand underlying dynamics between the response and the predictors.

In Chapter 5, we conclude with a summary of the contributions made in this thesis and discuss some interesting problems and extensions that could be addressed based on research carried out in this thesis.

Chapter 2

A Score Based Test for Functional Linear Concurrent Regression

2.1 Introduction

Functional linear concurrent regression model arises when the response and covariates are both functions of time (or any continuous index), and the value of the response at a particular time point is modeled as a linear combination of the covariates at that specific time point, where the coefficients of the functional covariates are functions of time (Ramsay and Silverman, 2005). One can view the functional linear concurrent regression model as a series of linear regression for each time point, with the assumption that the coefficient functions are smooth over time. Multiple methods exist in literature for estimation of these regression coefficient functions in functional linear concurrent regression and the closely related varying coefficient model (Hastie and Tibshirani, 1993) using basis functions with roughness penalty (Ramsay and Silverman, 2005), polynomial spline (Huang et al., 2002, 2004), local polynomial smoothing (Wu et al., 1998; Cai et al., 2000; Fan and Zhang, 2000), Bayesian modeling (Gelfand et al., 2003), covariance representation techniques (Şentürk and Nguyen, 2011), among many others. While estimation of the regression functions is an important problem, in many cases the primary interest might

be finding out whether a specific covariate is truly significant or not, i.e., to test for association between a predictor of interest and the response. For example, in the gait study data (Ramsay and Silverman, 2005), where there are longitudinal measurements of hip and knee angles taken on 39 children, the main purpose of the study is to understand how the joints in hip and knee interact during a gait cycle (Theologis, 2009). One natural question to ask here would be, whether the knee angles (response) are at all associated with the hip angles (covariate). Simply building a point-wise confidence interval of the estimated regression function does not answer the question of the overall significance of the covariate. Thus there is a need for developing testing methods to find out significant predictors in this setting.

Formally, our primary goal is to test the null hypothesis that the coefficient function corresponding to a predictor of interest is identically zero, versus the alternative hypothesis that the coefficient function is non-zero for some time point. Literature relating to such global testing in functional concurrent linear regression or closely related varying coefficient model can be traced back to Huang et al. (2002), Guo (2002) among many others. Huang et al. (2002) employed a resampling subject bootstrap method on an F-type statistic, whereas Guo (2002) used the connection between linear mixed effects models and smoothing splines, and subsequently used a generalized maximum likelihood ratio test to test for significance of predictors. Kim et al. (2018) extended the bootstrap based test (Huang et al., 2002) to general nonlinear functional concurrent model. Both of these tests rely on a subject-level bootstrap method to obtain the p-values making the test computationally intensive. Recently, Wang et al. (2018) developed a method for pointwise as well as global testing using empirical likelihood ratio tests. Their method, which is extremely general, uses a wild bootstrap procedure to perform the test for both dense and sparse functional data. Besides these global testing procedures one can also use confidence bands based methods e.g., Fan and Zhang (2000) for building simultaneous confidence bands for the underlying coefficient functions.

In this article, our goal is to build a classical likelihood based testing method for testing of the global effect of covariates, which is also computationally cheap. We model the unknown

regression coefficients using B-spline basis functions and derive an equivalent random effects model. Under such a framework, we show that our testing problem reduces to testing for zero variance components for a set of random effects. There are multiple existing methods in the literature for testing for zero variance components. Crainiceanu and Ruppert (2004), Greven et al. (2008), and Staicu et al. (2014) considered testing for variance components using likelihood ratio test (LRT) and restricted likelihood ratio test (RLRT). The main challenge of such tests is that the null distribution is different from the commonly used $0.5\chi_0^2 : 0.5\chi_1^2$ approximation or such mixtures of two chi-square distributions, which is used in Guo (2002). In this article, we propose a score based testing method that is computationally efficient. Our procedure is inspired from the work of Molenberghs and Verbeke (2007) which describes an approach of using a one-sided score test in constrained parameter space. The major advantage of working with the score test is, it does not require computations under the alternative. Zhang and Lin (2008) and Lin (1997) also used such one-sided score tests for variance component testing in generalized linear mixed models and longitudinal data. However, the methods mentioned above assumed that the responses are independent given the random effects and that the variances have some parametric form. In contrast, in our functional regression framework, we assume unknown non-trivial covariance structure and estimate the covariance function nonparametrically. The assumption of non-trivial dependence is crucial in functional data because of complex correlation structures that might be present in real data. We derive the asymptotic distribution of the test statistic under the null hypothesis; we show that the commonly used chi-squared approximation of the score test statistic is not appropriate in our situation. However, the null distribution of our test statistic is easy to simulate from. Thus the calculation of p-value for our testing procedure is computationally efficient. We show that asymptotically our testing procedure has the correct type I error rate. Using numerical simulations, we illustrate that our testing method has the desired type I error rates for finite sample sizes and that our proposed testing procedure has higher power than the bootstrapped F-test of Kim et al. (2018).

The rest of the article is organized as follows. In Section 2.2, we discuss our model specification,

present our testing method and derive theoretical properties related to our test statistic. In Section 2.3, we present a simulation study under various sampling design scenarios and give the simulation results. In Section 2.4, we demonstrate our proposed test by applying it to the two real data examples: gait data and calcium absorption study and summarize our findings. We conclude by a discussion about some limitations and some possible extensions of our work in Section 2.5.

2.2 Methodology

2.2.1 Modeling framework

Suppose that the observed data for the i th subject, $i = 1, \dots, n$, is $\{Y_i(t), X_{i1}(t), \dots, X_{ip}(t)\}$, where $Y(\cdot)$ is a functional response and $X_1(\cdot), \dots, X_p(\cdot)$ are the corresponding functional covariates. In practice, the functions for the i th subject are observed only on a finite set of points t_{ij} , $j = 1, \dots, m_i$. We assume that $t_{ij} \in \mathcal{T}$, a bounded and closed set. For the rest of the article, we assume $\mathcal{T} = [0, 1]$ without loss of generality. To start with, we will assume that $t_{ij} = t_j$, and that the covariates $X_{ik}(\cdot)$ are measured without error. We discuss the cases when the functions are observed on irregularly spaced grid, and with additional measurement errors in Section 2.2.5. We consider a functional linear concurrent regression model,

$$Y_i(t) = \beta_0(t) + \sum_{k=1}^p X_{ik}(t)\beta_k(t) + \epsilon_i(t),$$

where $\beta_0(t)$ and $\beta_k(t)$ ($k = 1, 2, \dots, p$) are smooth functions representing functional intercept and functional slope parameters, respectively. We assume $X_{ik}(\cdot)$ are independent and identically distributed (i.i.d.) copies of $X_k(\cdot)$ ($k = 1, 2, \dots, p$), where $X_k(\cdot)$ is a stochastic processes with finite second moment. For simplicity we illustrate our testing method for the single covariate model

$$Y_i(t) = \beta_0(t) + X_i(t)\beta_1(t) + \epsilon_i(t), \tag{2.1}$$

which can be easily extended to the multiple covariate situation above and this is discussed in Section 2.2.6. We further assume $\epsilon_i(\cdot)$ are i.i.d. copies of $\epsilon(\cdot)$, which is a mean zero Gaussian process plus some Gaussian white noise, that is, $\epsilon(t) = V(t) + w_t$, where $V(\cdot) \sim \mathcal{N}(0, G(\cdot, \cdot))$ and w_t are i.i.d. $\mathcal{N}(0, \sigma^2)$ random errors. Thus the covariance function of the error process is given by $\Sigma(s, t) = \text{cov}(\epsilon(s), \epsilon(t)) = G(s, t) + \sigma^2 I(s = t)$. Our primary interest lies in testing,

$$H_0 : \beta_1(t) = 0 \text{ for all } t \quad \text{versus} \quad H_1 : \beta_1(t) \neq 0 \text{ for some } t.$$

In general, testing H_0 is difficult since $\beta_1(t)$ is an infinite dimensional parameter. In this article we show that by modeling the coefficient function with splines and using a random effects model, the testing problem can be reduced to a variance component test. This reduction in parameter dimension not only helps in getting satisfactory performance of our testing method but also is computationally cheaper than doing the bootstrapped F test existing in literature. Subsequently we develop a one sided score test for testing our null hypothesis.

2.2.2 Equivalent random effects model

An usual method (Ramsay and Silverman, 2005) to estimate $\beta_0(t)$ and $\beta_1(t)$ in model (1) is by minimizing the penalized residual sum of squares,

$$\sum_{i=1}^n \|Y_i(\cdot) - \beta_0(\cdot) - X_i(\cdot)\beta_1(\cdot)\|_{F_2}^2 + \lambda_0 \int \{\beta_0^r(t)\}^2 dt + \lambda_1 \int \{\beta_1^r(t)\}^2 dt,$$

where λ_0, λ_1 are unknown penalty parameters penalizing r -th derivative of the coefficient functions and $\|\cdot\|_{F_2}$ denotes the functional L_2 norm. Suppose for $\ell = 0, 1$ $\{B_{k\ell}(t), k = 1, 2, \dots, k_\ell\}$ is a set of known basis functions. We write the unknown coefficient functions using basis function expansion as $\beta_\ell(t) = \sum_{k=1}^{K_\ell} b_{k\ell} B_{k\ell}(t) = \mathbf{B}_\ell^T(t) \mathbf{b}_\ell$, $\ell = 0, 1$, where $\mathbf{B}_\ell(t) = [B_{1\ell}(t), B_{2\ell}(t), \dots, B_{K_\ell\ell}(t)]^T$ and $\mathbf{b}_\ell = (b_{1\ell}, b_{2\ell}, \dots, b_{k_\ell\ell})^T$ is a vector of unknown coefficients. In this article, we use B-spline basis functions, however, other basis functions can be used as well. Thus we can write $X_i(t)\beta_1(t) = \sum_{k=1}^{k_1} b_{k1} X_i(t) B_{k1}(t) = \mathbf{X}_i^{*T}(t) \mathbf{b}_1$, where $\mathbf{X}_i^*(t) =$

$[X_i(t)B_{11}(t), X_i(t)B_{21}(t), \dots, X_i(t)B_{k_1}(t)]^T$. We can then rewrite our model (1) as $Y_i(t) = \mathbf{B}_0^T(t)\mathbf{b}_0 + \mathbf{X}_i^{*T}(t)\mathbf{b}_1 + \epsilon_i(t)$. The unknown basis coefficients can then be estimated by minimizing the penalized error sum of squares $\sum_{i=1}^n \|Y_i(\cdot) - \mathbf{B}_0^T(\cdot)\mathbf{b}_0 - \mathbf{X}_i^{*T}(\cdot)\mathbf{b}_1\|_2^2 + \lambda_0 \mathbf{b}_0^T \mathbb{P}_0 \mathbf{b}_0 + \lambda_1 \mathbf{b}_1^T \mathbb{P}_1 \mathbf{b}_1$, where \mathbb{P}_0 and \mathbb{P}_1 are the penalty matrices coming from penalizing the r -th derivative of the functions $\beta_0(t)$ and $\beta_1(t)$. In particular $\mathbb{P}_\ell = \int \mathbf{B}_\ell^r(t) \mathbf{B}_\ell^r(t)^T dt$ and thus $\int (\beta_\ell^r(t))^2 dt = \mathbf{b}_\ell^T \mathbb{P}_\ell \mathbf{b}_\ell$. Since we only observe data on a fine regular grid $S = \{t_1, t_2, \dots, t_m\}$ in practice, the minimization is carried out by minimizing

$$\sum_{i=1}^n \sum_{j=1}^m \{Y_i(t_j) - \mathbf{B}_0^T(t_j)\mathbf{b}_0 - \mathbf{X}_i^{*T}(t_j)\mathbf{b}_1\}^2 + \lambda_0 \mathbf{b}_0^T \mathbb{P}_0 \mathbf{b}_0 + \lambda_1 \mathbf{b}_1^T \mathbb{P}_1 \mathbf{b}_1.$$

Define $\mathbf{Y}_i = [Y_i(t_1), Y_i(t_2), \dots, Y_i(t_m)]^T$, $\mathbb{B}_0 = [\mathbf{B}_0(t_1) | \mathbf{B}_0(t_2) | \dots | \mathbf{B}_0(t_m)]^T$, $\mathbb{X}_i = [\mathbf{X}_i^*(t_1) | \mathbf{X}_i^*(t_2) | \dots | \mathbf{X}_i^*(t_m)]^T$. Thus the least square criterion for estimation is given by,

$$\sum_{i=1}^n \|\mathbf{Y}_i - \mathbb{B}_0 \mathbf{b}_0 - \mathbb{X}_i \mathbf{b}_1\|_2^2 + \lambda_0 \mathbf{b}_0^T \mathbb{P}_0 \mathbf{b}_0 + \lambda_1 \mathbf{b}_1^T \mathbb{P}_1 \mathbf{b}_1.$$

Now since the matrices $\mathbb{P}_0, \mathbb{P}_1$ are singular (for $r \geq 1$), the equivalent random effects model corresponding to this minimization problem would be rank deficient. As our primary interest lies in testing, we propose to penalize the coefficient functions directly, namely we use $r = 0$ and consequently $\mathbb{P}_\ell = \int \mathbf{B}_\ell(t) \mathbf{B}_\ell(t)^T dt$. Our simulations show we are able to maintain correct type I error rates of the proposed testing method using this strategy. It then follows that the normal equations are identical to those from the equivalent random effects model $\mathbf{Y}_i = \mathbb{B}_0 \mathbf{b}_0 + \mathbb{X}_i \mathbf{b}_1 + \zeta_i$, where $\zeta_i \sim \mathcal{N}_m(0, \sigma_\epsilon^2 \mathbb{I}_m)$, $\mathbf{b}_0 \sim \mathcal{N}_{k_0}(0, \sigma_0^2 \Sigma_0)$, $\mathbf{b}_1 \sim \mathcal{N}_{k_1}(0, \sigma_1^2 \Sigma_1)$ and $\mathbf{b}_0, \mathbf{b}_1, \zeta_i$ are independent. We denote $\Sigma_0 = \mathbb{P}_0^{-1}$, $\Sigma_1 = \mathbb{P}_1^{-1}$, $\sigma_0^2 = \frac{\sigma_\epsilon^2}{\lambda_0}$, $\sigma_1^2 = \frac{\sigma_\epsilon^2}{\lambda_1}$. Using Cholesky decomposition of Σ_0, Σ_1 and appropriately reparameterizing ($\mathbb{C}_0 = \mathbb{B}_0 \Sigma_0^{1/2}$, $\mathbb{Z}_i = \mathbb{X}_i \Sigma_1^{1/2}$, $\gamma_k = \Sigma_k^{-1/2} \mathbf{b}_k$), the model can be rewritten as $\mathbf{Y}_i = \mathbb{C}_0 \gamma_0 + \mathbb{Z}_i \gamma_1 + \zeta_i$, where $\zeta_i \sim \mathcal{N}_m(0, \sigma_\epsilon^2 \mathbb{I}_m)$, $\gamma_0 \sim \mathcal{N}_{k_0}(0, \sigma_0^2 \mathbb{I}_{k_0})$, $\gamma_1 \sim \mathcal{N}_{k_1}(0, \sigma_1^2 \mathbb{I}_{k_1})$, and all the random effects are independent. Thus our test H_0 can be carried out via testing of a single variance component, namely testing $H_0 : \sigma_1^2 = 0$ against the alternative $H_1 : \sigma_1^2 > 0$.

Now for our testing problem since the errors are not independent, to get the correct likelihood we need to use the true covariance kernel $\Sigma(s, t)$ for the residual vector ϵ_i s, which motivates us to use the random effects model

$$\mathbf{Y}_i = \mathbf{C}_0\gamma_0 + \mathbf{Z}_i\gamma_1 + \epsilon_i, \quad (2.2)$$

where $\epsilon_i \sim \mathcal{N}_m(0, \Sigma_{m \times m})$, $\gamma_0 \sim \mathcal{N}_{k_0}(0, \sigma_0^2 \mathbb{I}_{k_0})$, $\gamma_1 \sim \mathcal{N}_{k_1}(0, \sigma_1^2 \mathbb{I}_{k_1})$ and all of them are independent. Here $\Sigma_{m \times m}$ denotes the covariance kernel $\Sigma(s, t)$ evaluated at $S = \{t_1, t_2, \dots, t_m\}$. For the moment let us assume $\Sigma_{m \times m}$ to be known. Of course in reality $\Sigma_{m \times m}$ will be unknown and we will need to estimate it. We illustrate in Section 2.2.4 how to estimate $\Sigma_{m \times m}$ using functional principal component analysis (FPCA). Writing equation (2.2) in stacked form for $i = 1, 2, 3, \dots, n$ we have

$$\mathbf{Y} = \mathbb{B}\gamma_0 + \mathbb{Z}\gamma_1 + \mathcal{E}, \quad (2.3)$$

where $\mathbb{B} = [\mathbf{C}_0^T | \mathbf{C}_0^T | \dots | \mathbf{C}_0^T]^T$, $\mathbb{Z} = [\mathbf{Z}_1^T | \mathbf{Z}_2^T | \dots | \mathbf{Z}_n^T]^T$, $\mathbf{Y} = (\mathbf{Y}_1^T, \mathbf{Y}_2^T, \dots, \mathbf{Y}_n^T)^T$ and $\mathcal{E} = (\epsilon_1^T, \epsilon_2^T, \dots, \epsilon_n^T)^T$. $\mathcal{E} \sim \mathcal{N}_N(0, \Sigma)$ ($N = mn$), where the covariance matrix Σ is given by $\Sigma = \text{diag}\{\Sigma_{m \times m}, \Sigma_{m \times m}, \dots, \Sigma_{m \times m}\}$. Note that $\mathbb{Z} = \mathbb{X}\Sigma_1^{1/2}$ and $\mathbb{B} = \mathbb{B}_0\Sigma_0^{1/2}$, where \mathbb{X}, \mathbb{B}_0 are defined similarly by stacking \mathbb{X}_i and \mathbb{B}_0 s. So in this set up, we are interested in testing $H_0 : \sigma_1^2 = 0$ against the alternative $H_1 : \sigma_1^2 > 0$.

2.2.3 Testing method

We develop our testing method treating \mathbb{Z} as nonrandom (fixed). Namely our test is a conditional test based on observed \mathbb{Z} {i.e., observed $X_i(t)$ }. We show that our conditional testing method has the right levels under null which in turn ensures the unconditional test would also enjoy this property. Marginally $\mathbf{Y} \sim \mathcal{N}(0, \mathbb{V})$, which follows from equation (2.3) with $\mathbb{V} = \mathbb{V}(\tau_0, \tau_1) = \Sigma + \tau_0\mathbb{B}\mathbb{B}^T + \tau_1\mathbb{Z}\mathbb{Z}^T$, where $\tau_0 = \sigma_0^2$ and $\tau_1 = \sigma_1^2$. So the marginal log-likelihood of \mathbf{Y} (upto a constant) is :

$$L_{ML}(\tau_0, \tau_1) = -1/2(\ln |\mathbb{V}| + \mathbf{Y}^T\mathbb{V}^{-1}\mathbf{Y}). \quad (2.4)$$

Based on this likelihood, we want to test $H_0 : \tau_1 = 0$ vs $H_1 : \tau_1 > 0$.

Let $\boldsymbol{\theta} = (\tau_0, \tau_1)^T$ and $\tilde{\boldsymbol{\theta}}$ denote the maximum likelihood estimate (M.L.E) of $\boldsymbol{\theta}$ under H_0 .

The score function of τ_1 is

$$\begin{aligned} S_{\tau_1}(\tau_0, \tau_1) &= -1/2\{tr(\mathbb{V}^{-1}\mathbb{M}) - \mathbf{Y}^T\mathbb{V}^{-1}\mathbb{M}\mathbb{V}^{-1}\mathbf{Y}\} \\ &= -1/2\{tr(\mathbb{Z}^T\mathbb{V}^{-1}\mathbb{Z}) - (\mathbb{V}^{-1/2}\mathbf{Y})^T\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}(\mathbb{V}^{-1/2}\mathbf{Y})\}, \end{aligned}$$

where $\mathbb{M} = \mathbb{Z}\mathbb{Z}^T$. The information matrix $\mathbb{I}(\boldsymbol{\theta})$ corresponding to the likelihood in (2.4) is partitioned as,

$$\mathbb{I}(\boldsymbol{\theta}) = \begin{Bmatrix} I_{11}(\boldsymbol{\theta}) & I_{12}(\boldsymbol{\theta}) \\ I_{21}(\boldsymbol{\theta})^T & I_{22}(\boldsymbol{\theta}) \end{Bmatrix},$$

where $I_{11}(\boldsymbol{\theta}) = tr\{(\mathbb{B}^T\mathbb{V}^{-1}\mathbb{B})^2\}/2$, $I_{22}(\boldsymbol{\theta}) = tr\{(\mathbb{Z}^T\mathbb{V}^{-1}\mathbb{Z})^2\}/2$, $I_{21}(\boldsymbol{\theta}) = I_{12}(\boldsymbol{\theta}) = tr\{(\mathbb{B}^T\mathbb{V}^{-1}\mathbb{Z})(\mathbb{B}^T\mathbb{V}^{-1}\mathbb{Z})^T\}/2$. Then the classical score test statistic is given by,

$$T_S = \frac{S_{\tau_1}^2(\tilde{\boldsymbol{\theta}})}{I_{22}(\tilde{\boldsymbol{\theta}}) - I_{21}(\tilde{\boldsymbol{\theta}})^T I_{11}^{-1}(\tilde{\boldsymbol{\theta}}) I_{12}(\tilde{\boldsymbol{\theta}})}.$$

As the parameter space is constrained and the null hypothesis is on the boundary of the parameter space, following Molenberghs and Verbeke (2007) we define our one sided score test statistic as

$$T_S = \begin{cases} \frac{S_{\tau_1}^2(\tilde{\boldsymbol{\theta}})}{\Lambda(\tilde{\boldsymbol{\theta}})} & \text{if } S_{\tau_1}(\tilde{\boldsymbol{\theta}}) \geq 0 \\ 0 & \text{if } S_{\tau_1}(\tilde{\boldsymbol{\theta}}) < 0, \end{cases} \quad (2.5)$$

where $\Lambda(\tilde{\boldsymbol{\theta}}) = I_{22}(\tilde{\boldsymbol{\theta}}) - I_{21}(\tilde{\boldsymbol{\theta}})^T I_{11}^{-1}(\tilde{\boldsymbol{\theta}}) I_{12}(\tilde{\boldsymbol{\theta}})$. We assume the true covariance matrix Σ to be known for the time being for establishing asymptotic distribution of our test statistic but in reality it is generally unknown, so we will need to estimate it from data by some consistent estimator $\hat{\Sigma}$ and plug that in for Σ in T_S . Next we posit two theorems regarding distribution our test statistic.

Theorem 1. *Suppose the following conditions are true :*

a) The null hypothesis is true i.e., $H_0 : \tau_1 = 0$ holds and $\boldsymbol{\theta}_0 = (\tau_0^*, 0)$ is the true value of $\boldsymbol{\theta}$,

b) Σ be the true covariance matrix of the residual vector \mathcal{E} in equation (2.3).

$$\begin{aligned} \text{Then } T_S(\boldsymbol{\theta}_0, \Sigma) &= \frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \stackrel{d}{=} (1/2)^2 \frac{\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell\right)^2}{\Lambda_n(\boldsymbol{\theta}_0)} I\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell\right), \\ \text{where } x_\ell &\stackrel{iid}{\sim} \mathcal{N}(0, 1) \text{ and } \lambda_\ell \text{ are eigenvalues of } \mathbb{Z}^T \mathbb{V}(\boldsymbol{\theta}_0, \Sigma)^{-1} \mathbb{Z}/n \text{ and } \Lambda_n(\boldsymbol{\theta}_0) = \Lambda(\boldsymbol{\theta}_0)/n^2 \\ &= \frac{1}{2} \text{tr}\{(\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}/n)^2\} - \frac{\left[\frac{1}{2} \text{tr}\{(\mathbb{B}^T \mathbb{V}^{-1} \mathbb{Z}/n)(\mathbb{B}^T \mathbb{V}^{-1} \mathbb{Z}/n)^T\}\right]^2}{\frac{1}{2} \text{tr}\{(\mathbb{B}^T \mathbb{V}^{-1} \mathbb{B}/n)^2\}} \end{aligned}$$

The proof of the Theorem 1 is given in Appendix A.1.

Theorem 2. Suppose the following conditions are true :

a) The null hypothesis is true i.e $H_0 : \tau_1 = 0$ holds and $\boldsymbol{\theta}_0 = (\tau_0^*, 0)$ is the true value of $\boldsymbol{\theta}$,

b) $\tilde{\boldsymbol{\theta}}$ is \sqrt{n} consistent estimator of $\boldsymbol{\theta}_0$ under null and the estimator $\hat{\Sigma}$ is a consistent estimator of Σ in the sense $\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_2 = o_p(1)$ (spectral norm).

$$\text{Then } T_S(\tilde{\boldsymbol{\theta}}, \hat{\Sigma}) \xrightarrow{d} T_S(\boldsymbol{\theta}_0, \Sigma).$$

The proof is mainly based on application of Slutsky's theorem and matrix norm inequalities. A detailed proof is given in Appendix A.2. As mentioned in Theorem 1 the null distribution of the test statistic is given by $(1/2)^2 \frac{\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell\right)^2}{\Lambda_n(\boldsymbol{\theta}_0)} I\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell\right)$. Because $\boldsymbol{\theta}_0$ and λ_ℓ are unknown in reality, we approximate the null distribution using plug-in estimates of $\tilde{\boldsymbol{\theta}}$ and $\hat{\Sigma}$, i.e, we use the approximate null distribution

$$(1/2)^2 \frac{\left(\sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell\right)^2}{\Lambda_n(\tilde{\boldsymbol{\theta}})} I\left(\sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell\right), \quad (2.6)$$

where $\tilde{\lambda}_\ell$ are eigenvalues of $\mathbb{Z}^T \mathbb{V}(\tilde{\boldsymbol{\theta}}, \hat{\Sigma})^{-1} \mathbb{Z}/n$ and $x_\ell \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ for $\ell = 1, 2, \dots, k_1$. This is justified as it can be shown $\tilde{\lambda}_\ell \xrightarrow{p} \lambda_\ell$ and $\Lambda_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \Lambda_n(\boldsymbol{\theta}_0)$, see the proof in Appendix A.2. Our simulations show that we are able to get the correct type I error rates and good power of our test using the above strategy. Simulation from the null distribution of our test statistic is easy and computationally efficient, as we only need to calculate k_1 eigenvalues of the matrix

$\mathbb{Z}^T \mathbb{V}(\tilde{\boldsymbol{\theta}}, \hat{\Sigma})^{-1} \mathbb{Z}/n$, and simulate $x_\ell \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ for $\ell = 1, 2, \dots, k_1$. For calculation of $\mathbb{V}(\tilde{\boldsymbol{\theta}}, \hat{\Sigma})^{-1}$, we use the Woodbury matrix identity and also the fact $\hat{\Sigma}$ is block diagonal, which greatly speeds up the calculation. It is well known (Zhang and Lin, 2003), that the usual asymptotic χ^2 distribution of the score test do not work here, so the approximate null distribution in (6) is more appropriate. As the test statistic is one sided and in particular not continuous at zero, the p-value under null is asymptotically distributed as mixture distribution of degenerate one and $U(0, \alpha)$, where $\alpha = P_{H_0}\{S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\} = P(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell)$, a detailed proof is given in Appendix A.3. As k_1 increases it follows by application of CLT, $\alpha \rightarrow \frac{1}{2}$ and the null distribution of our test statistic asymptotically converges to $0.5\chi_0^2 : 0.5\chi_1^2$. In reality choice of k_1 will depend on the type of design (dense or sparse) and number of observed time points for each subjects. So such a convergence need not hold for subjects observed only on a finite set of points. Therefore, we use the approximate null distribution (2.6) to perform our test.

