Robust Estimation and Selection for Semivarying Coefficient Models

by

Mark Izuchukwu Uzochukwu

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

Auburn, Alabama
August 6, 2022

Keywords: Adaptive lasso, Wilcoxon rank estimation, asymptotic normality

Approved by

Asheber Abebe, Chair, Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics
Guanqun Cao, Assistant Professor of Mathematics and Statistics
Khodadadi M. Jeyhoon, Professor of Mechanical Engineering

Abstract

Varying coefficient models have gained popularity in recent years due to their flexibility in modeling more realistic problems. On the other hand, parametric models provide better interpretability. Model selection can be performed for both types of models. This dissertation is focused on estimation and variable selection for a semiparametric combination of the two types of models known as the semi-varying coefficient model. This model provides a flexible way to deal with various problems including problems that require spatiotemporal models. The approach used in this dissertation is based on rank estimation which provides a good balance between robustness and efficiency.

First, we consider a rank-based estimation of the varying coefficient functions for semi-varying coefficient model. The consistency and asymptotic normality of the proposed estimators are established. An extensive Monte-Carlo simulation study demonstrates the robustness and the efficiency of the proposed estimators compared to the least squares estimators. A backfitting algorithm is developed for estimating the parametric and nonparametric parts of the model in alternate steps. The semi-varying coefficient model was motivated by the popular COVID-19 where the rank-based estimation is used to provide accurate estimates of factors affecting the mortality rate. We use a real data example to show that the classical approach is highly affected by outliers in response space but not the rank-based method we propose in this dissertation. This is followed by variable selection method for semi-varying coefficient model. We develop a LASSO-type rank-based variable selection procedure to select and estimate coefficient functions.

Acknowledgments

Foremost, I would love to thank God Almighty for good health, patience and resilience to be able to round up this program.

I wish to express my sincere gratitude to my supervisor, Dr. Asheber Abebe who was always there and offered invaluable assistance. My sincere gratitude to my committee members Dr. Bertram Zinner, Dr. Peng Zeng, Dr. Guangun Cao and Dr. Jay Khodadadi, without whose knowledge and assistance this study would not have been a success and also Dr. Huybrechts Bindele, who's knowledge I was able to tap from.

To my wife and best friend, Chinememma your support throughout this work is inestimable. My friends Masuzyo and Ekene who also supported me when I needed them, with this medium, I appreciate you.

I will never forget my friends and colleagues Dr. Rachidi Salako, Dr. Fauziya Yakasai, Dr. Mariam Khachatran and Dr. Amarakristi Onyido who encouraged and supported me throughout this program.

My family, Ginika, Chidi, Chinwe, Emeka and Ogechukwu, I appreciate your love, I see you and I love you too.

Finally, I dedicate this thesis to my parents, Mr. and Mrs. Uzochukwu, who have always been there for me in every way possible, to make sure I got through this stage of education.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the most important properties in the estimation of a statistical model is robustness. This is because robustness underlies the strength of the statistical models including the tests and procedures that are involved in the complete statistical analysis of a model. For example, the mean is a less robust measure of central tendency compared to the median because the mean is very sensitive to extreme values at either tails of the distribution. Popular estimation procedures of statistical model such as linear regression, which emphasize the mean, assumes fixed parameters and that the data follow well-behaved distributions. In the absence of fixed parameters, other semi- and non-parametric models have emerged such as the varying coefficient model (VCM), the semi-varying coefficient model (SVCM), and the single-index varying coefficient model (SIVCM) that retain certain linearity and/or additive structures. Recent work has focused on robust estimation of such models are becoming more popular/useful due to the rise of big data that typically follows no known distribution.

The varying coefficient model (VCM) was pioneered by Cleveland, Grosse and Shyu (1991) to extend the applications of local regression techniques from one-dimensional to multi-dimensional setting. VCMs are generalized regression models whose coefficients vary as smooth function of other variables. As a variant of the VCMs, the semi-parametric or semi-varying coefficient model (SVCMs) allows for both fixed components and varying components that changes with other variables. The least squares (LS) method can be applied to both the VCM and SVCM. However, as in simple regression models, the presence of "outliers" imposes an unusual influence on the "least squares" estimators. There exist other methods that are not as sensitive to outlier such as the M-estimators, R-estimators, to mention a

few. The rank method is an example of a non-parametric method of estimation based on the minimization of a metric that uses the ranks of the errors.