2.2.4 Estimation of Covariance Matrix

In reality Σ is unknown, and we need a consistent estimator $\hat{\Sigma}$. In the context of functional data, we want to estimate $\Sigma(\cdot, \cdot)$ completely non-parametrically. If we had the original residuals ϵ_{ij} available, we could use functional principal component analysis (FPCA), e.g., Yao et al. (2005) or Zhang et al. (2007) to estimate $\Sigma(s, t)$. The error process $\epsilon(t)$ was defined as $\epsilon(t) = V(t) + w_t$. We assume the covariance kernel $G(s, t)$ of the smooth part $V(t)$ is a Mercer kernel (Mercer, 1909). Then by Mercer's theorem $G(s, t)$ must have a spectral decomposition

$$G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the ordered eigenvalues and $\phi_k(\cdot)$ s are corresponding eigenfunctions. Thus we have $\Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t) + \sigma^2 I(s = t)$. Given $\epsilon_{t_{ij}} = V(t_{ij}) + w_{ij}$, one could employ FPCA based methods to get $\hat{\phi}_k(\cdot)$, $\hat{\lambda}_k$ s and $\hat{\sigma}^2$. Hall et al. (2006) established L^2 convergence of the FPCA estimates, in particular of the eigenfunctions and eigenvalues, in both

the sparse and dense functional data setting under appropriate regularity conditions on the sampling design. Li and Hsing (2010) established uniform convergence rates for eigenfunctions and eigenvalues under more general framework where the number of observations for each function can be sampled at any rate relative to the sample size. More specifically Li and Hsing (2010) showed it is possible to get consistent estimators $\hat{\phi}_k(\cdot)$, $\hat{\lambda}_k$ and $\hat{\sigma}^2$ under both sparse and dense functional data settings. So a consistent estimator of $\Sigma(s, t)$ can be formed as

$$\hat{\Sigma}(s, t) = \sum_{k=1}^K \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t) + \hat{\sigma}^2 I(s = t),$$

where K is large enough for the convergence to hold and is typically chosen such that percent of variance explained (PVE) by the selected eigencomponents exceeds some pre-specified value such as 99% or 95%. In reality we don't have the original residuals ϵ_{ij} and use the full model (1) to obtain residuals $e_{ij} = Y_i(t_j) - \hat{Y}_i(t_j)$. Then treating e_{ij} as our original residuals, we obtain $\hat{\Sigma}(s, t)$ using FPCA. Our simulations show good results using this approach and we are able to maintain the correct levels of the test under both sparse and dense sampling design scenarios.

2.2.5 Extension to Sparse and Noisy Covariate

In developing our method we assumed that covariates are measured without noise and data is observed on a regular dense grid of points $S = \{t_1, t_2, \dots, t_m\} \subset \mathcal{T} = [0, 1]$. Although in reality data might be observed sparsely, and the covariates may be contaminated with measurement error. Our testing method can be extended to these situations in the following ways.

Case 1: Sparse design, no measurement error

We assume response $Y_i(t)$ and the covariate $X_i(t)$ are observed in $S_i = \{t_{i1}, t_{i2}, \dots, t_{im_i}\} \subset \mathcal{T} = [0, 1]$ for each $i = 1, 2, \dots, n$ and $\max_{1 \leq i \leq n} m_i \leq M$, for some fixed M . In this case the only difference in our model is that \mathbf{Y}_i is a $m_i \times 1$ dimensional vector and that $\boldsymbol{\epsilon}_i \sim \mathcal{N}_{m_i}(0, \Sigma_i)$ independently, where Σ_i is the covariance kernel $\Sigma(s, t)$ evaluated at $S_i = \{t_{i1}, t_{i2}, \dots, t_{im_i}\}$. So

our model described in (2.2) still holds with $\Sigma = \text{diag} \{ \Sigma_1, \Sigma_2, \dots, \Sigma_n \}$. As discussed in Section 2.2.4, if $\Sigma(s, t)$ can be consistently estimated by $\hat{\Sigma}(s, t)$ then $\hat{\Sigma}_i$ would be a consistent estimator of Σ_i and our testing method is still valid.

Case 2: Dense design with measurement error

Suppose now data $Y_i(t)$ and covariate $X_i(t)$ are observed in a fine regular grid $S = \{t_1, t_2, \dots, t_m\} \subset \mathcal{T} = [0, 1]$ and the covariate is observed with measurement error. Namely instead of observing $X_i(t_j)$ we observe $U_{ij} = X_i(t_j) + \delta_{ij}$, where δ_{ij} are i.i.d. mean zero random errors with variance κ^2 . There exists several methods to reconstruct the original curve $X_i(\cdot)$ from the observed curve with measurement error. Zhang et al. (2007) proposed to use local polynomial kernel smoothing technique for individual function reconstructions. They showed that under appropriate conditions and suitable choice of bandwidth, the smoothed trajectories, $\hat{X}_i(\cdot)$ will estimate the true curves $X_i(\cdot)$ with negligible error. Thus we can use the reconstructed curves $\hat{X}_i(\cdot)$ and the effect of such a substitution is asymptotically negligible.

Case 3 : Sparse design with measurement error

More generally, we consider the case where functional data is observed in irregular and sparse grid of points and covariate is observed with measurement error. Here we have observed response $\{(Y_i(t_{ij}), t_{ij}), j = 1, 2, \dots, m_i\}$ and observed covariate $\{(U(t_{ij}), t_{ij}), j = 1, 2, \dots, m_{1i}\}$. Again we assume that $U_{ij} = X_i(t_{ij}) + \delta_{ij}$, where δ_{ij} are i.i.d. mean zero random errors with variance κ^2 . In such a sparse sampling design, it is generally assumed (Kim et al., 2018) although the individual number of observations m_i is small, $\bigcup_{i=1}^n \bigcup_{j=1}^{m_i} t_{ij}$ is dense in $\mathcal{T} = [0, 1]$. Then reconstructing of the original curves from the observed sparse curves can be done by using FPCA methods of Yao et al. (2005) or Hall et al. (2006). The methods are based on estimating the mean and covariance functions using local linear smoothing, and subsequently estimating the eigenvalues and eigenfunctions from a spectral decomposition of estimated covariance matrix. As mentioned earlier, Li and Hsing (2010) proved uniform convergence of the mean, eigenvalues

and eigenfunctions for both dense and sparse design under suitable regularity conditions. For prediction of the scores Yao et al. (2005) introduced the PACE method which ensures the estimated scores asymptotically goes to BLUP of the original scores. Then these estimates can be put together using Karhunen-Loève expansion to get estimates $\hat{X}_i(\cdot)$ of the true curve $X_i(\cdot)$. We again use the PVE criterion to select the number of PC. So for sparse data observed on irregular grid and observed with measurement error, we employ FPCA to get $\hat{X}_i(t)$ and then use $\{Y_i(t_{ij}), \hat{X}_i(t_{ij}), j = 1, 2, \dots, m_i\}_{i=1}^n$ as our original data to perform our proposed test. Our simulations show that we are able to maintain correct type I error rate and obtain satisfactory power of our testing method using this strategy.

2.2.6 Extension to multiple covariates

Here, we illustrate how our testing method can also be applied in multiple covariate setting. In this case the likelihood is as in (2.4) and given by

$$L_{ML}(\tau_0, \tau_1, \dots, \tau_p) = -1/2(\ln |\mathbb{V}| + \mathbf{Y}^T \mathbb{V}^{-1} \mathbf{Y}), \quad (2.7)$$

with the only difference being $\mathbb{V} = \mathbb{V}(\tau_0, \tau_1, \dots, \tau_p) = \Sigma + \tau_0 \mathbb{B} \mathbb{B}^T + \tau_1 \mathbf{Z}_1 \mathbf{Z}_1^T + \dots + \tau_p \mathbf{Z}_p \mathbf{Z}_p^T$. So testing for the effect of $X_p(\cdot)$ here similarly reduces to testing for the variance component τ_p , namely, we test $H_0 : \tau_p = 0$ vs $H_1 : \tau_p > 0$. The score function $S_{\tau_p}(\tau_0, \tau_1, \dots, \tau_p)$ and the information matrix $\mathbb{I}(\boldsymbol{\theta})$ $\{\boldsymbol{\theta} = (\tau_0, \tau_1, \dots, \tau_p)^T\}$ are also modified accordingly taking into account the additional variance components, e.g., the information matrix $\mathbb{I}(\boldsymbol{\theta})$ now has to be partitioned as

$$\mathbb{I}(\boldsymbol{\theta}) = \begin{Bmatrix} \mathbb{I}_{11}(\boldsymbol{\theta})_{p \times p} & \mathbf{I}_{12}(\boldsymbol{\theta})_{p \times 1} \\ \mathbf{I}_{21}(\boldsymbol{\theta})_{1 \times p}^T & I_{22}(\boldsymbol{\theta})_{1 \times 1} \end{Bmatrix}.$$

Then our testing method can be applied as in Section 2.2.3 and the generalization of Theorem (1) and Theorem (2) to this multivariate setting remains valid. We have given examples of application of our testing method in multivariate setting, both in our simulation study in

Section 2.3 and real data application in Section 2.4, where we have considered two covariate scenarios and tested for effect of one covariate in presence of other.

2.3 Simulation Study

2.3.1 Study design

In this section, we investigate the performance of our testing method via simulation study. We evaluate our test in terms of type I error rate and power. We also compare our method to the existing bootstrapped F-test proposed by Kim et al. (2018). To this end, we consider the following two scenarios.

Scenario A (single covariate): We generate data from the model,

$$Y_i(t) = \beta_0(t) + X_i(t)\beta_1(t) + \epsilon_i(t),$$

where $\beta_0(t) = 1 + 2t + t^2$ and $\beta_1(t) = t/8$. The original covariate $X_i(\cdot)$ are i.i.d. copies of $X(\cdot)$, where $X(t) = a + b\sqrt{2}\sin(\pi t) + c\sqrt{2}\cos(\pi t)$, where $a \sim \mathcal{N}(0, 1)$, $b \sim \mathcal{N}(0, .85^2)$ and $c \sim \mathcal{N}(0, .70^2)$ and they are independent. As discussed in Section 2.2, we assume that we observe $X_i(t)$ with measurement error, i.e., we observe $U_i(t) = X_i(t) + \delta$, where $\delta \sim \mathcal{N}(0, .6^2)$. The error process $\epsilon_i(t)$ is generated as

$$\epsilon_i(t) = \xi_{i1}\sqrt{2}\cos(\pi t) + \xi_{i2}\sqrt{2}\sin(\pi t) + N(0, 0.9^2 I_{m_i}),$$

where $\xi_{i1} \stackrel{iid}{\sim} \mathcal{N}(0, 2)$ and $\xi_{i2} \stackrel{iid}{\sim} \mathcal{N}(0, 0.75^2)$. We consider the following sampling designs:

- Dense design: Functional data are observed in S for each subject, where S is the set of $m = 81$ equidistant time points in $\mathcal{T} = [0, 1]$.
- Sparse design: The response $Y_i(t)$ and noisy covariate $U_i(t)$ both are observed in random m_{Y_i} and m_{U_i} points in S where $m_{Y_i} \stackrel{iid}{\sim} \text{Uniform}\{20, 21, \dots, 31\}$ and also $m_{U_i} \stackrel{iid}{\sim}$

Uniform{20, 21, . . . , 31}.

We consider two sample sizes, $n = 100$ and 300 for comparison with the bootstrapped F test method. An additional simulation is done in this same set up for dense data with $m = 20$, $n = 39$ to illustrate the performance of the testing method for small sample size, as considered in the application section of this article.

Scenario B (Multiple Covariates):

We generate data from the model,

$$Y_i(t) = \beta_0(t) + X_{i1}(t)\beta_1(t) + X_{i2}(t)\beta_2(t) + \epsilon_i(t),$$

where $\beta_0(t) = 1 + 2t + t^2$, $\beta_1(t) = t/8$ and $\beta_2(t) = \sin(\pi t)$. The original covariates $X_{ik}(\cdot)$ are i.i.d. copies of $X_k(\cdot)$, where $X_k(t) = a_k + b_k \sqrt{2}\sin(\pi t) + c_k \sqrt{2}\cos(\pi t)$, where $a_k \sim \mathcal{N}(0, (2^{-.5(k-1)})^2)$, $b_k \sim \mathcal{N}(0, (0.85 \times 2^{-.5(k-1)})^2)$, and $c_k \sim \mathcal{N}(0, (0.70 \times 2^{-.5(k-1)})^2)$, and they are independent. The covariates $X_k(\cdot)$ are same as considered in Kim et al. (2018). We observe $X_{ik}(t)$ with measurement error, i.e., we observe $U_{ik}(t) = X_{ik}(t) + \delta_k$, where $\delta_k \sim \mathcal{N}(0, .6^2)$. The error process $\epsilon_i(t)$ are generated as in Scenario A described above. Similar sparse and dense design settings and sample sizes $n \in \{100, 300\}$ are considered. Here the main goal is to test $H_0 : \beta_2(t) = 0$ against the alternative $H_1 : \beta_2(t) \neq 0$. For each of the scenarios we use 1000 generated data sets to assess type I error and power. To model the regression functions, 12 cubic B-splines are used for all scenarios.

2.3.2 Simulation Results

Scenario A:

We first assess type I error of the test. We use nominal a level of $\alpha = 5\%$ for $n = 100$ and $n = 300$. The results are displayed in Table 2.1. The estimated standard errors are also given. We observe that our test maintains nominal type I error. For dense design, the test appears to be slightly conservative with estimated type I error below the nominal level of .05. This might

Table 2.1 Estimated type I error along with their standard error at $\alpha = 5\%$ level.

Simulation Scenario	Scenario A		Scenario B	
	$n = 100$	$n = 300$	$n = 100$	$n = 300$
Dense	.035 (.006)	.033 (.006)	.036 (.006)	.035 (.006)
Sparse	.066 (.008)	.048 (.007)	.059 (.007)	.051 (.007)

be due to using a small number of basis functions (12) for dense design ($m = 81$). For sparse design, as sample size increases, the size performance of our test improves, which is expected as our test is a large sample one and for the asymptotic convergence to the null distribution to hold we need larger sample size. The nominal level lies within two standard error limit of estimated type I error in this case.

Next, we study the power performance of our test for a fixed nominal level of $\alpha = 5\%$. To this end, we generate data in the simulation set up mentioned earlier with $\beta_1(t) = dt/8$. Then $d = 0$ corresponds to the null hypothesis and $d > 0$ captures the departure from the null hypothesis. We compare the power of our test to the bootstrapped-F test for $n \in \{100, 300\}$, and both dense and sparse sampling designs. For comparison of power with the bootstrapped method, we use the results from the simulation study conducted by Kim et al. (2018). The results of our study are displayed in Figure 2.1.

We note that across the sparse and dense design scenarios, our method produces higher power than the bootstrapped-F test method. This is expected as our method is a likelihood based method. We also observe that as the sample size increases or d increases the power of our test converges to one across all the settings faster than the bootstrapped-F test. Similar results for the additional simulation study is provided in Figure 2.2.

Scenario B:

The type I error rates for this scenario are also displayed in Table 2.1. Nominal levels of $\alpha = 5\%$

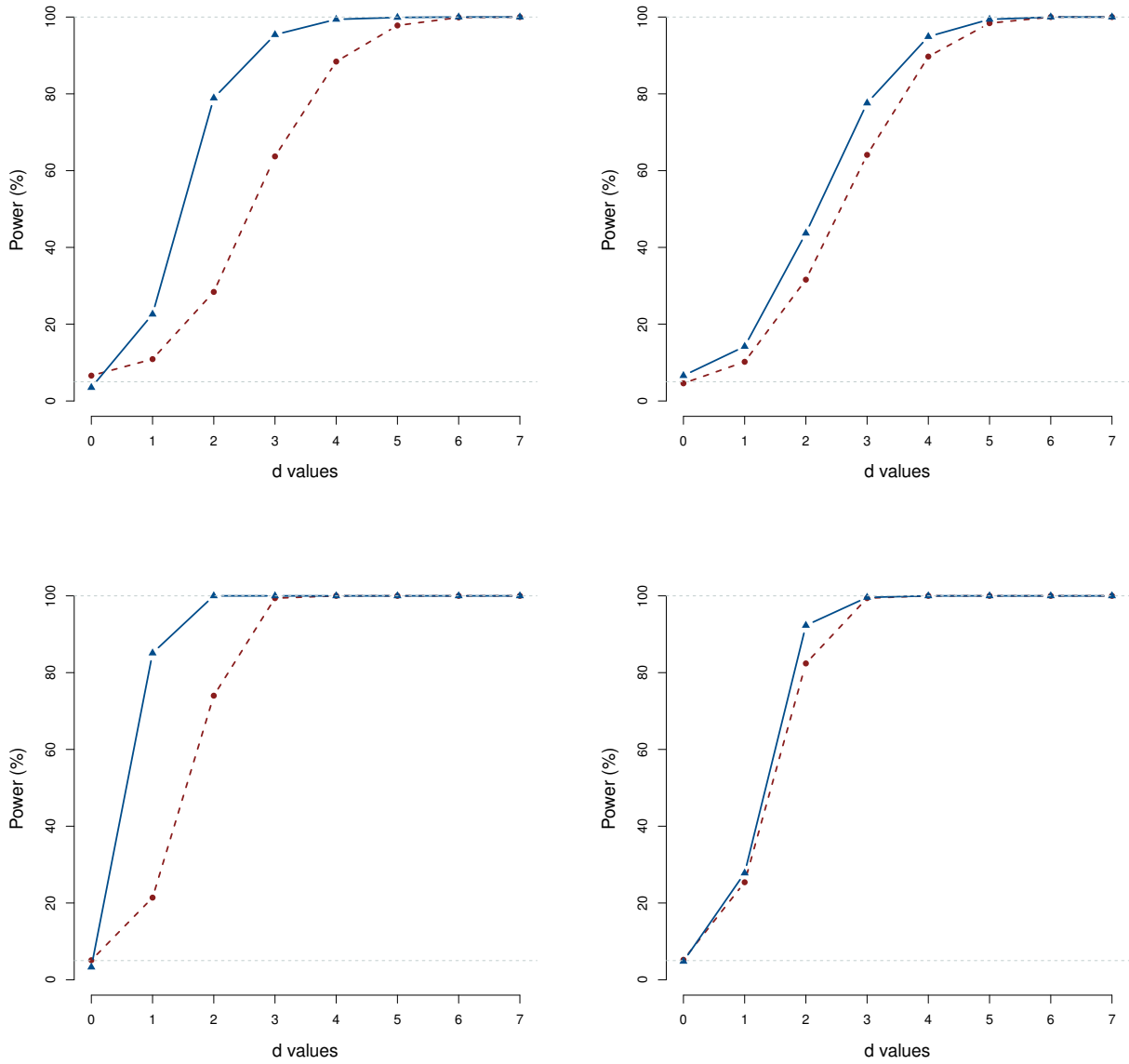


Figure 2.1 Results of simulation study as described in Section 2.3, Scenario A. Displayed are the power curves for our proposed procedure (solid line) and the bootstrap-F test based method (dashed line) for dense (left column) and sparse (right column) sampling designs with sample sizes $n = 100$ (top row) and $n = 300$ (bottom row).

for $n = 100$ and $n = 300$ are considered. Again we observe our test maintaining nominal type I error, and improvement in size performance with increasing sample size, particularly for sparse design. A similar evaluation of power performance is done as in Scenario A. Namely we generate data as in simulation **Scenario B** with $\beta_2^*(t) = d\beta_2(t)$. The power curve of our testing method for both dense and sparse settings for $n \in \{100, 300\}$ are displayed in Figure 2.3. We observe the power converging to one, as sample size or d increases, across all the sampling designs.

Our simulation results illustrate the proposed testing method is able to maintain the nominal type I error rate as well as capture the departure from null hypothesis efficiently even when data is observed sparsely and the covariates are observed with measurement error.

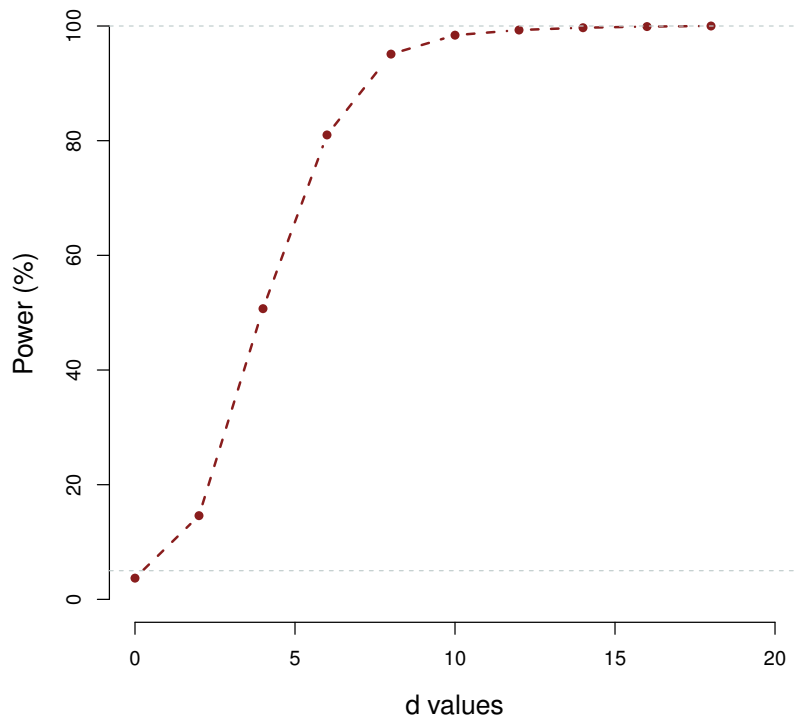


Figure 2.2 Power curve for simulation Scenario A, additional simulation study, $n=39$, $m=20$.

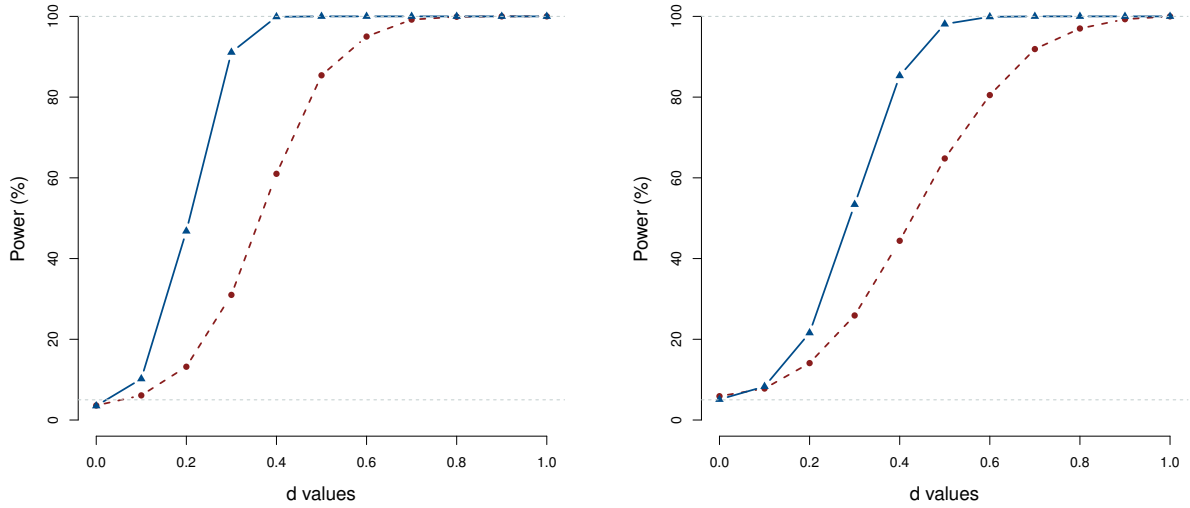


Figure 2.3 Results of simulation study as described in Section 2.3, Scenario B. Displayed are the power curves for our proposed procedure with sample sizes $n = 100$ (dashed line) and $n = 300$ (solid line) for dense (left column) and sparse (right column) sampling designs.

We have used 12 cubic B-splines basis to model the regression functions in all our simulations. In reality, the choice of the number of basis functions depends on the type of design (dense or sparse) and number of observed time points for each subjects. Using a large number of basis can result in loss of power while using a small number of basis can make the test conservative as we saw in case of dense design. To investigate the effect of number of basis on the performance of the test an additional simulation is carried out as in Scenario A for dense design, with number of basis functions $k_1 = 12, 14, 16$. The result is illustrated in Figure 2.4. The estimated type I error achieves the nominal level with an increase in the number of basis. We also observe a marginal decrease in the power as expected with a larger number of basis, which might be attributed to more parameters in the model.

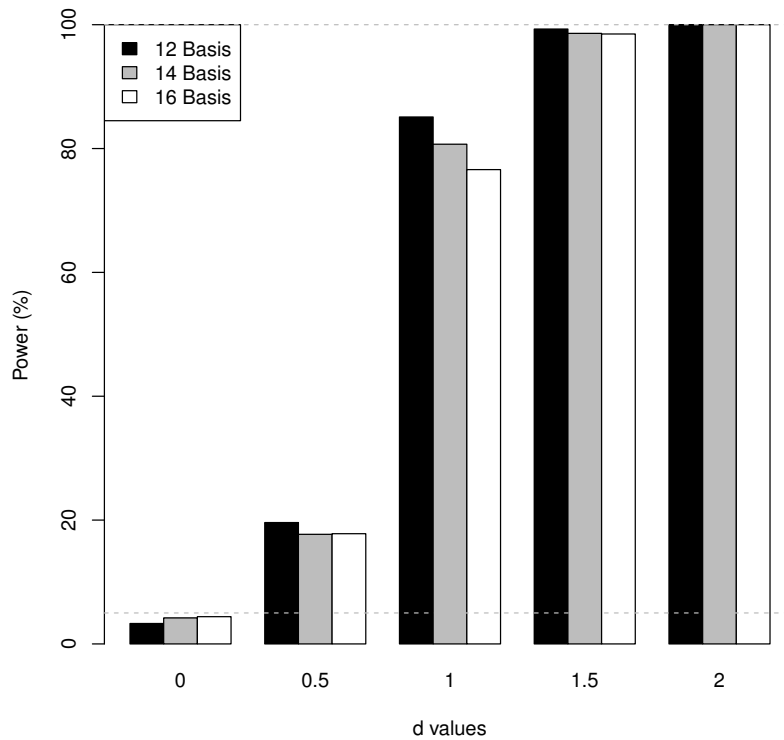


Figure 2.4 Effect of number of basis on the power of the test, simulation Scenario A, $n=300$.

2.4 Real Data Applications

Through simulations, we have shown our proposed testing method can identify significant covariates under both dense and sparse sampling design, even when the covariates are observed with measurement error. Next, we present two real data applications of our testing method to demonstrate its usefulness in identifying significant time varying covariates in practical problems. We first consider the study of gait deficiency which is a typical case of dense data with small measurement error, subsequently, we also apply our method to a study of dietary calcium absorption, where the data is sparse and measurement error is relatively higher.

2.4.1 Gait Data

In this study, the goal is to understand how the joints in hip and knee interact during a gait cycle (Theologis, 2009), (Ramsay and Silverman, 2005). Here, there are longitudinal measurements of hip and knee angles taken on 39 children on 20 equispaced evaluation points in $\mathcal{T} = [0, 1]$. Figure 2.5 displays the observed individual trajectories of the hip and knee angles. The study of

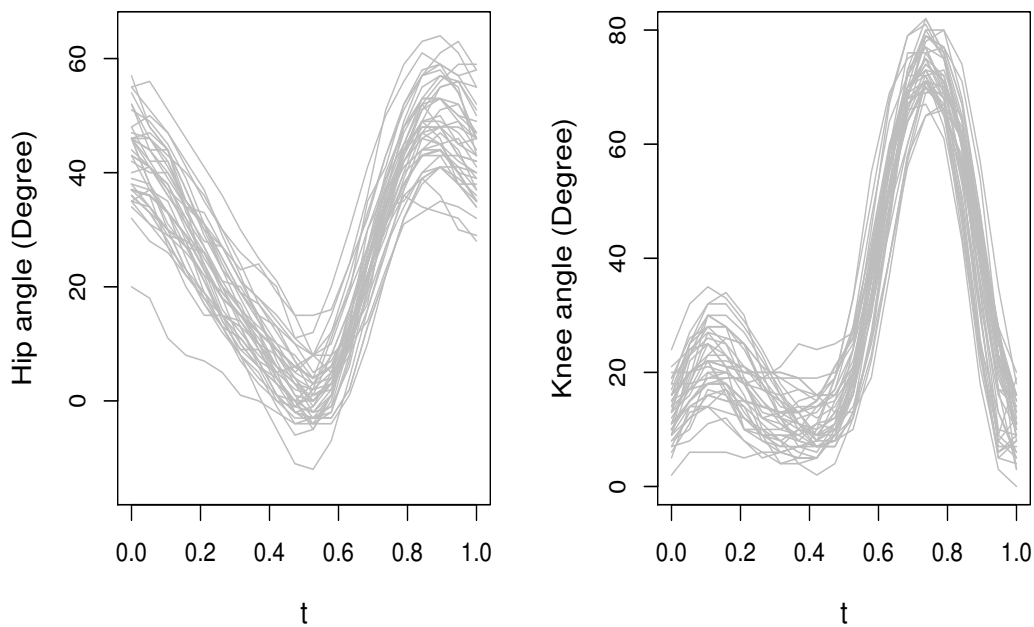


Figure 2.5 Measurement of hip angles and knee angles in the gait study.

gait is important as it helps to identify issues causing pain, and also implement and evaluate treatments to correct abnormalities. As discussed earlier, one natural question to ask here is whether the knee angles (response) are at all associated with the hip angles (covariate). In our terminology, here $Y_i(t)$ is the knee angle, we assume the hip angles $X_i(t)$ are observed with measurement error, i.e, we observe $U_i(t_{ij}) = X_i(t_{ij}) + \delta_{ij}$, where δ_{ij} are assumed to be white

noise. Kim et al. (2018) investigated the effect of knee angles on hip angles and found the effect to be linear. In our functional linear concurrent modeling setup, we are therefore interested in testing $H_0 : \beta_1(t) = 0$ against the alternative $H_1 : \beta_1(t) \neq 0$, where $\beta_1(t)$ denotes the linear concurrent effect of hip angles on knee angles. We use our proposed testing method with 14 cubic B-splines to model $\beta_0(t)$ and $\beta_1(t)$. The p-value of the proposed test is calculated to be .0004. So we reject the null hypothesis and conclude that knee angle at any fixed time point is associated with the hip angle at the same time point. Our findings match with that of Kim et al. (2018), who used the bootstrapped-F test method.

As the sample size ($n = 39$) for this data is small and our testing method is an asymptotic one, we further evaluate the performance of our proposed method using a simulation study that captures the feature of the gait data. This also enables us to see the power performance of our method. This is done similarly as that of Kim et al. (2018). We use a model that mimics the feature of the gait data, generate a large simulated data set and assess the power performance of our method on the simulated data. In particular, We generate the covariate $X_i(t)$ from a process with the mean and covariance functions that equal their estimated counterparts from the data using FPCA. We also estimate the parameters $\beta_0(t)$, $\beta_1(t)$ from the fit of our full model and estimate $\Sigma(s, t)$ by doing FPCA on the residuals obtained from the full model fit. Then we generate observations using the model $Y_i(t) = \hat{\beta}_0(t) + d\{\hat{X}_i(t)\hat{\beta}_1(t)\} + \epsilon_i(t)$, where $\epsilon_i(\cdot) \sim \mathcal{N}(0, \hat{\Sigma}(\cdot, \cdot))$. We simulate $n = 300$ response curves $Y_i(t)$ from the above set up. In the above set up d again plays the role of a parameter, which controls departure from the null hypothesis. We perform a power analysis simulating 1000 such data sets from the above scenario for various d . Figure 2.6 shows the power curve obtained from our analysis.

For $d = 0$, the type I error is .062, and nominal level $\alpha = .05$ is within its 2-standard errors limit. As d increases the power gradually increases and ultimately goes to one when $d = 0.35$, which ensures power is one at $d = 1$. This result further confirms our conclusion that knee angle during a gait cycle is associated with hip angle.

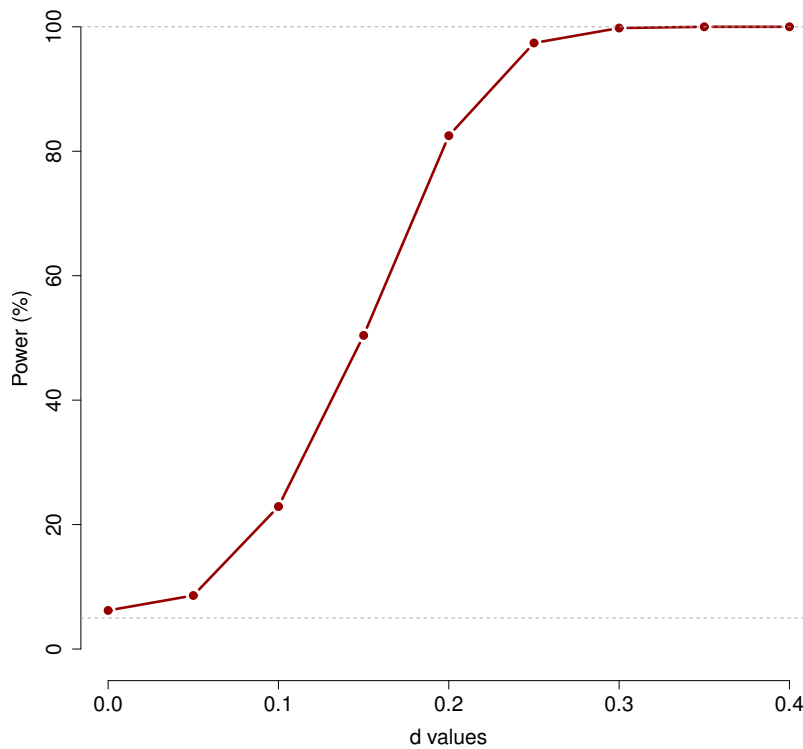


Figure 2.6 Results from simulation study mimicking gait data. Displayed is the power curve of our test.

2.4.2 Calcium Absorption Data

We consider the dietary calcium absorption study given in Davis (2002). In this study the subjects are a group of 188 patients. We have data on calcium absorption, dietary calcium intake and BMI of these patients at irregular intervals between 35 to 64 years of their ages. The number of repeated measurements for each patient is between 1 to 4. Figure 2.7 shows the individual curves of patients' calcium absorption and calcium intake along their ages.

We notice that this is a typical case of sparse data with relatively high measurement error. In this study, we are interested in finding whether calcium intake has any effect on calcium absorption in presence of BMI. We assume the covariates BMI $X_{i1}(t)$ and calcium intake $X_{i2}(t)$

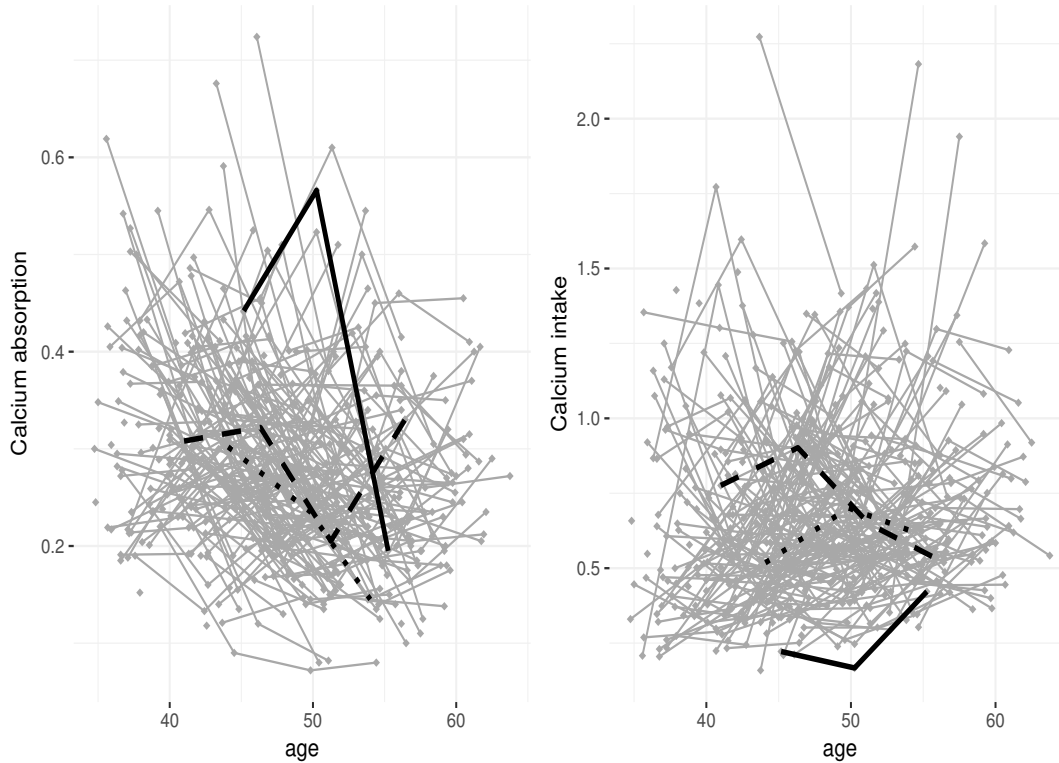


Figure 2.7 Observed calcium absorption and calcium intake in the calcium absorption study.

are observed with measurement error. So our observed data is $U_i(t_{ij}) = X_{i1}(t_{ij}) + \delta_{ij}$ and $V_i(t_{ij}) = X_{i2}(t_{ij}) + \nu_{ij}$.

Our response variable is calcium absorption $Y_i(t)$. Following the study in Kim et al. (2018) we assume a functional linear concurrent regression model, $Y_i(t) = \beta_0(t) + X_{i1}(t)\beta_1(t) + X_{i2}(t)\beta_2(t) + \epsilon_i(t)$, and want to test $H_0 : \beta_2(t) = 0$ against the alternative $H_1 : \beta_2(t) \neq 0$. First we use FPCA as discussed in Section 2.2.5 on noisy covariates to get their smooth counterparts at time points where we have $Y_i(t)$ available and then apply our one sided score test method for multiple covariates described as in Section 2.2.6. We use 12 cubic B-Splines to model $\beta_0(t)$, $\beta_1(t)$ and $\beta_2(t)$. The p-value of our test is calculated to be $< 10^{-5}$. Thus we conclude in presence of BMI, calcium intake of patients has a significant effect on calcium absorption which again matches with the findings in Kim et al. (2018) and other studies of dietary calcium absorption.

2.5 Discussion and Future Work

In this article, we have proposed a likelihood based method for testing of hypothesis in functional linear concurrent regression. We have formulated the problem as a test for variance component and have used a one sided score test approach. We have established the asymptotic null distribution of our test statistic under some standard assumptions. Through simulations, we have shown our proposed method maintains the nominal type I error rate and also yields higher power compared to the existing bootstrapped-F test, even when data is observed sparsely and with measurement error. We have successfully applied our method in finding significant covariates in two real data applications, namely the gait study and calcium absorption study. We note our method is a general one and can be applied for testing in longitudinal data setting too, where we have more flexibility in assuming a parametric form of error covariance structure and estimating it consistently from the data.

We have considered a Gaussian distribution for the functional error in our model, if the distribution is non-Gaussian, one can still use the score test statistic proposed in this article and employ a subject level bootstrap method for performing the test. Similarly, it would be of interest to explore how the score test approach can be extended to generalized functional concurrent models such as logistic regression. In developing our test, we have used a random effects formulation of the problem arising from directly penalizing the coefficient functions. It is also plausible to use penalty on r -th ($r > 0$) derivative of the coefficient functions. In this case, the main challenge is to handle the singularity of the resulting covariance matrices, which can be addressed by using mixed effects model and subsequently testing for variance component using variants of likelihood ratio tests (Crainiceanu and Ruppert, 2004). Another important work for the future could be to prove the consistency results of the FPCA approximation methods used in this article. Further, we would like to extend our testing method to nonparametric concurrent functional regression and more general function on function regression models and this is also a possible area for future research.

Software

Software in the form of R code illustrating implementation of the proposed method, together with the data set and complete documentation is available at GitHub (https://github.com/rahulfrodo/FLCM_Score).

Acknowledgements

We thank Dr. Janet Kim and Dr. Ana-Maria Staicu for kindly sharing the results of their method (Kim et al., 2018) with us.

Chapter 3

Variable Selection in Functional Linear Concurrent Regression

3.1 Introduction

Function on function regression is an active area of research in functional data with new statistical methods emerging frequently to address data where both the response variable and the covariates are functions over some continuous index such as time. Functional concurrent regression model is a special case of function on function regression, where the predictor variables influence the response variable only through their value at the current time point (Kim et al., 2018). The commonly used functional linear concurrent regression model assumes a linear relationship between the response and the predictors, where the value of the response at a particular time point is modeled as a linear combination of the covariates at that specific time point, and the coefficients of the functional covariates are univariate smooth functions over time (Ramsay and Silverman, 2005). Multiple methods exist in literature for estimation of these regression functions in functional linear concurrent regression and the closely related varying coefficient model (Hastie and Tibshirani, 1993), using kernel-local polynomial smoothing (Wu et al., 1998; Hoover et al., 1998; Fan and Zhang, 1999; Kauermann and Tutz, 1999), polynomial spline (Huang

et al., 2002, 2004), smoothing spline (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001; Eubank et al., 2004) among many others. Similar to classical scalar regression, when there are a large number of covariates present, the primary interest might be to select only the set influential variables and estimate their effects. While doing significance testing and building confidence bands can help for assessing the individual effect of a predictor, they are computationally infeasible, when the number of covariates is large. Thus arises the need to perform variable selection in functional linear concurrent regression.

Our research in this article is motivated by a fisheries footprint study where the goal is to identify important time varying socio-structural and economic drivers influencing fisheries footprint (Global Footprint Network's measure of total marine area required to produce the amount of seafood products a nation consumes) and to estimate their time varying effects. Although, a number of variable selection methods have been developed for scalar on function regression (Gertheiss et al., 2013; Fan et al., 2015) and function on scalar regression (Chen et al., 2016), literature for variable selection in functional linear concurrent regression is relatively sparse. Recently Goldsmith and Schwartz (2017) developed a variable selection method in functional linear concurrent model using a variational Bayes approach with sparsity being introduced through a spike and slab prior on the coefficients of the basis expansion of the regression functions. In this article we propose a variable selection method in functional linear concurrent regression extending the classically used variable selection methods like LASSO (Tibshirani, 1996) , SCAD (Fan and Li, 2001) and MCP (Zhang, 2010).

Our work is inspired from Gertheiss et al. (2013), where they show the variable selection problem in scalar on function regression scenario can be reduced to a group LASSO (Yuan and Lin, 2006) problem. We have shown in functional linear concurrent regression the variable selection problem can be addressed as a group LASSO, and their natural extension group SCAD or group MCP problem. Chen et al. (2016) also used group MCP for their variable selection in function on scalar regression. Our model is fundamentally different from them in the sense the covariates we consider are time varying functions. Our method is similar to Wang et al. (2008)

in which they use a group SCAD penalty for variable selection in varying coefficient models, but we propose a different penalty on the coefficient functions which simultaneously penalizes departure from sparsity as well as roughness of the coefficient functions, and our research shows there is much to be gained by using the group MCP penalty. A pre-whitening procedure similar to Chen et al. (2016) is employed to take into account temporal dependence present within functions. We also consider that the covariates might be contaminated with measurement error and therefore use functional principal component analysis (FPCA) to get denoised trajectories of the covariates which improves estimation accuracy of our approach. Through simulations, we illustrate our method, particularly with group SCAD or group MCP penalty, can pick out the relevant variables with very high accuracy and has minuscule false positive and false negative rate even when data is observed sparsely and is contaminated with measurement error. We demonstrate two real data applications of our method in the study of dietary calcium absorption (Davis, 2002) and the fisheries footprint study.

The rest of the article is organized as follows. In Section 3.2 we present our modeling framework and illustrate our variable selection method. In Section 3.3 we conduct a simulation study to evaluate the performance of our method and summarize the simulation results. In Section 3.4 we go back to the two real data examples; calcium absorption study and fisheries footprint study, apply our variable selection method to find out the influential covariates and present our findings. We conclude in Section 3.5 with a discussion about some limitations and possible extensions of our work.

3.2 Methodology

3.2.1 Modeling Framework and Variable Selection Method

Suppose that the observed data for the i -th subject is given by $\{Y_i(t), X_{i1}(t), X_{i2}(t), \dots, X_{ip}(t)\}$

for $(i = 1, 2, \dots, n)$, where $Y_i(\cdot)$ is a functional response and $X_{i1}(\cdot), X_{i2}(\cdot), \dots, X_{ip}(\cdot)$ are the corresponding functional covariates. We assume the covariates and the response are observed on a fine and regular grid of points $S = \{t_1, t_2, \dots, t_m\} \subset S = [0, T]$ for some $T > 0$, and the covariates are measured without any error. We discuss later in this section how our model and method can be easily extended to accommodate more general scenarios where the covariates are contaminated with measurement error and observed sparsely. We consider a functional linear concurrent regression model of the form,

$$Y_i(t) = \sum_{j=1}^p X_{ij}(t)\beta_j(t) + \epsilon_i(t), \quad (3.1)$$

where $\beta_j(t)$ ($j = 1, 2, \dots, p$) are smooth functions (finite second derivative) representing the functional regression parameters. We assume $X_{ij}(\cdot)$ are independent and identically distributed (i.i.d.) copies of $X_j(\cdot)$ ($j = 1, 2, \dots, p$), where $X_j(\cdot)$ s are underlying smooth stochastic processes. We further assume $\epsilon_i(\cdot)$ are i.i.d copies of $\epsilon(\cdot)$, which is a mean zero stochastic process. The model (3.1) in stacked form can be rewritten as $Y(t) = X(t)\beta(t) + \epsilon(t)$. Generally in functional linear concurrent regression, estimation is done (Ramsay and Silverman, 2005) by minimizing the penalized residual sum of square, $SSE(\beta) = \int r(t)^T r(t)dt + \sum_{j=1}^p \lambda_j \int (L_j \beta_j(t))^2 dt$, where $r(t) = Y(t) - X(t)\beta(t)$. For example when $L_j = I$, we minimize $\int r(t)^T r(t)dt + \sum_{j=1}^p \lambda_j \int (\beta_j(t))^2 dt$. Now suppose $\{\theta_{kj}(t), k = 1, 2, \dots, k_j\}$ is a set of known basis functions for $j = 1, 2, \dots, p$. We model the unknown coefficient functions using basis function expansion as $\beta_j(t) = \sum_{k=1}^{k_j} b_{kj} \theta_{kj}(t) = \boldsymbol{\theta}_j(t)^T \mathbf{b}_j$, where $\boldsymbol{\theta}_j(t) = [\theta_{1j}(t), \theta_{2j}(t), \dots, \theta_{k_j j}(t)]^T$ and $\mathbf{b}_j = (b_{1j}, b_{2j}, \dots, b_{k_j j})^T$ is a vector of unknown coefficients. In this article, we use B-spline basis functions, however, other basis functions can be used as well. Then the minimization in the example mentioned above, can be carried out by minimizing $\int \{Y(t) - X(t)\boldsymbol{\Theta}(t)\mathbf{b}\}^T \{Y(t) - X(t)\boldsymbol{\Theta}(t)\mathbf{b}\} dt + \mathbf{b}^T \mathbb{R} \mathbf{b}$. Here \mathbf{b} , $\boldsymbol{\Theta}(t)$ and penalty matrix \mathbb{R} are defined in stacked form as $\mathbf{b} = (\mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_p^T)^T$, $\boldsymbol{\Theta}(t) = \{\boldsymbol{\theta}_1(t)^T, \boldsymbol{\theta}_2(t)^T, \dots, \boldsymbol{\theta}_p(t)^T\}$ and $\mathbb{R} = \text{diag}(\mathbb{R}_1, \mathbb{R}_2, \dots, \mathbb{R}_p)$, where $\mathbb{R}_j = \lambda_j \mathbf{b}_j^T \{ \int \boldsymbol{\theta}_j(t) \boldsymbol{\theta}_j(t)^T dt \} \mathbf{b}_j$. For our variable selection method we define penalty on the regression functions $\beta_j(\cdot)$ as, $P_{\lambda, \psi} \{\beta_j(\cdot)\} = \lambda \int \beta_j(t)^2 dt +$

$\psi \int \{\beta_j''(t)\}^2 dt\}^{1/2} = \lambda \left(\mathbf{b}_j^T \mathbb{R}_j \mathbf{b}_j + \psi \mathbf{b}_j^T \mathbb{Q}_j \mathbf{b}_j \right)^{1/2} = \lambda \left(\mathbf{b}_j^T \mathbb{K}_{\psi,j} \mathbf{b}_j \right)^{1/2}$, where $\mathbb{K}_{\psi,j} = \mathbb{R}_j + \psi \mathbb{Q}_j$, $\mathbb{R}_j = \{\int \boldsymbol{\theta}_j(t) \boldsymbol{\theta}_j(t)^T dt\}$, $\mathbb{Q}_j = \{\int \boldsymbol{\theta}_j''(t) \boldsymbol{\theta}_j''(t)^T dt\}$. This penalty was originally proposed by Meier et al. (2009) and later used by Gertheiss et al. (2013) for their variable selection method in scalar on function regression. The parameter $\psi \geq 0$ controls the amount of penalization on the roughness penalty. The proposed penalty simultaneously penalizes departure from sparsity and roughness of the coefficient functions ensuring the resulting coefficient functions are smooth and small coefficient functions are shrunk to zero introducing sparsity. Subsequently, we propose to minimize the following penalized mean sum of square of the residuals for variable selection,

$$L(\mathbf{b}) = 1/n \int \{Y(t) - X(t)\boldsymbol{\Theta}(t)\mathbf{b}\}^T \{Y(t) - X(t)\boldsymbol{\Theta}(t)\mathbf{b}\} dt + \lambda \sum_{j=1}^p (\mathbf{b}_j^T \mathbb{K}_{\psi,j} \mathbf{b}_j)^{1/2}. \quad (3.2)$$

Since we assume data is observed in a dense equispaced grid, the variable selection in practice is carried out by minimizing the following equivalent criterion,

$$\sum_{i=1}^n \sum_{l=1}^m [Y_i(t_l) - \sum_{j=1}^p X_{ij}(t_l) \{\sum_{k=1}^{k_j} b_{kj} \theta_{kj}(t_l)\}]^2 + \lambda mn \sum_{j=1}^p (\mathbf{b}_j^T \mathbb{K}_{\psi,j} \mathbf{b}_j)^{1/2}. \quad (3.3)$$

Now using Cholesky decomposition of $\mathbb{K}_{\psi,j} = \mathbb{L}_{\psi,j} \mathbb{L}_{\psi,j}^T$ and denoting $\boldsymbol{\gamma}_j = \mathbb{L}_{\psi,j}^T \mathbf{b}_j$, the penalized sum of square of residuals can be reformulated as,

$$\begin{aligned} R(\boldsymbol{\gamma}) &= \sum_{i=1}^n \sum_{l=1}^m [Y_i(t_l) - \sum_{j=1}^p X_{ij}(t_l) \{\sum_{k=1}^{k_j} b_{kj} \theta_{kj}(t_l)\}]^2 + \lambda mn \sum_{j=1}^p (\mathbf{b}_j^T \mathbb{K}_{\psi,j} \mathbf{b}_j)^{1/2} \\ &= \sum_{i=1}^n \sum_{l=1}^m [Y_i(t_l) - \sum_{j=1}^p \mathbf{Z}_{ij}^*(t_l)^T \mathbf{b}_j]^2 + \lambda mn \sum_{j=1}^p (\mathbf{b}_j^T \mathbb{K}_{\psi,j} \mathbf{b}_j)^{1/2}, \quad \mathbf{Z}_{ij}^*(t_l)^T = X_{ij}(t_l) \times \boldsymbol{\theta}_j(t_l)^T \\ &= \sum_{i=1}^n \sum_{l=1}^m [Y_i(t_l) - \sum_{j=1}^p \tilde{\mathbf{Z}}_{ij}^{*}(t_l)^T \boldsymbol{\gamma}_j]^2 + \lambda mn \sum_{j=1}^p (\boldsymbol{\gamma}_j^T \boldsymbol{\gamma}_j)^{1/2}, \quad \text{where } \tilde{\mathbf{Z}}_{ij}^{*}(t_l) = \mathbb{L}_{\psi,j}^{-1} \mathbf{Z}_{ij}^*(t_l) \\ &= \sum_{i=1}^n \|\mathbf{Y}_i - \mathbb{Z}_i^* \boldsymbol{\gamma}\|_2^2 + \lambda mn \sum_{j=1}^p (\boldsymbol{\gamma}_j^T \boldsymbol{\gamma}_j)^{1/2} = \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \boldsymbol{\gamma}_j\|_2^2 + \lambda mn \sum_{j=1}^p (\boldsymbol{\gamma}_j^T \boldsymbol{\gamma}_j)^{1/2}, \end{aligned}$$

where $\mathbf{Y}_i = (Y_i(t_1), Y_i(t_2), \dots, Y_i(t_m))^T$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \dots, \boldsymbol{\gamma}_p^T)^T$ and \mathbb{Z}_i^* is defined as follows,

$$\mathbb{Z}_i^* = \begin{bmatrix} \tilde{\mathbf{Z}}_{i1}^*(t_1)^T & \tilde{\mathbf{Z}}_{i2}^*(t_1)^T & \tilde{\mathbf{Z}}_{i3}^*(t_1)^T & \dots & \tilde{\mathbf{Z}}_{ip}^*(t_1)^T \\ \tilde{\mathbf{Z}}_{i1}^*(t_2)^T & \tilde{\mathbf{Z}}_{i2}^*(t_2)^T & \tilde{\mathbf{Z}}_{i3}^*(t_2)^T & \dots & \tilde{\mathbf{Z}}_{ip}^*(t_2)^T \\ \dots & \dots & \dots & \dots & \dots \\ \tilde{\mathbf{Z}}_{i1}^*(t_m)^T & \tilde{\mathbf{Z}}_{i2}^*(t_m)^T & \tilde{\mathbf{Z}}_{i3}^*(t_m)^T & \dots & \tilde{\mathbf{Z}}_{ip}^*(t_m)^T \end{bmatrix}.$$

Here \mathbb{Z}_i^{*j} refers to the j th block column in this matrix. We recognize this minimization problem as performing a group LASSO (Yuan and Lin, 2006), where the grouping is introduced by covariates. In particular we obtain estimates of γ_j by minimizing similar penalized least square as in group LASSO namely;

$$\begin{aligned} \hat{\gamma} &= \underset{\gamma_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \gamma_j\|_2^2 + \lambda mn \sum_{j=1}^p (\gamma_j^T \gamma_j)^{1/2} \\ &= \underset{\gamma_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \gamma_j\|_2^2 + \lambda mn \sum_{j=1}^p \|\gamma_j\|_2 \\ &= \underset{\gamma_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \gamma_j\|_2^2 + mn \sum_{j=1}^p P_{LASSO,\lambda}(\|\gamma_j\|_2). \end{aligned} \quad (3.4)$$

We extend this group LASSO formulation to non convex penalties, which are known (Breheny and Huang, 2015; Mazumder et al., 2011) to produce sparser and less biased estimates, especially when there are large number of variables and magnitude of original parameter of interest is large. In particular we propose to use two non convex penalties; SCAD (Fan and Li, 2001) and MCP (Zhang, 2010). These two penalties have been shown to ensure selection consistency and estimation consistency under standard assumptions in scalar regression case. They also enjoy the so called oracle property in which they behave like oracle MLE asymptotically. This motivates us to use them in our functional variable selection context, and then the problem of variable selection reduces to a group SCAD or group MCP problem in our modeling set up as follows.