This dissertation focuses on applying the rank estimation procedure in the estimation of the fixed and varying coefficients in the SVCM. The Rank-based estimators were developed as a robust, nonparametric substitute to the classical likelihood or least squares estimators. The rank regression was first introduced by Jureckova (1971) as the zero of a rank-score function and Jaeckel (1972) as a minimizer of a rank dispersion function. Just over half a decade later, McKean and Hettmansperger (1978) developed a Newton step algorithm that led to viable computation of the rank-based estimates. Many authors have since then, worked to develop complete rank-based inference for linear models that were based on the rank-estimates, in ways analogous the way that traditional analysis is based on Least Squares estimates. Presently, the rank-based analysis is known to be a complete analysis parallel to the traditional least squares for general linear models. Rank regression works in a way similar to the way LS works, but instead of using the Euclidean distance, a different kind of distance function is used that is based on Jaeckel's dispersion function which we will discuss later in more detail. That is not all. The only assumption on the distribution of the errors for rank method is that it is continuous, which is one of the major improvements over LS and a reason why it provides robust estimators.

Our work follows the Hastie and Tibshirani (1993) model of varying coefficients. We follow the model in the context of the SVCM, which in addition to the varying coefficient part of the model, there is a non varying part, where the parameters are fixed. We apply the rank method to the SVCM to estimate the varying coefficients as well as the fixed coefficients, in addition, we also use adaptive group lasso to select the relevant variables in the parametric and nonparametric parts.

To the best of our knowledge, this dissertation is the first to investigate how the rank method applied to the estimation and variable selection of SVCMs accounts for the behavior of "outliers" in heavy-tailed error distributions. The rank estimators are found to be

asymptotically normal and consistent. In a simulation study, we used three error distributions, the standard normal, the t-distribution with 3 degrees of freedom (for heavy tails) and the contaminated normal distribution (for contamination) to compare rank and least squares estimates of fixed and varying coefficients of SVCMs. In doing this, we compare the estimated functions of rank and LS estimation as well as the true and false positives in the selection of the fixed coefficient part of the model for both methods. Our results reveal that the rank method is found to perform better than the LS method of estimation when the error distribution is heavy-tailed or contaminated. To further demonstrate findings from the analysis, we apply both approaches to a high-frequency spatial COVID-19 data for empirical comparison. We apply a LASSO-type penalty for variable selection to select the important fixed and time-varying variables from the data. The empirical validation results demonstrate that the rank method provides better estimates that the LS method of estimation for the high-frequency spatial COVID-19 data.

This dissertation adds to the literature in the robust estimation of semiparametric regression models, particularly semi-varying coefficient models that are useful for spatiotemporal modeling. Moreover, the dissertation demonstrates that the rank method is useful for estimating data from complex structures such as global COVID-19 data. This is further affirmation that the rank-based estimation and selection approach provides a useful alternative to the least squares based estimation and selection approach for SVCMs.

The rest of this dissertation is structured as follows: Chapter 2 gives detailed background information on the estimation of semiparametric models, including recent works in rank-based estimation and selection. Chapter 3 focuses on the theory of the rank method in estimating the unknown coefficients. Consistency, asymptotic normality, and selection consistency of the proposed estimators are proved. In Chapter 4, Monte Carlo simulations are done to show the edge of rank estimation over LS estimation. Chapter 5 gives an application of the proposed approach on a global high-frequency spatial COVID-19 data.

Chapter 2

Background

It is common knowledge that parametric statistical inference is often accompanied by some model assumptions, most importantly, linearity. Besides parsimony, this is helpful in efficiency of the computation of the model estimates. So much work has been put into linear models which makes their properties widely established, but in reality, there are only a few places where they can be applied. With the wrong assumptions, a researcher can be met with inconsistencies in the model, especially large bias. To solve this issue, many variants of parametric models have been proposed but each one with its own setbacks.