Group SCAD Method

In this method we perform variable selection and obtain estimates of γ as

$$\hat{\gamma} = \underset{\gamma_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \gamma_j\|_2^2 + mn \sum_{j=1}^p P_{SCAD, \lambda, \phi}(\|\gamma_j\|_2), \quad (3.5)$$

where $P_{SCAD, \lambda, \phi}(\|\gamma_j\|_2)$ is defined in the following way:

$$P_{SCAD, \lambda, \phi}(\|\gamma_j\|_2) = \begin{cases} \lambda \|\gamma_j\|_2 & \text{if } \|\gamma_j\|_2 \leq \lambda. \\ \frac{\lambda \phi \|\gamma_j\|_2 - .5(\|\gamma_j\|_2^2 + \lambda^2)}{\phi - 1} & \text{if } \lambda < \|\gamma_j\|_2 \leq \lambda \phi. \\ .5\lambda^2(\phi + 1) & \text{if } \|\gamma_j\|_2 > \lambda \phi. \end{cases}$$

Group MCP Method

For Group MCP method estimation of γ is done similarly,

$$\hat{\gamma} = \underset{\gamma_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|\mathbf{Y}_i - \sum_{j=1}^p \mathbb{Z}_i^{*j} \gamma_j\|_2^2 + mn \sum_{j=1}^p P_{MCP, \lambda, \phi}(\|\gamma_j\|_2), \quad (3.6)$$

where $P_{MCP, \lambda, \phi}(\|\gamma_j\|_2)$ is defined as,

$$P_{MCP, \lambda, \phi}(\|\gamma_j\|_2) = \begin{cases} \lambda \|\gamma_j\|_2 - \frac{\|\gamma_j\|_2^2}{2\phi} & \text{if } \|\gamma_j\|_2 \leq \lambda \phi. \\ .5\lambda^2 \phi & \text{if } \|\gamma_j\|_2 > \lambda \phi. \end{cases}$$

3.2.2 Incorporating Covariance Structure into variable selection

The variable selection method proposed in Section 3.2.1 does not account for possible correlation in error process. In reality, however, temporal correlation is more likely to be present within functions. While using an independent working correlation structure can yield consistent and unbiased estimates, incorporating the true covariance structure in the variable selection criterion (3.4), (3.5), or (3.6) may give definite gains in terms of performance as illustrated by Chen et al. (2016). We follow a similar pre whitening procedure employed by Chen et al. (2016) to take into account the underlying covariance structure. We assume the error process $\epsilon(t)$ has the form $\epsilon(t) = V(t) + w_t$, where $V(t)$ is a smooth mean zero stochastic process with

covariance kernel $G(s, t)$ and w_t is a white noise with variance σ^2 . The covariance function of the error process is then given by $\Sigma(s, t) = \text{cov}\{\epsilon(s), \epsilon(t)\} = G(s, t) + \sigma^2 I(s = t)$. For data observed on dense and regular grid the covariance matrix of the residual vector is the given by $\Sigma = \text{diag}\{\Sigma_{m \times m}, \Sigma_{m \times m}, \dots, \Sigma_{m \times m}\}$, where $\Sigma_{m \times m}$ denotes the covariance kernel $\Sigma(s, t)$ evaluated at $S = \{t_1, t_2, \dots, t_m\}$. Now if $\Sigma_{m \times m}$ is known, redefining \mathbf{Y}_i and \mathbb{Z}_i^{*j} as $\mathbf{Y}_i = \{\Sigma_{m \times m}^{-1/2}\} \mathbf{Y}_i$, $\mathbb{Z}_i^{*j} = \{\Sigma_{m \times m}^{-1/2}\} \mathbb{Z}_i^{*j}$, the same penalized criterion (4), (5) or (6) can be used to perform variable selection.

In reality Σ is unknown, and we need an estimator $\hat{\Sigma}$. In the context of functional data, we want to estimate $\Sigma(\cdot, \cdot)$ nonparametrically. If we had the original residuals ϵ_{ij} available, we could use functional principal component analysis (FPCA), e.g., Yao et al. (2005) or Zhang et al. (2007) to estimate $\Sigma(s, t)$. If the covariance kernel $G(s, t)$ of the smooth part $V(t)$ is a Mercer kernel (Mercer, 1909), by Mercer's theorem $G(s, t)$ must have a spectral decomposition

$$G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ are the ordered eigenvalues and $\phi_k(\cdot)$ s are the corresponding eigenfunctions. Thus we have the decomposition $\Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t) + \sigma^2 I(s = t)$. Given $\epsilon_{t_{ij}} = V(t_{ij}) + w_{ij}$, one could employ FPCA based methods to get $\hat{\phi}_k(\cdot)$, $\hat{\lambda}_k$ s and $\hat{\sigma}^2$. So an estimator of $\Sigma(s, t)$ can be formed as $\hat{\Sigma}(s, t) = \sum_{k=1}^K \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t) + \hat{\sigma}^2 I(s = t)$, where K is large enough for the convergence to hold and is typically chosen such that percent of variance explained (PVE) by the selected eigencomponents exceeds some pre-specified value such as 99% or 95%. In reality we don't have the original residuals ϵ_{ij} and use the full model (1) to obtain residuals $e_{ij} = Y_i(t_j) - \hat{Y}_i(t_j)$. Then treating e_{ij} as our original residuals, we obtain $\hat{\Sigma}(s, t)$ using FPCA.

Remark 1: We use cubic B-spline basis with the same number of basis functions to model the regression functions $\beta_j(t)$ s, where the number of basis is large so the basis is rich enough. For selection of the tuning parameter ψ (for smoothness) and the penalty parameter λ , we use the Extended Bayesian information criteria (EBIC) (Chen and Chen, 2008) corresponding to the

equivalent linear model of criterion (4), (5) or (6) and this has shown good performance in our simulation study. Chen and Chen (2008) established consistency of EBIC under standard assumptions and illustrated its superiority over other methods like cross-validation, AIC, and BIC, which tend to over select the variables. For tuning parameter ϕ we use the values 4 for SCAD and 3 for MCP, as proposed by the original authors. For model fitting we use ‘grpreg’ package (Breheny, 2019) in R.

Remark 2: In practice, we recommend standardizing the variables either using Euclidean norm (automatically performed in ‘grpreg’) or using FPCA based methods ($X_j^*(t) = \frac{X_j(t) - \mu_j(t)}{\sigma_j(t)}$), which is especially useful for highly sparse data where some B-splines might not have observed data on its support. This can help in faster convergence of the proposed method. We performed both the standardization methods in our simulation studies and obtained very similar results.

3.2.3 Extension to Sparse data and Noisy Covariates

More generally, we can consider the case where data is observed sparsely and covariates are observed with measurement error. This is most often the case for longitudinal data. Here the observed data is the response $\{(Y_i(t_{ij}), t_{ij}), j = 1, 2, \dots, m_i\}$ and the observed covariates $\{(U_1(t_{1ij}), t_{1ij}), j = 1, 2, \dots, m_{1i}\}, \{(U_2(t_{2ij}), t_{2ij}), j = 1, 2, \dots, m_{2i}\}, \dots, \{(U_p(t_{pij}), t_{pij}), j = 1, 2, \dots, m_{pi}\}$. Let us denote $U_k(t_{kij})$ s, ($k = 1, 2, 3, \dots, p$) by U_{ijk} . Here U_{ijk} s represent the observed covariates with measurement error, i.e., we have $U_{ijk} = X_k(t_{kij}) + e_{ijk}$ for $i = 1, 2, \dots, n$, $j = 1, 2, \dots, m_{ki}$ and $k = 1, 2, \dots, p$. The measurement error e_{ijk} are assumed to be white noises with zero mean and variance σ_k^2 . In sparse data set up it is generally assumed (Kim et al., 2018) although individual number of observations m_i is small, $\bigcup_{i=1}^n \bigcup_{j=1}^{m_i} t_{ij}$ is dense in $[0, T]$. Then we reconstruct the original curves from the observed sparse and noisy curves using FPCA methods (Yao et al., 2005) by estimating the eigenvalues and eigenfunctions corresponding to the original curves. Li and Hsing (2010) proved uniform convergence of the mean, eigenvalues and eigenfunctions associated with the curves for both dense and in particular sparse design under suitable regularity conditions. For prediction of the scores we use

PACE method as in Yao et al. (2005). Then these estimates are put together using Karhunen-Loève expansion (Karhunen, Loeve 1946) to get estimates $\hat{X}_{ik}(\cdot)$ of the true curves $X_{ik}(\cdot)$ as $\hat{X}_{ik}(t) = \hat{\mu}_k(t) + \sum_{s=1}^S \hat{\zeta}_{isk} \hat{\psi}_{sk}(t)$, where the number of eigenfunctions S to use is chosen using the percent of variance explained (PVE) criterion, which is the percentage of variance explained by the first few eigencomponents. Alternatively one can also use multivariate FPCA (Happ and Greven, 2018) instead of running FPCA on each predictor variable separately. Then for sparse data observed on irregular grid and observed with measurement error, we use $\{Y_i(t_{ij}), \hat{X}_{i1}(t_{ij}), \hat{X}_{i2}(t_{ij}), \dots, \hat{X}_{ip}(t_{ij})\}_{i=1}^n$ as our original data and use this data for performing variable selection.

3.3 Simulation Study

3.3.1 Simulation Set Up

In this section, we evaluate the performance of our variable selection method using a simulation study. To this end we generate data from the model,

$$Y_i(t) = \beta_0(t) + \sum_{j=1}^{20} X_{ij}(t)\beta_j(t) + \epsilon_i(t) \quad i = 1, 2, \dots, n, \quad t \in [0, 100].$$

The regression functions representing the dynamic effects are given by $\beta_0(t) = 8\sin(\pi t/50)$, $\beta_1(t) = 5\sin(\pi t/100)$, $\beta_2(t) = 4\sin(\pi t/50) + 4\cos(\pi t/50)$, $\beta_3(t) = 25e^{-t/20}$ and rest of the $\beta_j(t) = 0$ for $j = 4, 5, 6, \dots, 20$, i.e., the last 17 covariates are not relevant. The original covariates $X_{ij}(\cdot) \stackrel{iid}{\sim} X_j(\cdot)$, where $X_j(t)$ ($j = 1, 2, \dots, 20$) are given by $X_j(t) = a_j\sqrt{2}\sin(\pi jt/400) + b_j\sqrt{2}\cos(\pi jt/400)$, where $a_j \sim \mathcal{N}(50, (2)^2)$, $b_j \sim \mathcal{N}(50, (2)^2)$. We moreover assume that $X_{ij}(t)$ are observed with measurement error i.e., we observe $U_{ij}(t) = X_{ij}(t) + \delta_j$, where $\delta_j \sim \mathcal{N}(0, .6^2)$. The error process $\epsilon_i(t)$ is generated as follows;

$$\epsilon_i(t) = \xi_{i1}\cos(t) + \xi_{i2}\sin(t) + N(0, 1^2 I_{m_i}),$$

where $\xi_{i1} \stackrel{iid}{\sim} \mathcal{N}(0, .5^2)$ and $\xi_{i2} \stackrel{iid}{\sim} \mathcal{N}(0, 0.75^2)$. The response $Y_i(t)$ and noisy covariate $U_{ij}(t)$'s are observed sparsely in random m_i points in S , where S is the set of $m = 81$ equidistant time points in $[0, 100]$ and $m_i \stackrel{iid}{\sim} \text{Uniform}\{30, 31, \dots, 41\}$. Three sets of sample size $n \in \{100, 200, 400\}$ are considered. For each sample size, we use 500 replicated datasets for evaluation of our method.

3.3.2 Simulation Results

Our primary interest is selection (identification) of the relevant covariates $X_1(\cdot), X_2(\cdot), X_3(\cdot)$ and estimating their effects $\beta_1(t), \beta_2(t), \beta_3(t)$ accurately. As the covariates are observed sparsely and with measurement error, we apply FPCA as discussed in Section 3.2.3 with PVE = 99% and obtain the denoised curves $\hat{X}_{ij}(t)$ before applying our variable selection method. We apply the proposed variable selection method with and without the pre whitening procedure mentioned in Section 3.2.2. Table 3.1 and Table 3.2 display the selection percentage of each variable for each of the three selection methods discussed in Section 3.2 and for the three sample sizes $n = 100, 200, 400$, for the non pre whitened and pre whitened case respectively. We expect that the group LASSO selection method to have a higher false positive rate and use this as a benchmark for comparison. It can be seen from Table 3.1 and 3.2 that all the three methods (group LASSO, group SCAD, group MCP) pick out the three true covariates $X_1(\cdot), X_2(\cdot), X_3(\cdot)$; 100% of the time. The group LASSO method has a high false positive selection percentage as can be seen in both Table 3.1 and Table 3.2, with selection accuracy improving with increasing sample size. The group SCAD and group MCP method, on the other hand, have false selection percentage in the range of 0.2% – 1% for non pre whitened case and exactly 0% for pre whitened case. In other words, the group SCAD and group MCP are able to identify the true model using the pre whitening procedure. In scalar regression, SCAD and MCP are known to produce sparser solutions than LASSO due to its concave nature, and here also in the context of variable selection in functional linear concurrent regression we observe these two methods (their group extension) out performing LASSO. The average model sizes for each scenario are also given in Table 3.1, and the group SCAD and group MCP method produce smaller and closer values

to the true model size 3 (exactly 3 with pre whitening procedure in Table 3.2). These results also illustrate the benefit of pre-whitening and henceforth we have used pre-whitening as a pre processing step to perform variable selection using the proposed methods.

Next, as an assessment of the accuracy of the estimates $\hat{\beta}_k(t)$ ($k = 1, 2, 3$), we plot the true regression curves overlaid by their Monte Carlo (MC) mean estimate. MC point-wise confidence intervals (95%) (corresponding to 2.5 and 97.5 percentile) for each of the three curves are also plotted to assess variability of the estimates. Figure 3.1 displays this plot for $n = 200$, the plots for $n = 100, n = 400$ are similar with more accuracy and less variability for larger sample sizes.

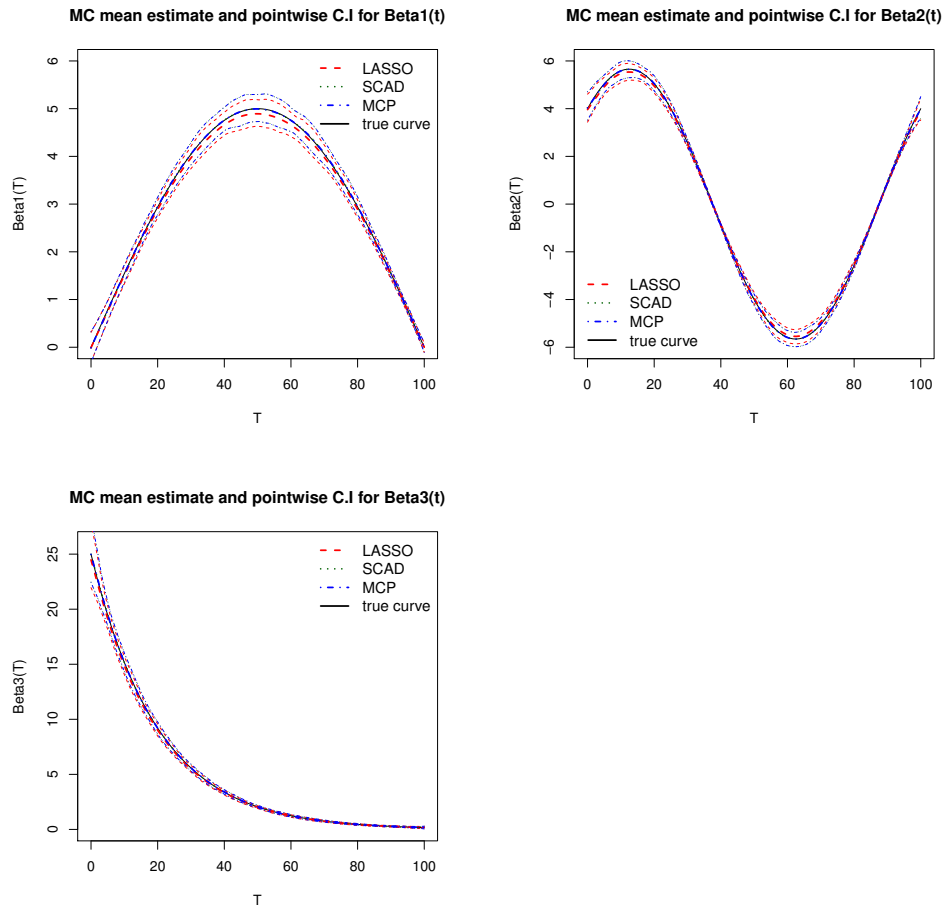


Figure 3.1 MC estimates and point wise confidence intervals of the coefficient functions ($n=200$).

Table 3.1 Comparison of selection percentages (%) of different variables and average model size, without pre whitening.

Sample Size	Method	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20	Avg Model Size
n=100	LASSO	100	100	100	16.4	18.4	16.6	10	14.4	15	15.6	17.2	15.2	13	14.2	17.8	16.4	15.4	16	14.4	13	5.59
	SCAD	100	100	100	0.6	0.4	0.2	1.0	1.2	0.4	0.4	0.8	0.4	0.2	0.4	.8	0.2	0.2	0.6	0	1	3.088
	MCP	100	100	100	0.6	0.2	0.2	1.0	1.0	0.4	0.4	0.6	0.4	0.2	0.2	0.4	0.2	0.2	0.6	0	1	3.076
n=200	LASSO	100	100	100	15.8	14.6	16.8	14	13.4	15.4	11.6	14.2	14.8	14.4	15	14.4	14	11.2	14.6	10.8	15	5.4
	SCAD	100	100	100	0.2	0.6	1	0.6	0.4	0.8	0.8	1.2	0.2	0.2	0	0.4	0.2	0.4	0.8	0.4	1	3.092
	MCP	100	100	100	0.2	0.6	0.8	0.4	0.4	0.8	0.6	1	0.2	0.2	0	0.4	0.2	0.4	0.8	0.2	0.8	3.08
n=400	LASSO	100	100	100	13.8	13.4	14.8	11.2	12.6	12.8	10.8	13.4	11	12.4	12.4	15.2	11.2	12.8	13.2	12	13.6	5.176
	SCAD	100	100	100	0.4	0	0.2	0.4	0.4	0	0	0	0	0.4	0	0.2	0.2	0.2	0.2	0	0	3.026
	MCP	100	100	100	0.4	0	0.2	0.4	0.2	0	0	0	0	0.4	0	0.2	0.2	0.2	0.2	0	0	3.024

Table 3.2 Comparison of selection percentages (%) of different variables and average model size, with pre whitening.

Sample Size	Method	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20	Avg Model Size	
n=100	LASSO	100	100	100	6.2	7.8	7.6	6	7.8	6.4	6.6	7.4	5.8	5.4	6.4	6.4	7.4	4.8	6	6	6.2	4.112	
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
n=200	LASSO	100	100	100	5.8	3	6	4.4	5.2	4.8	5.6	4.2	7.6	4.4	4.8	3.2	3.4	2.8	5.6	4.4	16	3.932	
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
n=400	LASSO	100	100	100	4.2	2.6	4.6	3.8	3.6	3	2.6	3.4	5	5.2	5.4	3.6	3.4	3.4	4.6	4.8	29	3.922	
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

Table 3.3 Comparison of MC absolute bias and mean square error.

Sample Size	Method	$\hat{\beta}_1(t)$		$\hat{\beta}_2(t)$		$\hat{\beta}_3(t)$	
		Bias	MSE	Bias	MSE	Bias	MSE
n=100	LASSO	0.083	0.033	0.092	0.048	0.112	0.191
	SCAD	0.011	0.025	0.015	0.038	0.022	0.165
	MCP	0.011	0.025	0.015	0.038	0.022	0.165
n=200	LASSO	0.061	0.017	0.069	0.024	0.092	0.109
	SCAD	0.007	0.013	0.008	0.019	0.010	0.091
	MCP	0.007	0.013	0.008	0.019	0.010	0.091
n=400	LASSO	0.047	0.009	0.051	0.013	0.070	0.063
	SCAD	0.004	0.007	0.004	0.010	0.010	0.050
	MCP	0.004	0.007	0.004	0.010	0.010	0.050

The group LASSO estimates (dashed line) have a larger bias which is again expected, as LASSO is known to have a relatively high bias when the magnitude of the regression coefficient is large. The group SCAD (dotted line) and group MCP (dashed-dotted line) estimates have almost identical accuracy and variability as seen from Figure 3.1; they have superimposed on each other and on the true curves represented by solid lines.

To further evaluate the performance of the estimates we calculate the absolute bias and the MC mean square error of the estimates averaged across 100 equally spaced grid points in $[0, 100]$, for all the selection methods and the three sample sizes. This is displayed in Table 3.3. We again observe group SCAD and group MCP method outperforming the group LASSO method, in terms of both absolute bias and mean square error, the performance of the estimators improving with increasing sample size. We compared these mean square errors of the estimates arising from pre-whitening procedure with the same from non pre whitening procedure and found these only to be marginally higher, which is expected due to the uncertainty associated with estimating the covariance matrix. The mean square errors appear to be converging to zero across all the three methods with increase in sample size indicating consistency of the estimators. The simulation results illustrate superior performance of the proposed group SCAD or group MCP based selection method in the context of functional linear concurrent model and are the recommended methods of this article.

3.4 Real Data Applications

In this section, we demonstrate application of our variable selection method in selection of influential time varying predictors in two real data studies. For performing variable selection, we use only the group SCAD and group MCP method along with the initial pre whitening procedure, as the group LASSO method yields a significantly higher false positive rate which is illustrated by our simulations. We first consider a small dietary calcium absorption dataset (3 time varying covariates) with added pseudo covariates as an illustration of our method. Addition of pseudo covariates is a popular way (Wang et al., 2008; Wu et al., 2007; Miller, 2002) of assessing false selection rate in real datasets. Pseudovariates can, therefore, be used effectively for tuning variable selection procedures. We show that our proposed method is able to select the relevant predictors and discard the pseudovariates successfully. Finally, we apply our variable selection method to the fisheries dataset to find out relevant socio-economic drivers influencing fisheries footprint of nations over time.

3.4.1 Study of Dietary Calcium Absorption

We consider the study of dietary calcium absorption in Davis (2002). In this study, the subjects are a group of 188 patients. We have data on calcium absorption ($Y(t)$), dietary calcium intake ($X_1(t)$), BMI ($X_2(t)$) and BSA (Body surface area) ($X_3(t)$) of these patients, at irregular time points between 35 to 64 years of their ages. At the beginning of the study patients aged between 35 to 45 years and subsequent observations were taken approximately every 5 years. The number of repeated measurements for each patient is between 1 to 4. Figure 3.2 displays the individual curves of patients' calcium absorption, calcium intake, BSA, BMI along their ages.

We are primarily interested in finding out which covariates influence calcium absorption profile of the patients. Kim et al. (2018) also investigated the effect of calcium intake on calcium absorption using an additive nonlinear functional concurrent model, and found the effect to be more or less linear while comparing to a functional linear concurrent model. So we use functional

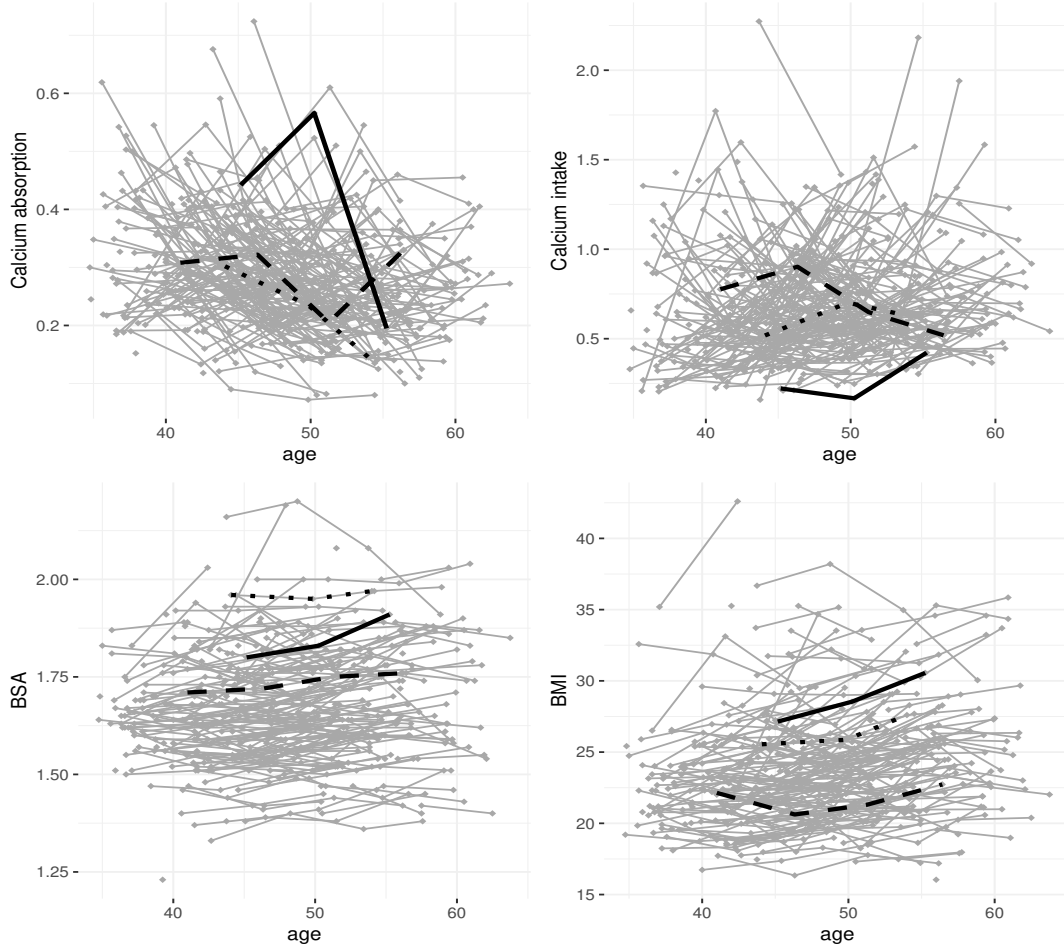


Figure 3.2 Calcium absorption and covariate profiles of patients along their ages.

linear concurrent regression to model the dependence of calcium absorption on calcium intake, BSA and BMI. As data is observed very sparsely and the original covariates might be observed with measurement error we apply FPCA methods (PVE = 95%) as discussed in Section 3.2.3 and get the denoised trajectories $\hat{X}_j(t)$ for $j = 1, 2, 3$. We expect that calcium intake out of the three covariates, will be associated with calcium absorption. To illustrate the selection performance and false positive rate of our variable selection method we further add 15 pseudo covariates by simulating from the following functional model. We generate $X_{ij}(\cdot) \stackrel{iid}{\sim} X_j(\cdot)$ where $X_j(t)$ ($j = 4, 5, \dots, 18$) are given by $X_j(t) = a_j\sqrt{2}\sin(\pi(j-3)t/200) + b_j\sqrt{2}\cos(\pi(j-3)t/200)$,

Table 3.4 Selection Percentages (%) of variables in Calcium absorption Study.

Method	Var 1(Calcium Intake)	Var 2 (BSA)	Var 3(BMI)	Max Var (4-18)
SCAD (FLCM)	100	0	0	0
MCP (FLCM)	100	0	0	0
BIC (LM)	100	100	0	9
Cp (LM)	100	100	2	33
PGEE (ind)	100	0	100	31
PGEE (AR-1)	100	0	100	31

where $a_j \sim \mathcal{N}(0, (2)^2)$, $b_j \sim \mathcal{N}(0, (2)^2)$. So in total, we have 18 covariates, where the first 3 are the denoised original covariates and rest are simulated predictors. Then we apply our variable selection method to $Y(t)$ and $\hat{X}_1(t), \hat{X}_2(t), \hat{X}_3(t), X_4(t), X_5(t), \dots, X_{18}(t)$. We repeat this a large number of times and observe which variables are being selected in each iteration. We expect our variable selection method to pick out the truly influential predictors and ignore the randomly generated functional covariates majority of the time. To illustrate the benefit of using our proposed variable selection method in functional regression model for this particular data we compare its performance to a backward selection method which uses model selection criterion like BIC or Mallows' Cp, under a linear model approach (using an independent working correlation structure), and to a penalized generalized estimating equations (PGEE) procedure (Wang et al., 2012) which was developed to analyze longitudinal data with a large number of covariates. We use the 'PGEE' package in R (Inan and Wang, 2017) for implementing the penalized generalized estimating equations procedure under two different working correlation structure (independent and AR (1)). Table 3.4 illustrates the selection percentage of each of the variables under different methods. We notice that both the proposed group SCAD and group MCP method for FLCM identify calcium intake ($X_1(t)$) as a significant predictor 100% of the time. All other variables including all the pseudo covariates are ignored in 100% of the iterations. On the other hand, BIC, Cp and PGEE exhibit a high false selection percentage for the pseudovariates. The case of selection of BSA and BMI too appears to be over selection. We further investigated this using the hypothesis testing method for FLCM proposed in Chapter 2,

and we found no evidence of BSA and BMI having significant effects on calcium absorption. This demonstrates when the underlying model is functional, use of naive variable selection methods using scalar regression techniques can lead to wrong inference.

As calcium intake is the only significant variable selected by both the proposed methods we want to estimate its effect and also get a measure of uncertainty of our estimate. For this purpose we use a subject-level bootstrap on our original data (no pseudo covariates added) while performing variable selection to come up with estimated regression curve $\hat{\beta}_1(t)$ and a pointwise confidence interval for the effect of calcium intake. This is displayed in Figure 3.3. We notice

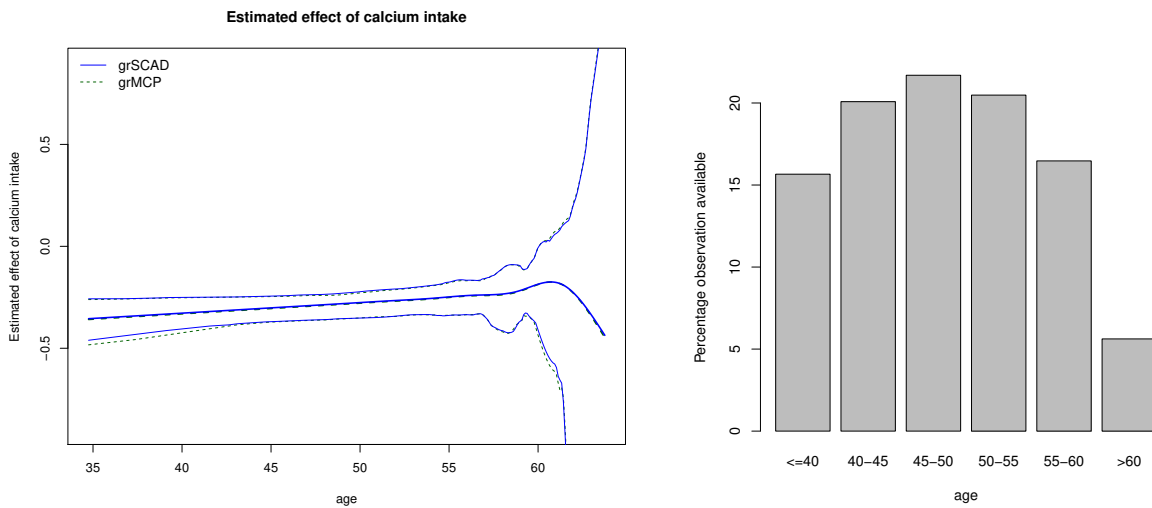


Figure 3.3 Bootstrap estimate and point wise confidence interval of the the effect of calcium intake (left panel) and percentage observation available in different age groups (right panel).

as calcium intake increase calcium absorption should decrease particularly until age 60 years, as $\hat{\beta}_1(t) < 0$ upto this age and the confidence interval strictly lies below zero, which might be due to dietary calcium saturation or due to interaction with some other elements in the body; although the overall magnitude of the effect seems to decrease with age. Above age 60, the estimate appear to have high variability associated with it, which is primarily because we have

very less observations (5.62%) above this mark (illustrated in Figure 3.3). The uncertainty in estimating $\hat{\Sigma}$ using FPCA and the uncertainty due to bootstrap is reflected in its variability. Hence some care should be taken in interpretation of the estimated regression curve beyond 60 years because of such high uncertainty.

3.4.2 Study of Fisheries Footprint

Production of fisheries is a source of protein as well as an economic livelihood across the world. Along with the increasing global population the importance of fish production and consumption has steadily increased through the modern era. Fisheries Footprint is defined as the Global Footprint Network's measure of total marine area required to sustain consumption levels of aquatic production of fish, crustacean (e.g., shrimp), shellfish, and seaweed from captures and aquaculture; so the fisheries footprint basically represents the coastal and marine area required to sustain the amount of seafood products a nation consumes. As pointed out by Longo and Clark (2016) the interaction between marine and social systems calls for further sociological analysis.

Over the last two decades, social scientists have accomplished much in advancing scholarly knowledge on the social drivers of ecological impact at a macro-scale. Such work is essential, as ecological problems are becoming increasingly interlinked and severe at a global or planetary scale (Steffen et al., 2011). Over time, for example, economic development, population structuring (e.g., urbanization or age structure), trade relations, and technological change are shown to affect measures of environmental impact across nations over time (Clark and Longo, 2019; Jorgenson and Clark, 2010; York et al., 2003). This body of literature centers on the ecological affects of globalization and modernization, under the socio-structural parameters of a capitalist economy. There is still much debate over the impacts of industrial and agricultural modernization on ecosystems. For example, development and resource economics literature (World Bank, 2007) advocate for the utilization of innovation and techno-improvements to improve marine system sustainability and food security (Valderrama and Anderson, 2010), while, on the other hand,

environmental sociologists demonstrate that such innovation, chiefly aquaculture, does not displace the deleterious ecological impacts of capture fisheries (Longo et al., 2019). Nevertheless, despite such progress, fisheries footprint remains an understudied metric, and its drivers are less understood in social research (Clark et al., 2018; Jorgenson et al., 2005).

The goal of this study, therefore, is to identify the relevant socio-economic drivers such as levels of economic development, population size, and transformations in food-system dynamics that influence fisheries footprint of nations over time and also to capture their time varying effects. Data for this study is collected from the World Bank, Fish StatJ of UN FAO, and Ecological Footprint Network for years between 1970-2009, across 136 nations. The main dependent variable of interest in this study is fisheries footprint. Figure 3.4 displays the fisheries footprint of the nations over the study years in the log scale. Fisheries footprint of three representative nation are plotted using solid, dashed and dotted lines. To capture the trend over the years, we plot the mean fisheries footprint of the nations along with their point wise 95% confidence interval. This is displayed in Figure 3.5. We notice an overall upward trend as well as heterogeneity across years. There are 20 independent time varying covariates in the study, broadly covering various sectors of population dynamics (e.g. population density, urban population, total population, working age population percentage etc), agriculture (e.g. tractor, agriculture value added etc), food consumption and other fisheries variables (e.g. meat consumption, aquaculture production tons etc), international relations (e.g. food export as percentage of merchandise, FDI inflow etc), and economy (e.g. GDP per capita at constant U.S. dollar , trade percentage of GDP etc). The full list of the variables are given in Table 3.5. The predictors here are also time varying and can be expected to have dynamic effects on fisheries footprint. Generally panel data methods like fixed effects or random effects modeling is used (Torres-Reyna, 2007; Clark and Longo, 2019) for analyzing such data where the effect the covariates are taken to be constant, since we are interested in dynamic effects of the covariates, the FLCM can be seen as a generalization of this approach with time varying effects of predictors. Therefore we use a functional linear concurrent regression model (3.1) discussed in this article to model the dynamic effects of the socio-economic

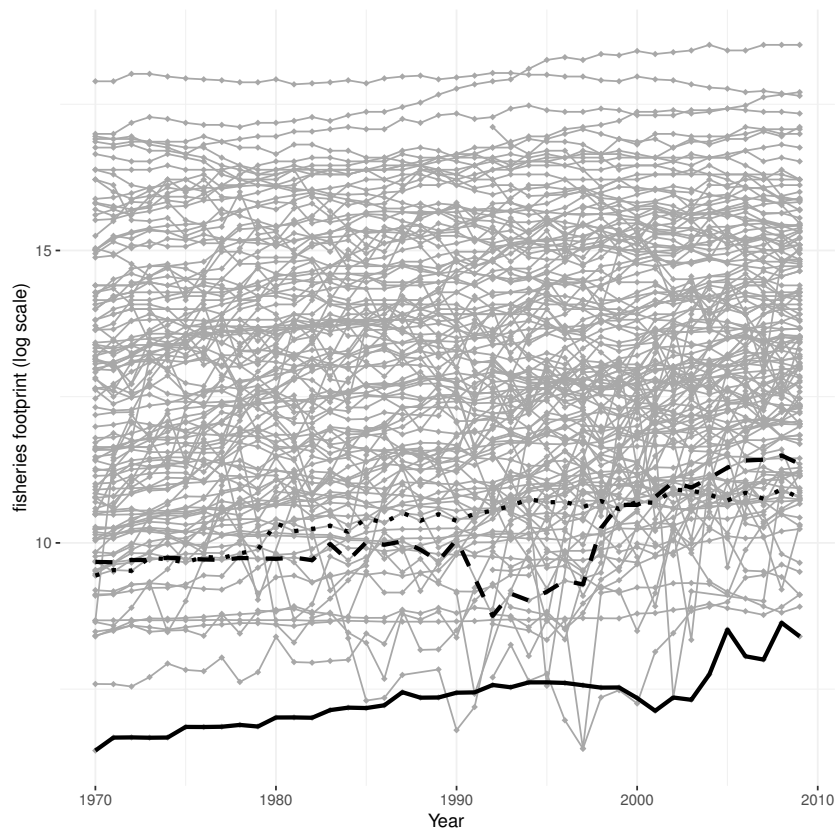


Figure 3.4 Fisheries footprint of the nations over year 1970-2009.

Table 3.5 List of covariates in the Fisheries Footprint Study.

Predictor Variables in the Fisheries Footprint study	
agriculture value added	aquaculture production tons
arable land hectares	arable land pct
exportsofgoodsandservicesofgdpn	fao livestock
FDI inflow	foodexportsofmerchandiseexprtst
foodimportsofmerchandiseimprtst	gdp pc 2010
manufacturing value pctGDP	meat consumption FAO
population 15_64 pct	population density
populationtotalsppoptotl	services value growth pct
tractors	trade pct GDP
urban pop	urban pop pct

Heterogeneity across years

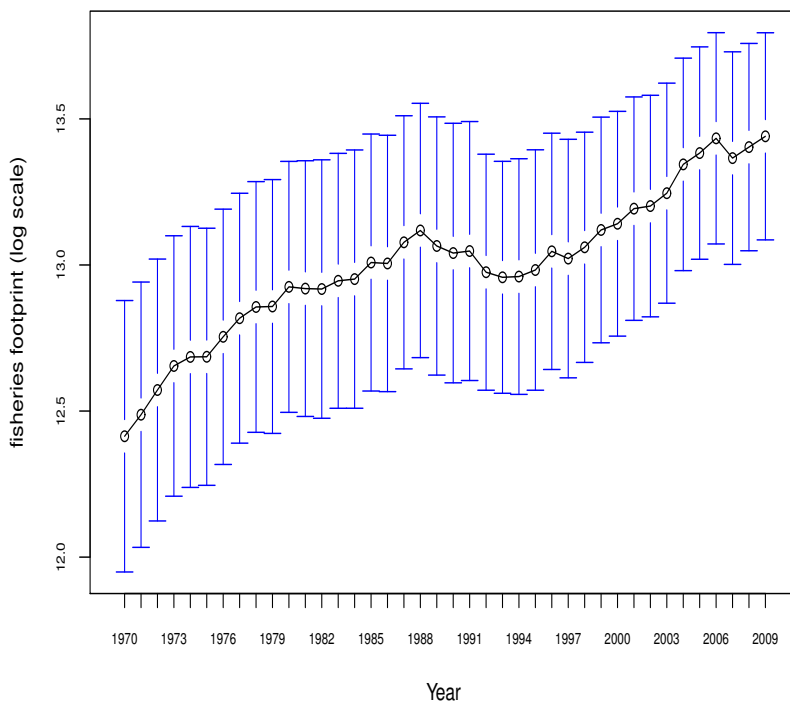


Figure 3.5 Mean fisheries footprint along with their 95% confidence interval.

predictors on fisheries footprint. The predictors in their original scale are also very large in magnitude, therefore converted into log scale; the covariates observed as percentages are used without any conversion. Before applying our variable selection method all the covariates are pre-processed using FPCA methods ($PVE = 95\%$) as discussed in Section 3.2.3.

We use the prewhitening procedure discussed in Section 3.2.2 and apply our proposed variable selection method. Out of 20 covariates group SCAD and group MCP both identify GDP per capita and urban population as the two significant predictors. Gross domestic product (GDP) is a measure of the market value of all the final goods and services produced in a specific time period, and GDP per capita is a measure of a country's economic output adjusting for its number of people. As the major economic indicator GDP per capita is associated with primary aspects of

economic growth, consumer behavior, trade and therefore is a key indicator of fisheries footprint of nations over time. Furthermore, GDP per capita is a common metric in extant social science research to operationalize the extent to which a nation is successfully developing according to the standards of the world, capitalist economy (Dietz and Jorgenson, 2013). Figure 3.6 shows the estimated regression curve for GDP per capita obtained by applying the group SCAD selection method. The estimate from the group MCP method is similar. We observe the net effect of GDP per capita on fisheries footprint to be positive and linear although the magnitude of the effect has decreased over time.

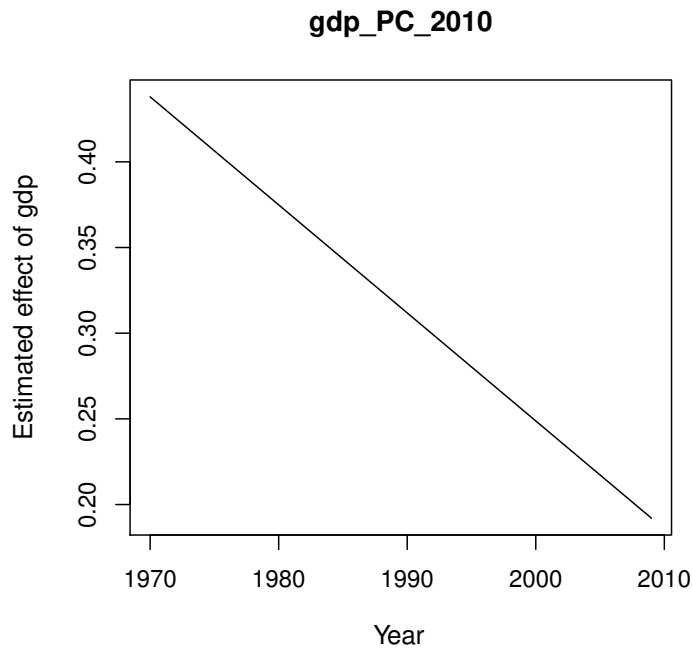


Figure 3.6 Estimate of the linear concurrent effect of GDP per capita on fisheries footprint (SCAD).

Urban population being the key market, also plays a crucial role in the total seafood consumption of nations and therefore influences fisheries footprint. Change in urban population reflects urbanization and urbanization have important effects on food security and farming

(Satterthwaite et al., 2010; Cohen and Garrett, 2010). Figure 3.7 shows the estimated regression curve illustrating the effect of urban population on fisheries footprint. Here also we notice the net effect of urban population on fisheries footprint to be positive, although the effect appears to be more or less constant with a very marginal decrease (in log scale). Here, it is important to

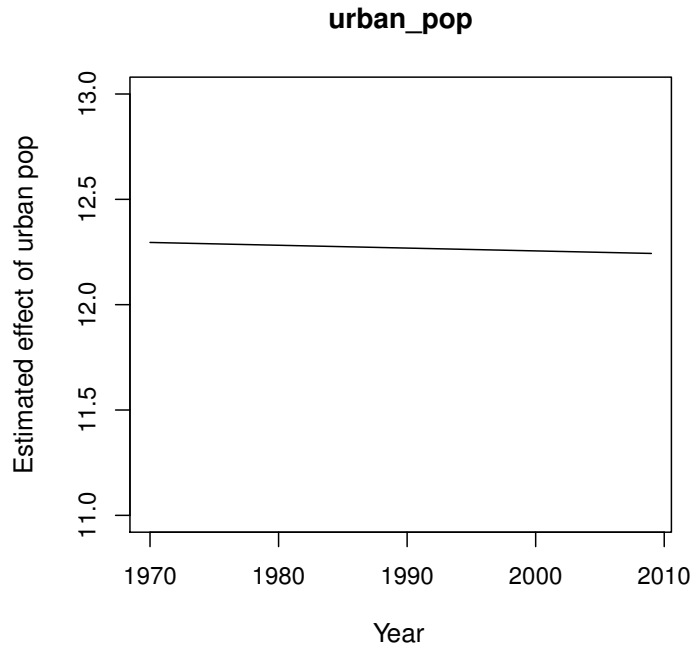


Figure 3.7 Estimate of the linear concurrent effect of urban population on fisheries footprint (SCAD).

note that fisheries footprint represents the metabolic potential of an ecosystem to reproduce itself ecologically. According to the Food and Agriculture Organization of the United Nations (FAO, 2016), about 58 percent of global fish stocks are currently fully exploited, and about 55 percent of ocean territory (conservatively) was subjected to industrial fishing in the past year (Kroodsma et al., 2018). Thus, there is declining metabolic potential for the expansion of capture fisheries, which likely helps to account for why variable effects were stronger in earlier, more ecologically productive decades. Both these variables are important in the sense they

represent the primary indicators in economics, food consumption, population dynamics, trade, etc; which directly interact with a nation’s need for seafood and therefore should influence fisheries footprint. It is therefore not surprising that countries having high GDP per capita and/or high urban population e.g., United States, Australia, Singapore, etc also have a high fisheries footprint. In Figure 3.8 we display the fisheries footprint, GDP per capita and urban population profile of the three representative countries mentioned earlier. We notice the overall trend in the fisheries footprint profile can be described well by their GDP per capita and urban population profile, both of which were shown to have a positive effect on fisheries footprint.

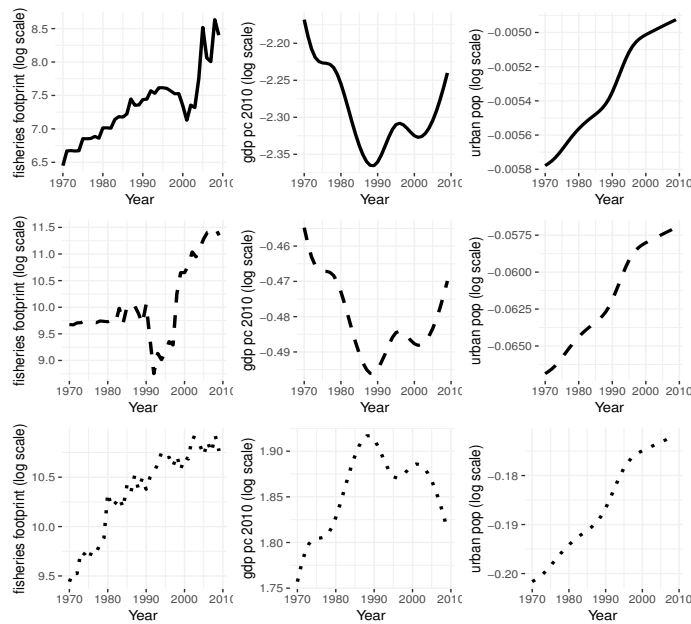


Figure 3.8 Profile of fisheries footprint, GDP per capita and urban population of three representative countries.

Remark 3: We have successfully applied our proposed variable selection method to find out the relevant time-varying predictors and their time-varying effects on fisheries footprint. It is very plausible that there might be country or region specific effects on fisheries footprint as revealed in study by Clark and Longo (2019), and one might be interested in estimating these effects. The proposed variable selection method for FLCM can be extended to handle such region specific

effects in its existing form.

Remark 4: We have considered concurrent effects of the predictors while some of the predictors might have lagged effects on fisheries footprint. For example in an economic crisis or recession, the predictors could very likely have reverberating impacts on development for a few years. As invested capital takes time to flow through the economy, considering such lagged effects would be interesting. We applied our variable selection method with lagged predictors present along with the original predictors (with lag window = 1, 3). We found out that for lag one, the proposed FMCP and FSCAD method select almost identical models with the FSCAD method selecting ‘services value growth pct’ as an additional variable. Considering a lag window of three years, the FSCAD method selects urban population lag instead of urban population while the FMCP method additionally selects ‘aquaculture production tons’ and ‘services value growth pct lag’ as influential covariates. These results indicate some of the predictors could have reverberating impacts and a more general framework like the historical functional regression model (Malfait and Ramsay, 2003) might be more suitable to model past effects of covariates on the response at current time point.

3.5 Discussion

In this article, we have proposed a variable selection method in functional linear concurrent regression extending the classically used penalized variable selection methods like LASSO, SCAD, and MCP. We have shown the problem can be addressed as a group LASSO and their natural extension group SCAD or group MCP problem. We have used a pre-whitening procedure to take into account the temporal dependence present within functions and through numerical simulations, have illustrated our proposed selection method with group SCAD or group MCP penalty can select the true underlying variables with high accuracy and has minuscule false positive and false negative rate even when data is observed sparsely, is contaminated with measurement error and the error process is highly non-stationary. We have illustrated usefulness of the proposed method by applying to two real datasets: the dietary calcium absorption study

data and the fisheries footprint data in identification of the relevant time-varying covariates. In this article we have used a resampling subject based bootstrap method to measure uncertainty of the regression functions estimates, theoretical properties corresponding to such bootstrap is something we would like to explore more deeply in the future.

There are many interesting research directions this work can head into. In real data, the dynamic effects of the predictors might always not be linear. In future, we would like to extend our variable selection method to nonparametric functional concurrent regression model (Maity, 2017), which is a more general and flexible model to capture complex relationships present between the response and covariates. As mentioned earlier it would be also of interest to consider the lagged effects of covariates through a more general historical functional regression model (Malfait and Ramsay, 2003).

In developing our method we assumed the covariates to be independent and identically distributed. In many cases this might not be a reasonable assumption. For example, in the fisheries footprint data some countries could be very similar and form clusters, on the other hand, they might not be even independent with the interplay of economies and other variables among nations. Even if the covariates are not independent over subjects, the variable selection criterion proposed in this article can still be used in practice as a penalized least square method. The heterogeneity present among the subjects can be addressed using interaction effect of covariates with regions, which can be clustered based on the level of affluence. This can be done similarly as in Clark and Longo (2019). Alternatively, one can also use subject specific functional random effects for covariates, especially if one is interested in individual specific trajectories. Functional linear mixed model (Liu et al., 2017) might be an appropriate choice in such situations. Extending the proposed variable selection method to such general functional regression models would be an extension of this work and remain an area for future research.

Software

Illustrations of implementation of our method using R are available at GitHub (https://github.com/rahulfrodo/FLCM_Selection).

Chapter 4

Variable Selection in Nonparametric Functional Concurrent Regression

4.1 Introduction

Function on function regression refers to the class of regression models in functional data analysis where both the response variable and the covariates are functions over any continuous domain such as time. Functional concurrent regression model is a special class of function on function regression, where the value of the response at the current time point depends only on the values of predictors at that specific time point. Functional linear concurrent model (Ramsay and Silverman, 2005) is the most commonly used concurrent regression model, where the dependence of the response on the covariates is assumed to be linear and is modeled using smooth univariate regression functions. These regression functions can capture the time varying effect of the covariates on the response. Estimation and inference in functional linear concurrent regression and closely related varying coefficient model (Hastie and Tibshirani, 1993) have been widely studied in the literature.

Although this is a useful model and finds applications in various branches of longitudinal data analysis, they are restrictive in the sense of assuming a linear relationship between response and

predictors. The nonparametric functional concurrent model (NPFCM) (Maity, 2017) overcomes the assumption of linear dependence between response and covariates; which may not be true in many real data applications, by specifying a nonparametric relationship which is more general and flexible for modeling purposes. Multiple methods exist in the literature for estimation and inference in functional concurrent regression models using smoothing splines (Kim et al., 2018), Gaussian process regression (Shi et al., 2005), local kernel smoothing techniques (Jiang et al., 2011) among many others. Analogous to scalar regression, when there are a large number of covariates present, one might be interested in only the true set of influential predictors and hence arises the need to perform variable selection.

A number of variable selection methods exist for scalar on function regression (Gertheiss et al., 2013; Fan et al., 2015) and function on scalar regression (Chen et al., 2016) models. Literature for variable selection in functional concurrent regression model, on the other hand, is relatively sparse. Goldsmith and Schwartz (2017) proposed a variable selection method in the functional linear concurrent model using a variational Bayes approach. A variable selection method for the functional linear concurrent regression model was proposed in Chapter 3 extending the classically used penalized variable selection methods. Both these methods are for the functional linear concurrent model, which as we mentioned, might not capture the true relationship between the response and predictors.

In this article, we propose a variable selection method for nonparametric functional concurrent regression, extending the classically used variable selection methods like LASSO (Tibshirani, 1996), SCAD (Fan and Li, 2001) and MCP (Zhang, 2010). To our knowledge, this is the first work in developing variable selection method for the nonparametric functional concurrent regression model. We have followed the approach in Chapter 3 and shown that for nonparametric functional concurrent regression the variable selection problem reduces to a group LASSO (Yuan and Lin, 2006), and their natural extension a group SCAD or a group MCP problem. Our model is very general in the sense it models the dependence of the functional response on the functional covariates by unknown nonparametric functions and requires no distributional assumption on

the data. Through simulations, we illustrate our variable selection method can successfully identify the true underlying predictors with negligible false positive rate and false negative rate and gives high out of sample prediction accuracy, even when the data is sparsely observed and contaminated with measurement error. We have also demonstrated application of our method in two real data problems: a dietary calcium absorption study (Davis, 2002) and a bike sharing data, in the identification of influential time varying predictors.

The rest of this article is organized in the following way. In Section 4.2 we present our modeling framework, set up the nonparametric functional concurrent regression model and illustrate our variable selection method. In Section 4.3 we present a simulation study to evaluate the performance of our method. In Section 4.4 we apply our variable selection method to find out the influential predictors in a dietary calcium absorption study and a bike sharing study, and summarize our findings. We conclude in Section 4.5 with a brief discussion on the contributions of our method and some possible extensions of this work.

4.2 Methodology

4.2.1 Modeling Framework and Variable Selection Method for NPFM

We assume that the observed data for the i th subject is $\{Y_i(t), X_{i1}(t), X_{i2}(t), \dots, X_{ip}(t)\}$ ($i = 1, 2, \dots, n$), where $Y(\cdot)$ represents the functional response and $X_1(\cdot), X_2(\cdot), \dots, X_p(\cdot)$ are the corresponding functional covariates. In practice, we observe both the response and the predictors only on finitely many time points over some closed and bounded interval. As in Chapter 2 and 3, in developing our method we assume the covariates and the response are observed on a dense and regular grid of points $S = \{t_1, t_2, \dots, t_m\} \subset S = [a, b]$ for some $a, b \in \mathbb{R}$, and the covariates are observed without any measurement error. We later illustrate how our method can be easily extended to accommodate more realistic scenarios where the covariates are contaminated with measurement error and are observed on irregularly spaced grid. We consider the following

nonparametric functional concurrent regression model,

$$Y_i(t) = \mu_Y(t) + \sum_{j=1}^p F_j\{X_{ij}(t), t\} + \epsilon_i(t), \quad \text{for } i = 1, \dots, n, \quad (4.1)$$

where $\mu_Y(t)$ is the functional intercept and $F_j\{X_j(t), t\}$ are smooth functions on $\mathbb{R} \times S$ capturing the concurrent effect of covariate $X_j(\cdot)$. If $E[F_j\{X_j(t), t\}] = 0$, the functions $F_j(\cdot, \cdot)$ s are identifiable and then $\mu_Y(\cdot)$ represents the marginal mean of the response curves. The covariates $X_{ij}(\cdot)$ are assumed to be independent and identically distributed (i.i.d.) copies of $X_j(\cdot)$ ($j = 1, 2, \dots, p$), where $X_j(\cdot)$ is a smooth stochastic processes. We further assume $\epsilon_i(\cdot)$ are i.i.d. copies of $\epsilon(\cdot)$ which is a mean zero stochastic process with unknown covariance structure. We model the intercept function $\mu_Y(t)$ in terms of some known univariate basis expansion as $\mu_Y(t) = \sum_{l=1}^{p_0} \theta_{\mu,l} B_{\mu,l}(t) = \mathbf{B}_\mu(t)^T \boldsymbol{\theta}_\mu$, where $\mathbf{B}_\mu(t) = [B_{\mu,1}(t), B_{\mu,2}(t), \dots, B_{\mu,p_0}(t)]^T$ and $\boldsymbol{\theta}_\mu = (\theta_{\mu,1}, \theta_{\mu,2}, \dots, \theta_{\mu,p_0})^T$ is a vector of unknown coefficients. In this article, we use B-spline basis functions for this purpose, however, other basis functions can be used as well. For modeling bivariate function $F_j(\cdot, \cdot)$ we express them in terms of bi-variate basis expansion using tensor product of univariate basis functions (McLean et al., 2014; Kim et al., 2018). Now suppose for $j = 1, \dots, p$, $\{B_{X_j,k}(x)\}_{k=1}^{K_j}$ and $\{B_{T_j,\ell}(t)\}_{\ell=1}^{L_j}$ be a set of known basis functions over x and t respectively. Then we model $F_j(\cdot, \cdot)$ using tensor product of these two basis functions as $F_j\{X_{ij}(t), t\} = \sum_{k=1}^{K_j} \sum_{\ell=1}^{L_j} \theta_{j,k,\ell} B_{X_j,k}\{X_{ij}(t)\} B_{T_j,\ell}(t) = \mathbf{Z}_{ij}(t)^T \boldsymbol{\theta}_j$, where we denote the $K_j L_j$ dimensional vector of $B_{X_j,k}\{X_{ij}(t)\} B_{T_j,\ell}(t)$'s as $\mathbf{Z}_{ij}(t)$, and $\boldsymbol{\theta}_j$ is the vector of unknown basis coefficients $\theta_{j,k,\ell}$'s. For our variable selection purpose we don't include 1 as a basis in $\{B_{X_j,k}(x)\}_{k=1}^{K_j}$. So we are able to estimate the joint effect of x, t on $Y(t)$ adjusted for the main effect of t , which gets included in the functional intercept $\mu_Y(t)$. Therefore w.l.g by modeling $F_j\{x, t\}$ we basically model the main effect of x and interaction effect of x, t . Based on the basis expansions model (4.1) be reformulated as,

$$Y_i(t) = \mathbf{B}_\mu(t)^T \boldsymbol{\theta}_\mu + \sum_{j=1}^p \mathbf{Z}_{ij}(t)^T \boldsymbol{\theta}_j + \epsilon_i(t). \quad (4.2)$$

Since we assume data is observed on a dense and regular grid of points $S = \{t_1, t_2, \dots, t_m\}$ the corresponding linear model to the above functional model (2) can be written in stacked form as,

$$\mathbf{Y}_i = \mathbb{B}_\mu \boldsymbol{\theta}_\mu + \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j + \boldsymbol{\epsilon}_i, \quad (4.3)$$

where $\mathbf{Y}_i = (Y_i(t_1), Y_i(t_2), \dots, Y_i(t_m))^T$, $\mathbb{Z}_{ij} = (\mathbf{Z}_{ij}(t_1), \mathbf{Z}_{ij}(t_2), \dots, \mathbf{Z}_{ij}(t_m))^T$, and $\boldsymbol{\epsilon}_i = (\epsilon_i(t_1), \epsilon_i(t_2), \dots, \epsilon_i(t_m))^T$. Now a covariate $X_j(t)$ taking a value x influences response $Y(t)$ unless $F_j\{x, t\} = \alpha_j(t) \forall x$, which in terms of the basis representation we use, is equivalent to the condition $\boldsymbol{\theta}_j = \mathbf{0}$. Therefore, we propose to minimize the following penalized residual sum of squares criterion to perform variable selection.

$$\begin{aligned} R(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\| \mathbf{Y}_i - \mathbb{B}_\mu \boldsymbol{\theta}_\mu - \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j \right\|_2^2 + \lambda mn \sum_{j=1}^p (\boldsymbol{\theta}_j^T \boldsymbol{\theta}_j)^{1/2} \\ &= \sum_{i=1}^n \left\| \mathbf{Y}_i - \mathbb{B}_\mu \boldsymbol{\theta}_\mu - \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j \right\|_2^2 + \lambda mn \sum_{j=1}^p \|\boldsymbol{\theta}_j\|_2 \\ &= \sum_{i=1}^n \left\| \mathbf{Y}_i - \mathbb{B}_\mu \boldsymbol{\theta}_\mu - \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j \right\|_2^2 + mn \sum_{j=1}^p P_{LASSO, \lambda}(\|\boldsymbol{\theta}_j\|_2). \end{aligned} \quad (4.4)$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \dots, \boldsymbol{\theta}_p^T)^T$. This minimization problem is exactly same as performing a group LASSO (Yuan and Lin, 2006), where the grouping is introduced by covariates Z_j and depending on whether $\boldsymbol{\theta}_j = \mathbf{0}$ or not, $X_j(\cdot)$ will be either present or absent in the model. The tuning parameter for this minimization are the penalty parameter λ and number of basis functions K_j, L_j which controls the smoothness of the function $F_j(\cdot, \cdot)$ in x and t directions respectively. Instead of directly using penalties on the smoothness of the functions, we restrict the number of basis functions to be small which ensures the resulting functions are smooth. We now extend this group LASSO formulation to two non convex penalties, as it was illustrated in Chapter 3 that group LASSO can have a very high false positive rate for variable selection in functional linear concurrent regression. It was also noticed that they produce biased estimates which in turn can affect the prediction accuracy of model. The non convex penalties SCAD (Fan and Li,

2001) and MCP (Zhang, 2010) on the other hand are known to produce sparser solutions in scalar regression literature and also enjoy the so called oracle property in which they behave like oracle MLE asymptotically. Subsequently we propose a group SCAD and group MCP method extending the group LASSO formulation for performing variable selection in nonparametric functional concurrent regression.

Group SCAD Method

We perform variable selection and obtain estimate of $\boldsymbol{\theta}$ in the following way,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|Y_i - \mathbb{B}_\mu \boldsymbol{\theta}_\mu - \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j\|_2^2 + mn \sum_{j=1}^p P_{SCAD,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2), \quad (4.5)$$

where $P_{SCAD,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2)$ is defined in the following way:

$$P_{SCAD,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2) = \begin{cases} \lambda \|\boldsymbol{\theta}_j\|_2 & \text{if } \|\boldsymbol{\theta}_j\|_2 \leq \lambda. \\ \frac{\lambda \phi \|\boldsymbol{\theta}_j\|_2 - .5(\|\boldsymbol{\theta}_j\|_2^2 + \lambda^2)}{\phi - 1} & \text{if } \lambda < \|\boldsymbol{\theta}_j\|_2 \leq \lambda \phi. \\ .5\lambda^2(\phi + 1) & \text{if } \|\boldsymbol{\theta}_j\|_2 > \lambda \phi. \end{cases}$$

Group MCP Method

For Group MCP method we estimate $\boldsymbol{\theta}$ as,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}_j, j=1,2,\dots,p}{\operatorname{argmin}} \sum_{i=1}^n \|Y_i - \mathbb{B}_\mu \boldsymbol{\theta}_\mu - \sum_{j=1}^p \mathbb{Z}_{ij} \boldsymbol{\theta}_j\|_2^2 + mn \sum_{j=1}^p P_{MCP,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2), \quad (4.6)$$

where $P_{MCP,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2)$ is defined as :

$$P_{MCP,\lambda,\phi}(\|\boldsymbol{\theta}_j\|_2) = \begin{cases} \lambda \|\boldsymbol{\theta}_j\|_2 - \frac{\|\boldsymbol{\theta}_j\|_2^2}{2\phi} & \text{if } \|\boldsymbol{\theta}_j\|_2 \leq \lambda \phi. \\ .5\lambda^2 \phi & \text{if } \|\boldsymbol{\theta}_j\|_2 > \lambda \phi. \end{cases}$$

Remark 1: For the selection of the tuning parameter K_j , L_j and the penalty parameter λ we

use the Extended BIC (EBIC) (Chen and Chen, 2008) criterion of a corresponding Gaussian likelihood. Although we make no distributional assumption this has shown good performance as illustrated in our simulation studies. Some other criteria or data-driven methods can be used depending on the problem. For computational simplicity, while fitting our model we use $K_j = K, L_j = L$ for $j = 1, \dots, p$, which is a standard assumption for computational tractability and assumes all the nonparametric effects have more or less same level of smoothness in both variables respectively. For the tuning parameter ϕ we use the values 4 for SCAD and 3 for MCP, which were proposed by the original authors of these methods. For model fitting purposes we use ‘grpreg’ package (Breheny, 2019) in R.

4.2.2 Extension to Sparse and Noisy Data

More generally, we consider the case where data is observed on an irregular and sparse grid and covariates are possibly contaminated with measurement error. For longitudinal data, this is most often the scenario. Here we have response $\{(Y_i(t_{ij}), t_{ij}), j = 1, 2, \dots, m_i\}$ and the observed covariates given by $\{(U_1(t_{1ij}), t_{1ij}), j = 1, 2, \dots, m_{1i}\}, \{(U_2(t_{2ij}), t_{2ij}), j = 1, 2, \dots, m_{2i}\}, \dots, \{(U_p(t_{pij}), t_{pij}), j = 1, 2, \dots, m_{pi}\}$. We denote $U_k(t_{kij})$ s, ($k = 1, 2, 3, \dots, p$) by U_{ijk} which represent the observed covariate with measurement error i.e., we have $U_{ijk} = X_k(t_{kij}) + e_{ijk}$ for $i = 1, 2, \dots, n, j = 1, 2, \dots, m_{ki}$ and $k = 1, 2, \dots, p$. The measurement error e_{ijk} are assumed to be white noises with zero mean and variance σ_k^2 . In sparse data set up, a generally assumption (Kim et al., 2018) is, although individual number of observations m_i is small, $\bigcup_{i=1}^n \bigcup_{j=1}^{m_i} t_{ij}$ is dense in $[0, T]$. Then using FPCA methods (Yao et al., 2005) it is possible to reconstruct the original curves from the observed sparse and noisy curves, by estimating the eigenvalues and eigenfunctions corresponding to the original curves. For the prediction of the scores, we can use PACE method of Yao et al. (2005). Then these estimates can be put together using Karhunen-Loève expansion (Karhunen, Loeve 1946) to get estimates $\hat{X}_{ik}(\cdot)$ of the true curves $X_{ik}(\cdot)$ as $\hat{X}_{ik}(t) = \hat{\mu}_k(t) + \sum_{s=1}^S \hat{\zeta}_{isk} \hat{\psi}_{sk}(t)$, where the number of eigenfunctions S to use is typically chosen such percent of variance explained (PVE) exceeds some specified value. So for sparse data observed on an irregular grid and contaminated with measurement

error, we use $\{Y_i(t_{ij}), \hat{X}_{i1}(t_{ij}), \hat{X}_{i2}(t_{ij}), \dots, \hat{X}_{ip}(t_{ij})\}_{i=1}^n$ as our original data and use this data for performing variable selection. Our simulations show good performance in terms of selection accuracy and out of sample prediction performance using this strategy and henceforth throughout the article we pre-process the observed covariates before applying our variable selection method.

4.3 Simulation Study

4.3.1 Simulation Set Up

In this section, we investigate the performance of the proposed variable selection method using simulation study. We evaluate our method in terms of selection accuracy and out of sample prediction performance. We also present a comparison of our nonparametric concurrent variable selection method with variable selection method in functional linear concurrent model (FLCM) proposed in Chapter 3. To this end, we consider the following two scenarios.

Scenario A : We generate data from the model,

$$Y_i(t) = \mu_Y(t) + \sum_{j=1}^{20} F_j\{X_{ij}(t), t\} + \epsilon_i(t), \text{ for } i = 1, \dots, n, \quad t \in [0, 100],$$

where we have $\mu_Y(t) = 1 + 2(t/100) + (t/100)^2$, $F_1\{x, t\} = 2x(t/100)$, $F_2\{x, t\} = 5x \sin(\pi t/100)$, $F_3\{x, t\} = 8 \sin(xt/100)$ and $F_j\{x, t\} = 0$ for $j = 4, 5, 6, \dots, 20$, i.e., the last 17 covariates do not have any effect on the response. The original covariates $X_{ij}(\cdot) \stackrel{iid}{\sim} X_j(\cdot)$, where $X_j(t)$ ($j = 1, 2, \dots, 20$) are given by $X_j(t) = a_j\sqrt{2}\sin(\pi jt/400) + b_j\sqrt{2}\cos(\pi jt/400)$, where $a_j \sim \mathcal{N}(0, (2)^2)$, $b_j \sim \mathcal{N}(0, (2)^2)$. Moreover following the discussion in section 2.2.5 we assume that covariates $X_{ij}(t)$ are observed with measurement error, i.e., we observe $U_{ij}(t) = X_{ij}(t) + \delta_j$, where $\delta_j \sim \mathcal{N}(0, .6^2)$. The error process $\epsilon_i(t)$ is generated as follows;

$$\epsilon_i(t) = \xi_{i1}\cos(t) + \xi_{i2}\sin(t) + N(0, 1^2 I_{m_i}),$$

where $\xi_{i1} \stackrel{iid}{\sim} \mathcal{N}(0, .5^2)$ and $\xi_{i2} \stackrel{iid}{\sim} \mathcal{N}(0, 0.75^2)$. The response $Y_i(t)$ and noisy covariate $U_{ij}(t)$ s are observed sparsely in random m_i points in S , where S is the set of $m = 81$ equidistant time points in $[0, 100]$ and $m_i \stackrel{iid}{\sim} Uniform\{30, 31, \dots, 41\}$. Three sets of sample size $n \in \{200, 400, 600\}$ are considered. For each set 80% of data is used for model fitting and rest for validation.

Scenario B : We generate data from the model,

$$Y_i(t) = \mu_Y(t) + \sum_{j=1}^{20} F_j\{X_{ij}(t), t\} + \epsilon_i(t), \text{ for } i = 1, \dots, n, \quad t \in [1, 101],$$

where $\mu_Y(t) = 1 + 2(t/100) + (t/100)^2$, $F_1\{x, t\} = 2x(t/100)$, $F_2\{x, t\} = 5x \sin(\pi t/100)$, $F_3\{x, t\} = 8 \sin(xt/100) - (16(100/t)^2) ((-12t/100) \cos(12t/100) + \sin(12t/100))$ and $F_j\{x, t\} = 0$ for $j = 4, 5, 6, \dots, 20$. The rest of the design is exactly same as of scenario A. This scenario is mainly used to compare the performance of non-parametric variable selection method proposed in this article to that of variable selection method in FLCM. For this purpose we consider sample size $n = 200$. For both the scenarios use 500 generated data sets from the specified models to asses the performance of the proposed variable selection method.

4.3.2 Simulation Results

Scenario A:

We first asses the selection accuracy of the proposed variable selection method. As the covariates are observed sparsely and with measurement error, we apply FPCA method as discussed in Section 4.2.2 and obtain the denoised curves $\hat{X}_{ij}(t)$ before applying our variable selection method. We obtain the selection percentage of each variable for each of the three selection methods discussed in Section 4.2, and for the three sample sizes $n = 160, 320, 480$. This is displayed in Table 4.1. We observe all the three selection methods (group LASSO, group SCAD, group MCP) pick out the nonparametric effect of $X_1(\cdot), X_2(\cdot), X_3(\cdot)$; 100% of the times. The group lasso method, as expected has a high false selection percentage ranging from 0.4% to as high as 36% for some variables. The group SCAD and group MCP method, on the other hand, show zero false positive rate across all the sample sizes for our simulation, for all the variables. In other words, they are able to identify the true model 100% of the time. The average model sizes for

Table 4.1 Comparison of selection percentages (%) of different variables and mean model size, Scenario A.

Sample Size	Method	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10	Var11	Var12	Var13	Var14	Var15	Var16	Var17	Var18	Var19	Var20	Avg Model Size	
n=160	LASSO	100	100	100	36.0	27.2	25.8	19.8	18.4	13.6	10.8	9.6	7.2	6.4	4.2	4.6	4.2	3.2	1.6	2.4	0.14	4.964	
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
n=320	LASSO	100	100	100	32.4	23.0	22.0	18.2	11.4	9.6	8.8	7.2	7.4	2.8	2.6	3.8	2.0	.8	1.0	0.4	0.4	0.4	4.538
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
n=480	LASSO	100	100	100	31.2	26.2	20.4	16.2	13.4	8.8	8.8	6.6	4.6	4.8	2.2	3.2	0.8	2.0	0.6	0.4	0.4	0.4	4.506
	SCAD	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
	MCP	100	100	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3

each scenario are also given in Table 4.1. The group lasso method produces an average model size in the range of 4.5 – 4.96 for the three different sample sizes with performance getting better with increasing sample size. The group SCAD and group MCP method both produce mean model size 3 on account of successfully identifying the true model each time. Next, we compare the methods in terms of their out of sample prediction performance. For this purpose, we define a criterion similar to R-square based on the median of squared deviations of the predicted curves from the actual curves. For normalizing the measure we use the median of the squared deviations of the actual curve from their mean. The measure is defined as follows,

$$R_{Med}^2 = 1 - \frac{\sum_{i=1}^n \text{Median}_{j=1,2,\dots,m_i} \{Y_i(t_{ij}) - \hat{Y}_i(t_{ij})\}^2}{\sum_{i=1}^n \text{Median}_{j=1,2,\dots,m_i} \{Y_i(t_{ij}) - \bar{Y}(\dots)\}^2}$$

We tabulate the quantiles and 10, 90-th percentile of this out of sample R-Square criterion based on our simulation study. This is displayed in Table 4.2. We notice for all the three methods, the median of out of sample R_{Med}^2 is around 98% – 99% and increasing with training sample size. This indicates a satisfactory out of sample prediction performance of all the three variable selection methods. This simulation study illustrates that the proposed group SCAD and group MCP selection method for NPFM can not only identify the underlying model successfully but also yield high out of sample prediction accuracy.

Table 4.2 Quantiles of out of sample R_{Med}^2 based on MC simulation, Scenario A.

Sample Size (train)	R_{Med}^2 (test) Method	10%	25%	50%	75%	90%
n=160	LASSO	0.823	0.980	0.986	0.988	0.990
	SCAD	0.646	0.981	0.987	0.989	0.991
	MCP	0.646	0.981	0.987	0.989	0.991
n=320	LASSO	0.862	0.985	0.989	0.990	0.991
	SCAD	0.904	0.986	0.989	0.991	0.991
	MCP	0.904	0.986	0.989	0.991	0.991
n=480	LASSO	0.912	0.988	0.989	0.991	0.992
	SCAD	0.909	0.988	0.990	0.991	0.992
	MCP	0.909	0.988	0.990	0.991	0.992

Scenario B:

In this scenario, the primary goal is to illustrate the benefit of using the non-parametric variable selection method proposed in this article by comparing it the with the variable selection method in functional linear concurrent model proposed in Chapter 3. In particular, we want to demonstrate what one can gain using the nonparametric method in terms of selection accuracy and prediction performance when there are nonlinear effects of predictors present. As in Scenario A, we obtain the denoised curves $\hat{X}_{ij}(t)$ before applying our variable selection method. Here covariate $X_1(\cdot), X_2(\cdot)$ have a linear effect on the response, whereas $X_3(\cdot)$ have a nonlinear effect. We obtain the selection percentage of each variable for each of the three selection methods discussed in this article and compare them to that of variable selection method in FLCM. The results are displayed in Table 4.3. We observe that for the variable selection method in FLCM, the effect of $X_3(\cdot)$ is captured in 99.8% case for group LASSO penalty, in 96.2% case for group SCAD penalty and in 89.4% case for group MCP penalty. In comparison, the variable selection method for NPFCM captures the nonlinear effect of $X_3(\cdot)$ in 100% cases for all the penalties, also the group SCAD and group MCP penalty are able to identify the true underlying model 100% of the time. This illustrates how one can avoid making a false negative decision, using the variable selection method for nonparametric functional concurrent model (NPFCM).

Table 4.3 Comparison of selection percentages (%) of different variables, Scenario B.

Method	Penalty	Var 1	Var 2	Var 3	Max (Var 4-Var 20)	Avg Model Size
FLCM	LASSO	100	100	99.8	2.8	3.32
	SCAD	100	100	96.2	0.4	2.968
	MCP	100	100	89.4	0.2	2.896
NPFCM	LASSO	100	100	100	17.4	4.712
	SCAD	100	100	100	0	3
	MCP	100	100	100	0	3

Next, a similar comparison is made between the methods in terms of their out of sample prediction performance. The FLCM intuitively should only be able to capture linear component

of the nonlinear effect and thus may not be suitable for prediction. The results are displayed in Table 4.4. We notice that the median out of sample R_{Med}^2 is around 98.8% for variable selection method in FLCM, whereas for variable selection method in NPFCM this is 99.9%. So in terms of out of sample prediction accuracy also, the variable selection method in non-parametric functional concurrent regression produces better performance.

Table 4.4 Quantiles of out of sample R_{Med}^2 based on MC simulation, Scenario B.

Method	$R_{Med}^2(\text{test})$ Penalty	10%	25%	50%	75%	90%
FLCM	LASSO	0.986	0.987	0.988	0.989	0.990
	SCAD	0.986	0.987	0.988	0.989	0.990
	MCP	0.986	0.987	0.988	0.989	0.990
NPFCM	LASSO/SCAD/MCP	0.999	0.999	0.999	0.999	0.999

The simulation results illustrate the proposed variable selection method in NPFCM can identify the underlying variables with high accuracy and also gives satisfactory out of sample prediction performance. The group SCAD and group MCP based selection methods have minuscule false positive and false negative rates, and as in Chapter 3 are the recommended methods in this article.

4.4 Real Data Applications

In this section, we demonstrate application of our variable selection method in NPFCM for selection of functional covariates in two real data studies. We first consider the calcium absorption study with added pseudovariates as an illustration of our method. Next, we apply the proposed method for selection of influential weather variables in a capital bike sharing data and estimate their concurrent nonparametric effects on casual bike rentals.

4.4.1 Study of Dietary Calcium Absorption

We consider the study of dietary calcium absorption given in Davis (2002). As discussed in the

Chapter 3, we have data on calcium absorption $Y(t)$, dietary calcium intake $X_1(t)$, BMI $X_2(t)$ and BSA (Body surface area) $X_3(t)$ of 188 patients at irregular time-points between 35 to 64 years of their ages. For each patient, the number of repeated measurements varies between 1 to 4. Figure 3.2 in Chapter 3 displays the individual curves of patients' calcium absorption, calcium intake, BSA, BMI along their ages. We are primarily interested in finding out which predictors influence calcium absorption $Y(t)$. As data is observed sparsely and the original covariates might be observed with measurement error, we apply FPCA methods as discussed in Section 4.2.2 and get the denoised trajectories $\hat{X}_j(t)$ for $j = 1, 2, 3$. We expect calcium intake to be associated with calcium absorption, as demonstrated by our proposed for variable selection method for FLCM in Chapter 3. To further illustrate the selection performance and false positive rate of our variable selection method, we add 15 pseudo covariates by simulating from the following functional model. We generate $X_{ij}(\cdot) \stackrel{iid}{\sim} X_j(\cdot)$, where $X_j(t)$ ($j = 4, 5, \dots, 18$) are given by $X_j(t) = a_j\sqrt{2}\sin(\pi(j-3)t/200) + b_j\sqrt{2}\cos(\pi(j-3)t/200)$, where $a_j \sim \mathcal{N}(0, (2)^2)$, $b_j \sim \mathcal{N}(0, (2)^2)$. Finally we apply the variable selection method for NPFCM to $Y(t)$ and $\hat{X}_1(t), \hat{X}_2(t), \hat{X}_3(t), X_4(t), X_5(t), \dots, X_{18}(t)$. We repeat the whole procedure a large number of times. Table 4.5 displays the selection percentage of each of the variables. We notice that both the group SCAD and group MCP method identifies calcium intake $X_1(t)$ as a significant predictor 100% of the time. All other variables including the randomly generated predictors are ignored in 100% of the iterations. So both the group SCAD and group MCP method do a good job in selecting the true predictor set and discarding the pseudovariables. As calcium intake is selected by both the variable selection methods we want to estimate its non-parametric effect. Therefore we use the proposed method for performing variable selection on the original data without adding any pseudovariables. The estimated effect calcium intake (adjusted for the main effect of t) is displayed in Figure 4.2. We notice that in this case, the effect of calcium intake is more or less linear which matches with prior findings of Kim et al. (2018). For fixed and smaller values of t , as x increases, the effect is decreasing in most of the regions indicating as calcium intake increase dietary calcium absorption should decrease at earlier ages, but for

Table 4.5 Selection Percentages (%) of variables in Calcium absorption Study using NPFCM.

Method	Var 1(Calcium Intake)	Var 2 (BSA)	Var 3(BMI)	Max Var (4-18)
SCAD	100	0	0	0
MCP	100	0	0	0

older ages (above 60 years) the effect seems to be increasing indicating an increase in dietary calcium absorption with increasing calcium intake. These findings match our results in Chapter 3. One needs to be careful in interpreting the result beyond 60 years, as there might be high uncertainty associated with the estimate because of very less number of observations above this mark as illustrated in Chapter 3.

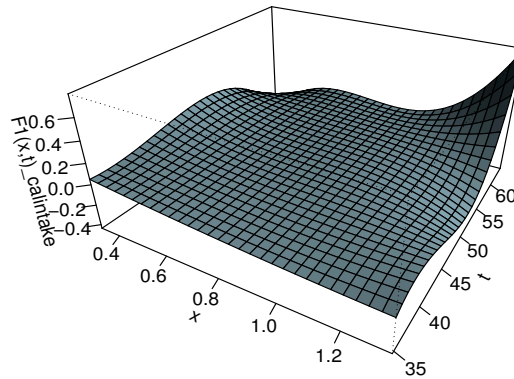


Figure 4.1 Estimate of non-parametric effect of calcium intake on calcium absorption from proposed variable selection method in NPFCM.

4.4.2 Study of Bike Sharing Data

Next, we illustrate another application of our variable selection method on the capital bike

share study (Fanaee-T and Gama, 2014). The data were originally obtained from the Capital Bike Share system in Washington, D.C., and contains casual bike rentals data along with other weather information such as temperature (temp), feels-like temperature (atemp), humidity (hum) and windspeed (windspeed) on an hourly basis. Casual bike rentals are defined as rentals to cyclists without membership in the Capital Bike Share program. Since bike rentals can have different dynamics on weekends, compared to weekdays; following the analysis of Kim et al. (2018) we only restrict our attention to casual bike rentals on Saturdays. The counts of casual bike rentals are available during the period from January 1, 2011, to December 31, 2012, on a total of 105 Saturdays on an hourly basis barring some exceptions (8 missing). Figure 4.2 displays the daily curves of hourly casual bike rentals and other weather variables. It is

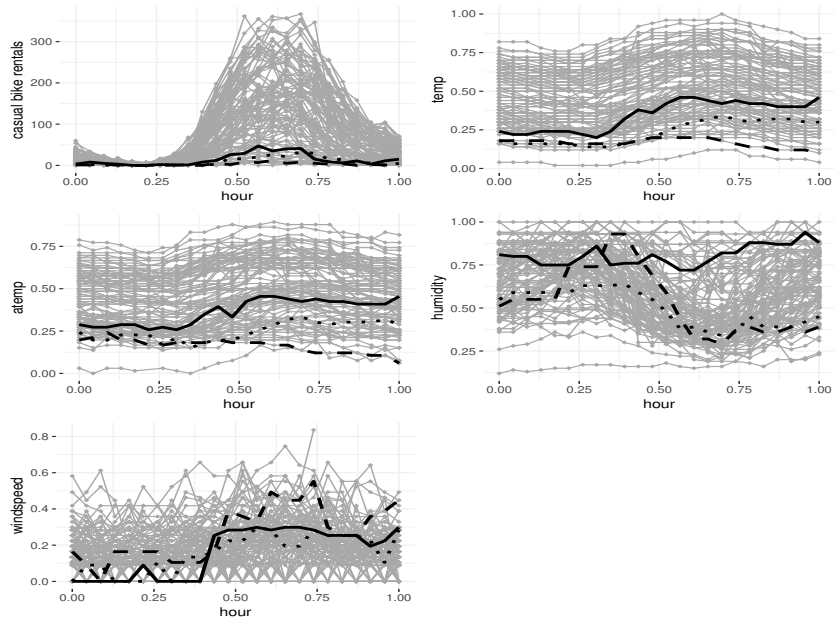


Figure 4.2 Hourly casual Bike rentals and weather variables profile in bike sharing data.

plausible that the weather variables have real time dynamic effects in determining the number of bike rentals and we are primarily interested in selecting the relevant covariates as well as estimating their non-parametric concurrent effect. The hourly bike rentals $Y(t)$ are discrete and

their distribution is possibly skewed, therefore we log transform the response prior to analysis ($Y(\cdot) = \log\{Y(\cdot) + 1\}$). Before applying our variable selection method, we add 16 random functional covariates to the existing 4 weather variables by generating from the covariate model $X_j(t) = a_j\sqrt{2}\sin(2\pi(j-4)t) + b_j\sqrt{2}\cos(2\pi(j-4)t)$, where $a_j \sim \mathcal{N}(0, (2)^2)$, $b_j \sim \mathcal{N}(0, (2)^2)$, for $j = 5, 6, \dots, 20$. This is again done in order to assess the false selection performance of the proposed method. We pre-process the weather variables $X_1(t), X_2(t), X_3(t), X_4(t)$ using FPCA methods as discussed in Section 4.2.2 and then apply our variable selection method to $Y(t)$ and $\hat{X}_1(t), \hat{X}_2(t), \hat{X}_3(t), \hat{X}_4(t), X_5(t), \dots, X_{20}(t)$. The whole procedure is repeated several times. Table 4.6 displays the selection percentage of each of the predictors for the group SCAD and group MCP method. We see the group SCAD method selects feels-like temperature $X_2(t)$,

Table 4.6 Selection Percentages (%) of variables in Bike Sharing Study.

Method	Var1 (temp)	Var2 (atemp)	Var3 (hum)	Var4 (windspeed)	Max Var (5-20)
SCAD	0	100	100	100	0
MCP	100	0	100	100	0

humidity $X_3(t)$ and windspeed $X_4(t)$ 100% of the time, whereas the group MCP method selects temperature $X_1(t)$, humidity $X_3(t)$ and windspeed $X_4(t)$ 100% of the time. The covariates temperature $X_1(t)$ and feels-like temperature $X_2(t)$ should be highly correlated and it is ideal that the variable selection method is picking only one of them consistently. In terms of the false positive rate, we see that both the group SCAD and group MCP method discards the pseudovariates in all the iterations which is extremely satisfactory indicating an almost negligible false selection rate of the proposed variable selection method in NPFCM, which is what we observed in our simulations. Next to estimate the nonparametric concurrent effects of the weather variables, we use the group MCP method for performing variable selection on the actual data (without any random covariates added). The estimated concurrent effects of the selected weather variables (adjusted for the main effect of t) are displayed in Figure 4.3. The estimates from the group SCAD method are similar. We observe that, the effect of temperature on bike rentals is

maximum when the temperature is in the mid-range and during midday as this should be an ideal condition for biking. For too cold or too warm temperature the effect of temperature

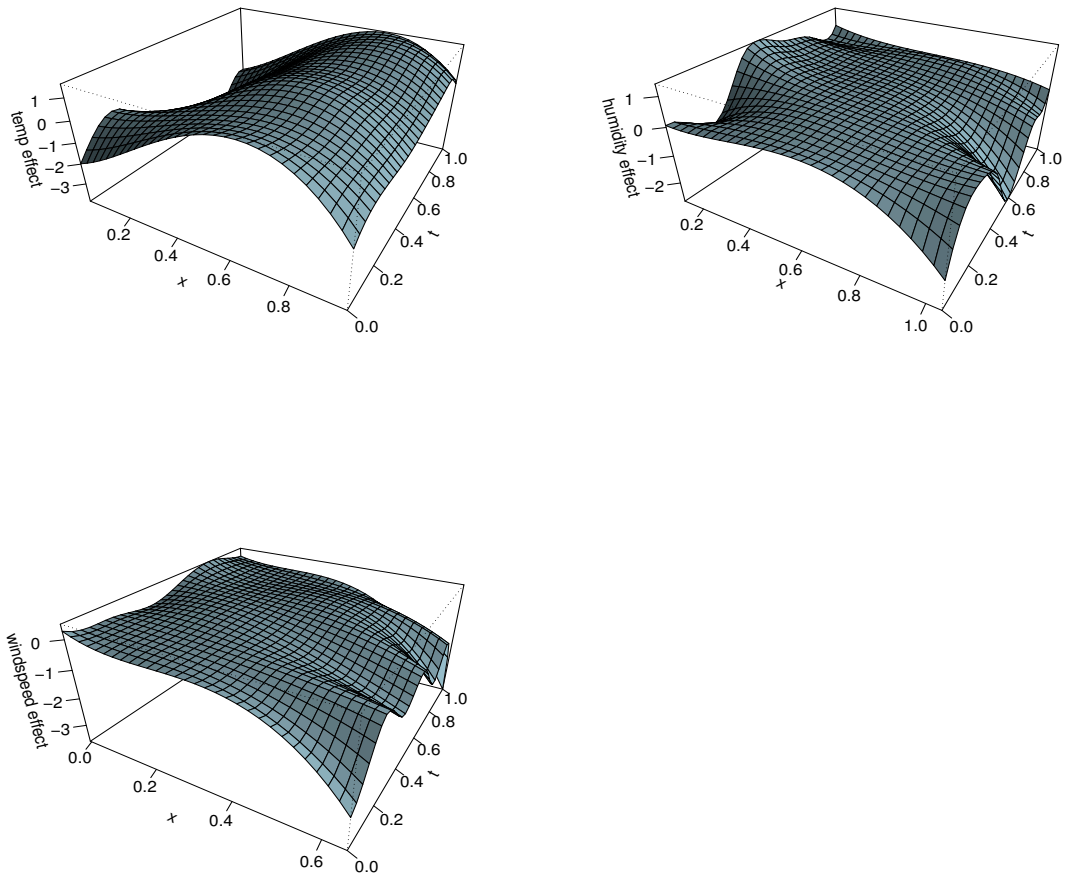


Figure 4.3 Estimated non-parametric concurrent effect of the weather variables. Displayed are the effects of temperature (top left), humidity (top right) and windspeed (bottom row).

on bike rental decreases, which is natural. For mid-range to higher temperatures, the effect of temperature on casual bike rentals is more or less uniform throughout the day, which is expected

in weekends and good weather in Washington, DC.

For humidity, there is a negative association for each fixed timepoint as its effect decrease with increasing humidity. This was also observed in Kim et al. (2018). For lower humidity the effect increases during midday and is highly positive, and for higher humidity we see a sharp drop in the effect, indicating in humid days there should be a fall in bike rentals, which is again what one would naturally expect in such extreme weather condition. For windspeed, the effect is more or less constant across the domain except in higher windspeed regions where bikers might face difficulties. In such high windspeed regions during midday (plausibly the hottest time of the day), there is a noticeable dip which can be attributed to the difficulty one might face while cycling during high-velocity winds.

Our findings match with that of Gebhart and Noland (2014) who made similar observations using scalar regression techniques on the capital bike sharing data. Our analysis using NPFCM gives a more comprehensive understanding of the effect of the weather variables.

Our data analysis illustrates with the variable selection method in NPFCM, it is possible to identify the underlying true predictors as the source of major variations in the response and capture their concurrent effect which can be helpful to understand the dynamics between the functional predictors and the functional response.

4.5 Discussion

In this article, we have proposed a variable selection method in non-parametric functional concurrent regression, extending the classically used penalized variable selection methods like LASSO, SCAD, and MCP. Through numerical simulations, we have shown the proposed variable selection method for NPFCM, particularly with group SCAD or group MCP penalty not only identifies the true underlying variables with minimal false positive and minuscule false negative rate but also performs extremely well in terms of out of sample prediction accuracy; even when data is observed sparsely, is contaminated with measurement error and the error process is highly non stationary. Our research also illustrates there might be situations, where the covariates

might have complex nonlinear effects and using the variable selection method for nonparametric functional concurrent regression model is more suitable than using the variable selection method in FLCM in such cases. We have demonstrated application of our method on a dietary calcium absorption study and a bike sharing study, for selection of influential covariates and estimating their underlying nonparametric effects.

We have used an independent working correlation structure in our proposed variable selection method for NPFCM, and this has shown good performance in our simulation study in terms of selection accuracy. It is plausible to consider a pre-whitening method to take into account temporal dependence within functions which might yield gain in terms of predictive inference. This can be done similarly as in chapter 3.

There are plenty of research problems that remain to be explored based on this current work. It is of particular interest to study small sample properties of the estimated non-parametric functions and get a measure of their uncertainty. Resampling subject bootstrap technique could be one way to handle this but gets computationally intractable with increasing dimensionality of the feature space. Perturbation based bootstrap (Das et al., 2019) procedure could be interesting to explore in such situations. For distribution free predictive inference split conformal prediction bands (Lei et al., 2018) can be explored by extending the conformal prediction approach for functional response (Lei et al., 2015).

The way we have developed our method is a general one and can be extended to other regression scenarios e.g., the historical functional model where we can consider past effects of covariates on the response at current time point. It would be also interesting to develop methods to take into account interactions that might be present between the predictors. Extending the proposed method to these general class of functional regression models is certainly plausible and remain areas for future research.

Software

Illustrations of implementation of our method using R along with the datasets discussed are available on Github at (https://github.com/rahulfrodo/NPFCM_Selection).

Chapter 5

Conclusion

The domain and application of functional data are rapidly increasing with data coming in from financial markets, climate models, wearable devices and biomedical research. Function on function regression models are set to play a key role in analyzing such data. In this dissertation, we have developed hypothesis testing and variable selection methods for functional concurrent regression models which will be useful for dynamic modeling of such data. The developed methods can be applied for inference when both the covariate and response are functional in nature.

The proposed hypothesis testing method is useful for testing the global linear concurrent effect of functional predictors on functional response. For high dimensional functional data with a large number of covariates, we have proposed a variable selection method for functional linear concurrent regression which can be used to identify the underlying set of predictors while simultaneously estimating their dynamic effects. To take into account possible nonlinear effect of the predictors, we have extended this method to a variable selection method for nonparametric functional concurrent model. The proposed methods in terms of reproducible R code are available at Github.

There are multiple possible research directions that could be explored based on our work. It will be interesting to consider the lagged effect of predictors in both our testing and variable

selection problem as many predictors could have impacts that take time to materialize. Functional historical model could be considered for this purpose allowing both linear and nonlinear effects. In many applications, there might be significant interaction at play between the predictors. Accounting for these interaction effects and going beyond additive modeling would be worth exploring in the context of hypothesis testing and variable selection problem in function on function regression models. Functional single index model can be useful for this purpose.

Developing theory and methods for correlated functional predictors is also something that could be looked into. When the predictors are highly correlated we would want our variable selection methods to pick out only one representative covariate. Extending existing methods for scalar regression for such situations is certainly worth consideration.

In many real data applications, the predictors could have subject specific variability. These can be addressed using subject specific effects following a functional mixed modeling approach. The main challenge will be to handle the computational complexity of optimization with increasing dimensionality. Using Bayesian methods for this purpose is, therefore, worth consideration.

Generalized functional regression model for non-continuous functional responses is another field where the proposed methods could be extended to. The proposed variable selection methods in this dissertation can be readily applied in such models, by replacing the least square loss with an appropriate negative log-likelihood function. This could have potentially interesting applications in biological studies and material science where the response of interest might be binary or integer-valued.

Finally, a new direction of research can be developing methods for inference in function on function regression model which integrates multiple outcomes (both discrete and continuous) from different sources. Such multi-omics data would require substantial advancement of existing methods in functional regression literature and remains an area for future research.

BIBLIOGRAPHY

- Breheeny, P. (2019). Package `grpreg`.
- Breheeny, P. and Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and computing* **25**, 173–187.
- Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association* **95**, 941–956.
- Chen, J. and Chen, Z. (2008). Extended bayesian information criteria for model selection with large model spaces. *Biometrika* **95**, 759–771.
- Chen, Y., Goldsmith, J., and Ogden, R. T. (2016). Variable selection in function-on-scalar regression. *Stat* **5**, 88–101.
- Chiang, C.-T., Rice, J. A., and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association* **96**, 605–619.
- Clark, T. P. and Longo, S. B. (2019). Examining the effect of economic development, region, and time period on the fisheries footprints of nations (1961-2010). *International Journal of Comparative Sociology* page 0020715219869976.
- Clark, T. P., Longo, S. B., Clark, B., and Jorgenson, A. K. (2018). Socio-structural drivers, fisheries footprints, and seafood consumption: A comparative international study, 1961-2012. *Journal of rural studies* **57**, 140–146.
- Cohen, M. J. and Garrett, J. L. (2010). The food price crisis and urban food (in) security. *Environment and Urbanization* **22**, 467–482.
- Crainiceanu, C. M. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 165–185.
- Das, D., Gregory, K., Lahiri, S., et al. (2019). Perturbation bootstrap in adaptive lasso. *The Annals of Statistics* **47**, 2080–2116.
- Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*.
- Dietz, T. and Jorgenson, A. (2013). *Structural human ecology: new essays in risk, energy, and sustainability*. Washington State University Press.
- Eubank, R., Huang, C., Maldonado, Y. M., Wang, N., Wang, S., and Buchanan, R. (2004). Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**, 653–667.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* **96**, 1348–1360.

- Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics* pages 1491–1518.
- Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* **27**, 715–731.
- Fan, Y., James, G. M., Radchenko, P., et al. (2015). Functional additive regression. *The Annals of Statistics* **43**, 2296–2325.
- Fanaee-T, H. and Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence* **2**, 113–127.
- FAO (2016). State of the worlds fisheries and aquaculture.
- Gebhart, K. and Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in washington, dc. *Transportation* **41**, 1205–1225.
- Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* **98**, 387–396.
- Gertheiss, J., Maity, A., and Staicu, A.-M. (2013). Variable selection in generalized functional linear models. *Stat* **2**, 86–101.
- Goldsmith, J. and Schwartz, J. E. (2017). Variable selection in the functional linear concurrent model. *Statistics in medicine* **36**, 2237–2250.
- Greven, S., Crainiceanu, C. M., Küchenhoff, H., and Peters, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics* **17**, 870–891.
- Guo, W. (2002). Functional mixed effects models. *Biometrics* **58**, 121–128.
- Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics* **34**, 1493–1517.
- Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association* **113**, 649–659.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 757–796.
- Hoover, D. R., Rice, J. A., Wu, C. O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85**, 809–822.
- Huang, J. Z., Wu, C. O., and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* **89**, 111–128.
- Huang, J. Z., Wu, C. O., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* pages 763–788.

- Inan, G. and Wang, L. (2017). Pgee: An r package for analysis of longitudinal data with high-dimensional covariates. *The R Journal* **9**, 393–402.
- Jiang, C.-R., Wang, J.-L., et al. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39**, 362–388.
- Jorgenson, A. K. and Clark, B. (2010). Assessing the temporal stability of the population/environment relationship in comparative perspective: a cross-national panel study of carbon dioxide emissions, 1960–2005. *Population and Environment* **32**, 27–41.
- Jorgenson, A. K., Rice, J., and Crowe, J. (2005). Unpacking the ecological footprint of nations. *International Journal of Comparative Sociology* **46**, 241–260.
- Kauermann, G. and Tutz, G. (1999). On model diagnostics using varying coefficient models. *Biometrika* **86**, 119–128.
- Kim, J., Maity, A., and Staicu, A.-M. (2018). Additive nonlinear functional concurrent model. *Statistics and Its Interface* **11**, 669–685.
- Kim, J. S., Staicu, A.-M., Maity, A., Carroll, R. J., and Ruppert, D. (2018). Additive function-on-function regression. *Journal of Computational and Graphical Statistics* **27**, 234–244.
- Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T. D., Block, B. A., et al. (2018). Tracking the global footprint of fisheries. *Science* **359**, 904–908.
- Lei, J., GSell, M., Rinaldo, A., Tibshirani, R. J., and Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association* **113**, 1094–1111.
- Lei, J., Rinaldo, A., and Wasserman, L. (2015). A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence* **74**, 29–43.
- Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38**, 3321–3351.
- Lin, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–326.
- Liu, B., Wang, L., and Cao, J. (2017). Estimating functional linear mixed-effects regression models. *Computational Statistics & Data Analysis* **106**, 153–164.
- Longo, S. B. and Clark, B. (2016). An ocean of troubles: advancing marine sociology. *Social Problems* **63**, 463–479.
- Longo, S. B., Clark, B., York, R., and Jorgenson, A. K. (2019). Aquaculture and the displacement of fisheries captures. *Conservation Biology* .

- Maity, A. (2017). Nonparametric functional concurrent regression models. *Wiley Interdisciplinary Reviews: Computational Statistics* **9**,
- Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics* **31**, 115–128.
- Mazumder, R., Friedman, J. H., and Hastie, T. (2011). Sparsenet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association* **106**, 1125–1138.
- McLean, M. W., Hooker, G., Staicu, A.-M., Scheipl, F., and Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics* **23**, 249–269.
- Meier, L., Van de Geer, S., Bühlmann, P., et al. (2009). High-dimensional additive modeling. *The Annals of Statistics* **37**, 3779–3821.
- Mercer, J. (1909). Functions of positive and negative type, and their connection with the theory of integral equations. *Phil. Trans. R. Soc. Lond. A* **209**, 415–446.
- Miller, A. (2002). *Subset selection in regression*. Chapman and Hall/CRC.
- Molenberghs, G. and Verbeke, G. (2007). Likelihood ratio, score, and wald tests in a constrained parameter space. *The American Statistician* **61**, 22–27.
- Ramsay, J. and Silverman, B. (2005). *Functional data analysis*.
- Satterthwaite, D., McGranahan, G., and Tacoli, C. (2010). Urbanization and its implications for food and farming. *Philosophical transactions of the royal society B: biological sciences* **365**, 2809–2820.
- Şentürk, D. and Nguyen, D. V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica* **21**, 1831.
- Shi, J. Q., Murray-Smith, R., and Titterton, D. (2005). Hierarchical gaussian process mixtures for regression. *Statistics and computing* **15**, 31–41.
- Staicu, A.-M., Li, Y., Crainiceanu, C. M., and Ruppert, D. (2014). Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics* **41**, 932–949.
- Steffen, W., Persson, Å., Deutsch, L., Zalasiewicz, J., Williams, M., Richardson, K., Crumley, C., Crutzen, P., Folke, C., Gordon, L., et al. (2011). The anthropocene: From global change to planetary stewardship. *Ambio* **40**, 739.
- Theologis, T. (2009). *Childrens orthopaedics and fractures* (chapter 6).
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* pages 267–288.
- Torres-Reyna, O. (2007). *Panel data analysis fixed and random effects using stata (v. 4.2)*. Data & Statistical Services, Princeton University .

- Valderrama, D. and Anderson, J. L. (2010). Market interactions between aquaculture and common-property fisheries: Recent evidence from the bristol bay sockeye salmon fishery in alaska. *Journal of Environmental Economics and Management* **59**, 115–128.
- Wang, B. and Shi, J. Q. (2014). Generalized gaussian process regression model for non-gaussian functional data. *Journal of the American Statistical Association* **109**, 1123–1133.
- Wang, H., Zhong, P.-S., Cui, Y., and Li, Y. (2018). Unified empirical likelihood ratio tests for functional concurrent linear models and the phase transition from sparse to dense functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **80**, 343–364.
- Wang, L., Li, H., and Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association* **103**, 1556–1569.
- Wang, L., Zhou, J., and Qu, A. (2012). Penalized generalized estimating equations for high-dimensional longitudinal data analysis. *Biometrics* **68**, 353–360.
- World Bank (2007). Changing the face of waters: the promise and challenge of sustainable aquaculture.
- Wu, C. O., Chiang, C.-T., and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American statistical Association* **93**, 1388–1402.
- Wu, Y., Boos, D. D., and Stefanski, L. A. (2007). Controlling variable selection by the addition of pseudovariables. *Journal of the American Statistical Association* **102**, 235–243.
- Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.
- Yao, F., Müller, H.-G., Wang, J.-L., et al. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.
- York, R., Rosa, E. A., Dietz, T., et al. (2003). Footprints on the earth: The environmental consequences of modernity. *American sociological review* **68**, 279–300.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 49–67.
- Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of statistics* **38**, 894–942.
- Zhang, D. and Lin, X. (2003). Hypothesis testing in semiparametric additive mixed models. *Biostatistics* **4**, 57–74.
- Zhang, D. and Lin, X. (2008). *Variance Component Testing in Generalized Linear Mixed Models for Longitudinal/Clustered Data and other Related Topics*, pages 19–36.
- Zhang, J.-T., Chen, J., et al. (2007). Statistical inferences for functional data. *The Annals of Statistics* **35**, 1052–1079.

APPENDIX

Appendix A

Proofs and Additional Results

A.1 Proof of Theorem 1

Suppose that the conditions (a) and (b) of Theorem 1 hold, i.e., the null hypothesis $H_0 : \tau_1 = 0$ holds, $\boldsymbol{\theta}_0 = (\tau_0^*, 0)$ is the true value of $\boldsymbol{\theta}$ and Σ is the true covariance matrix of residual vector \mathcal{E} . Denote $\mathbb{V} = \mathbb{V}(\boldsymbol{\theta}_0, \Sigma)$, $\mathbf{Y} \sim \mathcal{N}(0, \mathbb{V})$ under null. Then $\mathbb{W} = \mathbb{V}^{-1/2}\mathbf{Y} \sim \mathcal{N}(0, \mathbb{I})$. Now $S_{\tau_1}(\boldsymbol{\theta}_0) = -1/2\{tr(\mathbb{Z}^T\mathbb{V}^{-1}\mathbb{Z}) - (\mathbb{V}^{-1/2}\mathbf{Y})^T\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}(\mathbb{V}^{-1/2}\mathbf{Y})\}$. Hence

$$\begin{aligned} \frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} &= \frac{[1/2\{tr(\mathbb{Z}^T\mathbb{V}^{-1}\mathbb{Z}) - (\mathbb{V}^{-1/2}\mathbf{Y})^T\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}(\mathbb{V}^{-1/2}\mathbf{Y})\}]^2}{\Lambda(\boldsymbol{\theta}_0)} \\ &= \frac{[1/2\{tr(\frac{\mathbb{Z}^T\mathbb{V}^{-1}\mathbb{Z}}{n}) - (\mathbb{V}^{-1/2}\mathbf{Y})^T\frac{\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}}{n}(\mathbb{V}^{-1/2}\mathbf{Y})\}]^2}{\frac{\Lambda(\boldsymbol{\theta}_0)}{n^2}} \\ &= \frac{[1/2\{\sum_{\ell=1}^{k_1} \lambda_\ell - (\mathbb{W})^T(\frac{\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}}{n})(\mathbb{W})\}]^2}{\Lambda_n(\boldsymbol{\theta}_0)} \\ &= \frac{[1/2\{\sum_{\ell=1}^{k_1} \lambda_\ell - (\mathbb{W})^T\mathbb{U}\mathbb{D}\mathbb{U}^T(\mathbb{W})\}]^2}{\Lambda_n(\boldsymbol{\theta}_0)}. \end{aligned}$$

Now we use the spectral decomposition $\frac{\mathbb{V}^{-1/2}\mathbb{Z}\mathbb{Z}^T\mathbb{V}^{-1/2}}{n} = \mathbb{U}\mathbb{D}\mathbb{U}^T$, where \mathbb{D} is a diagonal matrix and \mathbb{U} is orthogonal. Then using this decomposition we can rewrite the score test statistic above

in the following manner.

$$\begin{aligned}
\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} &= \frac{\left[1/2\{\sum_{\ell=1}^{k_1} \lambda_\ell - (\mathbf{U}^T \mathbb{W})^T \mathbb{D}(\mathbf{U}^T \mathbb{W})\}\right]^2}{\Lambda_n(\boldsymbol{\theta}_0)} \\
&= \frac{\{1/2(\sum_{\ell=1}^{k_1} \lambda_\ell - \mathbb{X}^T \mathbb{D} \mathbb{X})\}^2}{\Lambda_n(\boldsymbol{\theta}_0)}. \\
&= \frac{\{1/2(\sum_{\ell=1}^{k_1} \lambda_\ell - \sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2)\}^2}{\Lambda_n(\boldsymbol{\theta}_0)}.
\end{aligned}$$

This follows from the fact $\mathbb{X} = \mathbf{U}^T \mathbb{W} \sim \mathcal{N}(0, \mathbb{I})$ as \mathbf{U} is an orthogonal matrix and nonzero eigenvalues of $\frac{\mathbb{V}^{-1/2} \mathbb{Z} \mathbb{Z}^T \mathbb{V}^{-1/2}}{n}$ and $\frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n}$ are same.

Therefore we have shown $\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} \stackrel{d}{=} \frac{\{1/2(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell)\}^2}{\Lambda_n(\boldsymbol{\theta}_0)}$ under null, which completes the proof of our Theorem 1 namely:

$$T_S(\boldsymbol{\theta}_0, \Sigma) = \frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \stackrel{d}{=} (1/2)^2 \frac{\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell\right)^2}{\Lambda_n(\boldsymbol{\theta}_0)} I\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell\right).$$

Where $x_\ell \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and λ_ℓ are eigenvalues of $\frac{\mathbb{Z}^T \mathbb{V}(\boldsymbol{\theta}_0, \Sigma)^{-1} \mathbb{Z}}{n}$.

A.2 Proof of Theorem 2

Suppose $\tilde{\boldsymbol{\theta}}$ (MLE under null) and $\hat{\Sigma}$ are consistent estimators of $\boldsymbol{\theta}_0$ and Σ in the sense $\tilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ and $\|\hat{\Sigma}^{-1} - \Sigma^{-1}\|_2 = o_p(1)$ (spectral norm). We want to show $T_S(\tilde{\boldsymbol{\theta}}, \hat{\Sigma}) \xrightarrow{d} T_S(\boldsymbol{\theta}_0, \Sigma)$.

We note that in the proof of Theorem 1,

$$\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} = \frac{[1/2\{tr(\frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n}) - (\mathbb{V}^{-1/2} \mathbf{Y})^T \frac{\mathbb{V}^{-1/2} \mathbb{Z} \mathbb{Z}^T \mathbb{V}^{-1/2}}{n} (\mathbb{V}^{-1/2} \mathbf{Y})\}]^2}{\Lambda_n(\boldsymbol{\theta}_0)} = \frac{S_{\tau_1, n}^2(\boldsymbol{\theta}_0)}{\Lambda_n(\boldsymbol{\theta}_0)},$$

where $S_{\tau_1, n}(\boldsymbol{\theta}_0) = S_{\tau_1}(\boldsymbol{\theta}_0)/n$ and $\mathbb{V} = V(\boldsymbol{\theta}_0, \Sigma) = \Sigma + \tau_0^* \mathbb{B} \mathbb{B}^T$. Thus we have $T_S(\boldsymbol{\theta}_0, \Sigma) = \frac{S_{\tau_1, n}^2(\boldsymbol{\theta}_0)}{\Lambda_n(\boldsymbol{\theta}_0)} I(S_{\tau_1, n}(\boldsymbol{\theta}_0) \geq 0)$. We now show $S_{\tau_1, n}(\tilde{\boldsymbol{\theta}}) \xrightarrow{d} S_{\tau_1, n}(\boldsymbol{\theta}_0)$ and $\Lambda_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \Lambda_n(\boldsymbol{\theta}_0)$, which then by Slutsky's theorem proves $T_S(\tilde{\boldsymbol{\theta}}, \hat{\Sigma}) \xrightarrow{d} T_S(\boldsymbol{\theta}_0, \Sigma)$.

$$S_{\tau_1, n}(\tilde{\boldsymbol{\theta}}) = [-1/2\{tr(\frac{\mathbb{Z}^T \tilde{\mathbb{V}}^{-1} \mathbb{Z}}{n}) - (\mathbf{Y})^T \frac{\tilde{\mathbb{V}}^{-1} \mathbb{Z} \mathbb{Z}^T \tilde{\mathbb{V}}^{-1}}{n} (\mathbf{Y})\}].$$
 Where $\tilde{\mathbb{V}} = V(\tilde{\boldsymbol{\theta}}, \hat{\Sigma}) = \hat{\Sigma} + \tilde{\tau}_0^* \mathbb{B} \mathbb{B}^T$.

It is enough to show $\|\frac{\mathbb{Z}^T \tilde{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n}\|_2 = o_p(1)$ and $\|\frac{\tilde{\mathbb{V}}^{-1} \mathbb{Z} \mathbb{Z}^T \tilde{\mathbb{V}}^{-1}}{n} - \frac{\mathbb{V}^{-1} \mathbb{Z} \mathbb{Z}^T \mathbb{V}^{-1}}{n}\|_2 = o_p(1)$ to

conclude $S_{\tau_1, n}(\tilde{\boldsymbol{\theta}}) \xrightarrow{d} S_{\tau_1, n}(\boldsymbol{\theta}_0)$.

Using Woodbury matrix identity, we can write

$$\frac{\mathbb{Z}^T \tilde{\mathbb{V}}^{-1} \mathbb{Z}}{n} = \frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{Z}}{n} - \tilde{\tau}_0^* \left(\frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right) \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} \left(\frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^T. \quad (\text{A.1})$$

We prove that each part in the r.h.s of (A.1) converges to the corresponding part of the right hand side, with τ_0 and Σ in place of $\tilde{\tau}_0^*$ and $\hat{\Sigma}$, and also show that each individual piece is bounded.

We observe $\left\| \frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \Sigma^{-1} \mathbb{Z}}{n} \right\|_2 \leq \left\| \frac{\mathbb{Z}^T}{\sqrt{n}} \right\|_2 \left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2 \left\| \frac{\mathbb{Z}}{\sqrt{n}} \right\|_2 \leq \left\| \frac{\mathbb{Z}^T}{\sqrt{n}} \right\|_F \left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2 \left\| \frac{\mathbb{Z}}{\sqrt{n}} \right\|_F$.

Here $\|\cdot\|_F$ denotes the Frobenius norm and $\|\mathbb{A}\|_2 \leq \|\mathbb{A}\|_F$. Note as defined in the main article

$\mathbb{Z} = \mathbb{X} \Sigma_1^{1/2}$. We then have $\left\| \frac{\mathbb{Z}^T}{\sqrt{n}} \right\|_F = \left\| \frac{\mathbb{Z}}{\sqrt{n}} \right\|_F = \left\| \frac{\mathbb{X} \Sigma_1^{1/2}}{\sqrt{n}} \right\|_F \leq \left\| \Sigma_1^{1/2} \right\|_F \sqrt{\text{tr}(\frac{\mathbb{X}^T \mathbb{X}}{n})} < D$ (some

$D > 0$), as $\left\| \Sigma_1^{1/2} \right\|_F$ is bounded; $\Sigma_1^{1/2}$ being a fixed $k_1 \times k_1$ matrix. Also note that $\text{tr}(\mathbb{X}^T \mathbb{X}/n) = \left\| \mathbb{X} \right\|_F^2/n = \sum_{j=1}^m \sum_{i=1}^n X_i^2(t_j)/n \sum_{k=1}^{k_1} B_{k1}(t_j)^2 \rightarrow \sum_{j=1}^m \mu_2(t_j) \sum_{k=1}^{k_1} B_{k1}(t_j)^2 < \infty$ by S.L.L.N,

as X_i 's were assumed to be i.i.d. with finite second moment. So $\left\| \frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \Sigma^{-1} \mathbb{Z}}{n} \right\|_2 = o_p(1)$,

as $\left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2 = o_p(1)$ and the other terms are bounded as shown above.

Now similarly $\left\| \frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} - \frac{\mathbb{Z}^T \Sigma^{-1} \mathbb{B}}{n} \right\|_2 \leq \left\| \Sigma_1^{1/2} \right\|_F \sqrt{\text{tr}(\frac{\mathbb{X}^T \mathbb{X}}{n})} \left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2 \left\| \Sigma_0^{1/2} \right\|_F \sqrt{\text{tr}(\frac{\mathbb{B}^T \mathbb{B}}{n})} = o_p(1)$. ($\mathbb{B} = \mathbb{B} \Sigma_0^{1/2}$, $\text{tr}(\mathbb{B}^T \mathbb{B}/n) = \left\| \mathbb{B} \right\|_F^2/n = \sum_{j=1}^m \sum_{i=1}^n 1/n \sum_{k=1}^{k_0} B_{k0}(t_j)^2 = \sum_{j=1}^m \sum_{k=1}^{k_0} B_{k0}(t_j)^2 < \infty$). Also using sub multiplicativity of norms, $\left\| \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} - \left(\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right)^{-1} \right\|_2 \leq \left\| \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} \right\|_2 \left\| \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} - \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right\|_2 \left\| \left(\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right)^{-1} \right\|_2$.

The middle term $\left\| \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} - \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right\|_2 \leq |\tilde{\tau}_0^* - \tau_0| \left\| \frac{\mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right\|_2 + |\tau_0| \left\| \frac{\mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} - \frac{\mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right\|_2 \leq |\tilde{\tau}_0^* - \tau_0| \text{tr}(\mathbb{B}^T \mathbb{B}/n) \left\| \Sigma_0^{1/2} \right\|_F^2 \left\| \hat{\Sigma}^{-1} \right\|_2 + |\tau_0| \text{tr}(\mathbb{B}^T \mathbb{B}/n) \left\| \Sigma_0^{1/2} \right\|_F^2 \left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2 = o_p(1)$, as

$\left\| \hat{\Sigma}^{-1} - \Sigma^{-1} \right\|_2$ and $|\tilde{\tau}_0^* - \tau_0|$ are $o_p(1)$ (by assumption) and rest of the terms are bounded as

shown earlier. We have used the fact that $\left\| \hat{\Sigma}^{-1} \right\|_2$ is bounded because minimum eigenvalue

of $\hat{\Sigma}$ is > 0 . This is because $\hat{\Sigma}$ is block diagonal, i.e, $\hat{\Sigma} = \text{diag} \{ \hat{\Sigma}_{m \times m}, \hat{\Sigma}_{m \times m}, \dots, \hat{\Sigma}_{m \times m} \}$ and

$\hat{\Sigma}(s, t) = \sum_{k=1}^K \lambda_k \hat{\phi}_k(s) \hat{\phi}_k(t) + \hat{\sigma}^2 I(s = t)$. All the inequalities follow using application of tri-

angle inequality and sub multiplicativity of norms. For showing $\left\| \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} \right\|_2$ and

$\left\| \left(\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right)^{-1} \right\|_2$ are bounded, it suffices to show minimum eigenvalues of $\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n}$

and $\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n}$ are positive. Now $\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} = \frac{\mathbb{I}}{n} + \frac{n \tilde{\tau}_0^* \mathbb{C}_{0 \times m}^T \hat{\Sigma}_{m \times m}^{-1} \mathbb{C}_{0 \times m}}{n} = \frac{\mathbb{I}}{n} +$

$\tilde{\tau}_0^* \mathbb{C}_{k_0 \times m}^T \hat{\Sigma}_{m \times m}^{-1} \mathbb{C}_{0m \times k_0}$. Here $\mathbb{C}_0 = \mathbb{B}_0 \Sigma_0^{1/2}$ is full column rank, so $\mathbb{C}_{0k_0 \times m}^T \hat{\Sigma}_{m \times m}^{-1} \mathbb{C}_{0m \times k_0}$ is p.d matrix (not depending on n) and therefore minimum eigenvalue of $\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n}$ is positive. In similar way it can be shown, minimum eigenvalues of $\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n}$ are positive. This completes the proof that $\| \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} - \left(\frac{\mathbb{I}}{n} + \frac{\tau_0 \mathbb{B}^T \Sigma^{-1} \mathbb{B}}{n} \right)^{-1} \|_2 = o_p(1)$.

So we have shown all the terms in r.h.s of (A.1) converge to the corresponding counterparts with true values $\boldsymbol{\theta}_0$ and Σ . Note that all the individual terms are bounded, i.e., $\| \frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{Z}}{n} \|_2$, $\| \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} \|_2$ are bounded with the bound not depending on n . These can be shown exactly in the same way using sub multiplicativity of spectral norms, relation between spectral and Frobenius norm and the bounds established during the proof of the convergence part. So by repeated use of triangle inequality from (A.1) we have $\| \frac{\mathbb{Z}^T \hat{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n} \|_2 = o_p(1)$.

Next to complete the proof we have to show $\| \frac{\hat{\mathbb{V}}^{-1} \mathbb{Z} \mathbb{Z}^T \hat{\mathbb{V}}^{-1}}{n} - \frac{\mathbb{V}^{-1} \mathbb{Z} \mathbb{Z}^T \mathbb{V}^{-1}}{n} \|_2 = o_p(1)$ and this is again done using similar kind of techniques. Let us denote $\frac{\mathbb{V}^{-1} \mathbb{Z}}{\sqrt{n}} = \mathbb{A}$, $\frac{\hat{\mathbb{V}}^{-1} \mathbb{Z}}{\sqrt{n}} = \tilde{\mathbb{A}}$. Then we are trying to show $\| \tilde{\mathbb{A}} \tilde{\mathbb{A}}^T - \mathbb{A} \mathbb{A}^T \|_2 = o_p(1)$. Now using triangle and norm inequalities we have $\| \tilde{\mathbb{A}} \tilde{\mathbb{A}}^T - \mathbb{A} \mathbb{A}^T \|_2 \leq \| \tilde{\mathbb{A}} \|_2 \| \tilde{\mathbb{A}}^T - \mathbb{A}^T \|_2 + \| \tilde{\mathbb{A}}^T - \mathbb{A}^T \|_2 \| \mathbb{A} \|_2$. It suffices to show that $\| \tilde{\mathbb{A}} \|_2, \| \mathbb{A} \|_2$ are bounded and $\| \tilde{\mathbb{A}}^T - \mathbb{A}^T \|_2 = o_p(1)$. Note $\| \mathbb{A} \|_2 = \| \mathbb{A}^T \|_2 = \| \frac{\mathbb{Z}^T \hat{\mathbb{V}}^{-1} \mathbb{Z}}{\sqrt{n}} \|_2$ and thus we have

$$\frac{\mathbb{Z}^T \hat{\mathbb{V}}^{-1}}{\sqrt{n}} = \frac{\mathbb{Z}^T \hat{\Sigma}^{-1}}{\sqrt{n}} - \tilde{\tau}_0^* \left(\frac{\mathbb{Z}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right) \left(\frac{\mathbb{I}}{n} + \frac{\tilde{\tau}_0^* \mathbb{B}^T \hat{\Sigma}^{-1} \mathbb{B}}{n} \right)^{-1} \left(\frac{\mathbb{B}^T \hat{\Sigma}^{-1}}{\sqrt{n}} \right). \quad (\text{A.2})$$

Now as in (A.1), it can be shown similarly that all the terms in r.h.s of (A.2) are bounded and each term converges to the corresponding counterpart with τ_0 and Σ in place of $\tilde{\tau}_0^*$ and $\hat{\Sigma}$. This would then imply $\| \tilde{\mathbb{A}} \|_2, \| \mathbb{A} \|_2$ are bounded and $\| \tilde{\mathbb{A}}^T - \mathbb{A}^T \|_2 = o_p(1)$. So we have $\| \frac{\hat{\mathbb{V}}^{-1} \mathbb{Z} \mathbb{Z}^T \hat{\mathbb{V}}^{-1}}{n} - \frac{\mathbb{V}^{-1} \mathbb{Z} \mathbb{Z}^T \mathbb{V}^{-1}}{n} \|_2 = o_p(1)$ and along with $\| \frac{\mathbb{Z}^T \hat{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n} \|_2 = o_p(1)$, this proves $S_{\tau_1, n}(\tilde{\boldsymbol{\theta}}) \xrightarrow{d} S_{\tau_1, n}(\boldsymbol{\theta}_0)$. The proof $\Lambda_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \Lambda_n(\boldsymbol{\theta}_0)$ is again based on using similar set of techniques applied so far. We note $\Lambda_n(\boldsymbol{\theta}_0) = \Lambda(\boldsymbol{\theta}_0)/n^2 = \frac{1}{2} \text{tr} \{ (\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}/n)^2 \} - \frac{[\frac{1}{2} \text{tr} \{ (\mathbb{B}^T \mathbb{V}^{-1} \mathbb{Z}/n)(\mathbb{B}^T \mathbb{V}^{-1} \mathbb{Z}/n)^T \}]^2}{\frac{1}{2} \text{tr} \{ (\mathbb{B}^T \mathbb{V}^{-1} \mathbb{B}/n)^2 \}}$. It was already shown that $\| \frac{\mathbb{Z}^T \hat{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n} \|_2 = o_p(1)$, similarly it can be shown $\| \frac{\mathbb{B}^T \hat{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{B}^T \mathbb{V}^{-1} \mathbb{Z}}{n} \|_2 = o_p(1)$ and $\| \frac{\mathbb{B}^T \hat{\mathbb{V}}^{-1} \mathbb{B}}{n} - \frac{\mathbb{B}^T \mathbb{V}^{-1} \mathbb{B}}{n} \|_2 = o_p(1)$; which along with the facts, eigenval-

ues of square of a matrix are squares of it's eigenvalues, and trace is the sum of eigenvalues, proves $\Lambda_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \Lambda_n(\boldsymbol{\theta}_0)$. Therefore by Slutsky's theorem $T_S(\tilde{\boldsymbol{\theta}}, \hat{\Sigma}) \xrightarrow{d} T_S(\boldsymbol{\theta}_0, \Sigma) \stackrel{d}{=} (1/2)^2 \frac{(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell)^2}{\Lambda_n(\boldsymbol{\theta}_0)} I(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell)$, where $x_\ell \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ and λ_ℓ are eigenvalues of $\frac{\mathbb{Z}^T \mathbb{V}(\boldsymbol{\theta}_0, \Sigma)^{-1} \mathbb{Z}}{n}$.

As $\|\frac{\mathbb{Z}^T \tilde{\mathbb{V}}^{-1} \mathbb{Z}}{n} - \frac{\mathbb{Z}^T \mathbb{V}^{-1} \mathbb{Z}}{n}\|_2 = o_p(1)$ and $\Lambda_n(\tilde{\boldsymbol{\theta}}) \xrightarrow{p} \Lambda_n(\boldsymbol{\theta}_0)$, we approximate the null distribution using $(1/2)^2 \frac{(\sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell)^2}{\Lambda_n(\tilde{\boldsymbol{\theta}})} I(\sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \tilde{\lambda}_\ell)$, where $\tilde{\lambda}_\ell$ are eigenvalues of $\frac{\mathbb{Z}^T \mathbb{V}(\tilde{\boldsymbol{\theta}}, \hat{\Sigma})^{-1} \mathbb{Z}}{n}$ and $x_\ell \stackrel{iid}{\sim} \mathcal{N}(0, 1)$ for $\ell = 1, 2, \dots, k_1$.

A.3 Asymptotic distribution of p-value under null

Case 1: $a = 0$

$$\begin{aligned} P(T_S(\tilde{\boldsymbol{\theta}}) \geq a) &\rightarrow P(T_S(\boldsymbol{\theta}_0) \geq a) \\ &= P(T_S(\boldsymbol{\theta}_0) \geq 0 | S_{\tau_1}(\boldsymbol{\theta}_0) < 0)(1 - \alpha) + P(T_S(\boldsymbol{\theta}_0) \geq 0 | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0)\alpha \\ &= P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq 0 | S_{\tau_1}(\boldsymbol{\theta}_0) < 0\right)(1 - \alpha) \\ &\quad + P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq 0 | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right)\alpha \\ &= 1 \times (1 - \alpha) + 1 \times \alpha = 1. \end{aligned}$$

where $\alpha = P_{H_0}(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) = P_{H_0}(S_{\tau_1, n}(\boldsymbol{\theta}_0) \geq 0) = P(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell)$.

Case 2: $a > 0$

$$\begin{aligned} P(T_S(\tilde{\boldsymbol{\theta}}) \geq a) &\rightarrow P(T_S(\boldsymbol{\theta}_0) \geq a) \\ &= P(T_S(\boldsymbol{\theta}_0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) < 0)(1 - \alpha) + P(T_S(\boldsymbol{\theta}_0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0)\alpha \\ &= P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) < 0\right)(1 - \alpha) \\ &\quad + P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) > 0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right)\alpha \\ &= 0 \times (1 - \alpha) + \alpha \times P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right) \end{aligned}$$

$$\begin{aligned}
&= \alpha \times P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a | S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right) \\
&= \alpha \times \frac{P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a \cap S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right)}{P(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0)} \\
&= \alpha \times \frac{P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a \cap S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right)}{\alpha} \\
&= P\left(\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) \geq a \cap S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0\right) \\
&= P(S_{\tau_1}^2(\boldsymbol{\theta}_0) \geq a\Lambda(\boldsymbol{\theta}_0) \cap S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) = P(S_{\tau_1, n}(\boldsymbol{\theta}_0) \geq \sqrt{a\Lambda_n(\boldsymbol{\theta}_0)}) \\
&= P\left(1/2\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 - \sum_{\ell=1}^{k_1} \lambda_\ell\right) \geq \sqrt{a\Lambda_n(\boldsymbol{\theta}_0)}\right) \\
&= P\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell + 2\sqrt{\Lambda_n(\boldsymbol{\theta}_0)}\sqrt{a}\right) \leq P\left(\sum_{\ell=1}^{k_1} \lambda_\ell x_\ell^2 \geq \sum_{\ell=1}^{k_1} \lambda_\ell\right) = \alpha.
\end{aligned}$$

From this intuitively it is clear that under null p-value asymptotically takes value from a mixture distribution of uniform(0, α) and a degenerate distribution at one. Next we prove this fact. Again all the calculations are done assuming null is true.

Let P_v stand for p-value. We can define the p-value as $P_v(c) = P(T_s > T_{obs} | T_{obs} = c)$.

Then from the previous discussion it follows that

$$P(P_v(T_{obs}) = 1) = P(T_{obs} = 0) = P(S_{\tau_1}(\tilde{\boldsymbol{\theta}}) \leq 0) \rightarrow P(S_{\tau_1}(\boldsymbol{\theta}_0) \leq 0) = 1 - \alpha.$$

Recall that $T_s(\boldsymbol{\theta}_0) = \frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)} I(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0)$. So for $b > 0$, $P(T_s \geq b) = \alpha \overline{G(b)}$, where $G(b)$ is distribution function of the continuous part of T_S namely $\frac{S_{\tau_1}^2(\boldsymbol{\theta}_0)}{\Lambda(\boldsymbol{\theta}_0)}$ and $\overline{G(b)} = 1 - G(b)$.

Hence we have for $0 \leq d \leq \alpha < 1$,

$$\begin{aligned}
P(p_v(T_{obs}) \leq d) &= P(P(T_S \geq T_{obs}) \leq d) \\
&= P(P(T_S \geq T_{obs}) \leq d | T_{obs} > 0) P(T_{obs} > 0) \\
&\rightarrow P(P(T_S \geq T_{obs}) \leq d | T_{obs} > 0) \alpha.
\end{aligned}$$

The last convergence follows since $P(T_{obs} > 0) = P(S_{obs} > 0) \rightarrow P(S_{\tau_1}(\boldsymbol{\theta}_0) > 0) = P(S_{\tau_1}(\boldsymbol{\theta}_0) \geq 0) = \alpha$, asymptotically under null, as $\tilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$ and $S_{\tau_1}(\cdot)$ is a continuous random variable. Thus it is easy to see that

$$\begin{aligned} P(p_v(T_{obs}) \leq d) &\rightarrow P(\alpha \overline{G(T_{obs})} \leq d)\alpha \\ &= P(\overline{G(T_{obs})} \leq d/\alpha)\alpha \\ &= (d/\alpha)\alpha = d. \end{aligned}$$

The last line follows because $\left(\overline{G(T_{obs})}\right) \sim U(0, 1)$, as G is a continuous distribution function. So under null p-value is asymptotically an $\alpha:(1 - \alpha)$ mixture of uniform(0, α) and degenerate distribution at one.