Nonparametric models, however, need no assumption(s) on the model specifications, but may fail to incorporate some prior information, leading to estimators with large variance. An example is the "curse of dimensionality" which poses a huge setback for non-parametric models when the data has excessively high dimension. Aiming at fixing this limitation, Huber (1985) proposed the projection pursuit method, Li (1991) the sliced inverse regression method, Hardle and Stoker (1990) the single index models, to name a few. In theses methods, the response variable is modelled as an unknown function of the covariates and unknown parameters.

$$y = g(X^T\beta_1, \cdots, X^T\beta_q, \epsilon), \tag{2.1}$$

where y is the response, X, a p-dimensional matrix of covariates and ϵ, the error. The model is very impractical especially if $q \geqslant 3$, in fact, there are several limitations that surround model (2.1) that even its interpretabilty can even be an issue.

To help mitigate this problem surrounding the traditional parametric models, an alternative approach is to relax some of the conditions imposed. To that effect, Breiman and Friedman (1985) and Hastie and Tibshirani (1990) proposed additive models, where the

response y is related to a sum of univariate (unkown) functions of each of the covariates. Hastie and Tibshirani (1993), Fan and Zhang (1999, 2000), Chang et al. (2001), introduced varying coefficient models, where the regression coefficients are smooth functions of a variable. Wong et. al. (2008) proposed the single-index varying coefficient model, where the regression coefficients are univariate (unknown) functions of linear combination of several variables. Friedman (1991), Gu and Wahba (1992), and Stone et al. (1997) proposed the low-dimensional interaction models, and Wahba (1984), Green and Silverman (1994), partially linear models, to name a few.

Among the semiparametric models just mentioned, the varying coefficient models arise in many applications. They have been successfully used in multi-dimensional nonparametric regression, nonlinear time series models, generalized linear models, analysis of longitudinal, functional, and survival data, and financial and economic data. When coefficients are varying with time, the varying coefficient model has the form

$$Y_i = \mathbf{X}_i^T \boldsymbol{\alpha}(t_i) + \epsilon_i \ , \tag{2.2}$$

where $\epsilon_i \in \mathbb{R}$ represent random errors, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \alpha_2(\cdot), ..., \alpha_p(\cdot))^T \in \mathbb{R}^p$ is an unknown smooth function vector, and $\mathbf{X}_i$ is a matrix of time varying covariates.

It is important to note that this type of model can be adjusted to fit a scenario where some covariates are seen to change with time while others are fixed and are said to be time-invariant. This is precisely the model of interest in this dissertation. The motivation for this is the analysis of excess mortality from COVID-19 based on a global high-frequency spatiotemporal data. Excess mortality is assumed to depend on several variables where some are not time variant and only changes with the country considered while others varied with time for each specific country. For this COVID-19 excess mortality analysis, the data contained variables that changed with time including the number of vaccinated people, number of deaths, number of COVID-19 cases, and many more. There were also variables that were assumed to be unchanged over the relatively short period of the COVID-19 pandemic

including the proportion of people above 70 years of age, the country's GDP per capita, median age and so on. The semi-varying coefficient model studied in this dissertation allows for incorporating fixed-effects variables and time varying coefficients in the same structure.

The semi-varying coefficient model studied in this work, has the form

$$Y_i = \mathbf{X}_i^T \boldsymbol{\alpha}(t_i) + \mathbf{Z}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{2.3}$$

where $\epsilon_i \in \mathbb{R}$, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \alpha_2(\cdot), ..., \alpha_p(\cdot))^T \in \mathbb{R}^p$ is an unknown smooth function vector, $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)^T$ the constant coefficient vector, with true values $\boldsymbol{\alpha}_0(\cdot)$ and $\boldsymbol{\beta}_0$, respectively. $\mathbf{X}$ is the matrix of time varying covariates, while $\mathbf{Z}$ is the matrix of time invariant covariates. To understand why this model is of interest, we used the response Y_i as the mortality, to show it's relationship with the covariates $(\mathbf{X}, \mathbf{Y})$
which we discuss in detail in chapter 4.

Over the years, regression models have been commonly used in analysis of data taken from numerous sources such as geostatistics that arise often in environmental studies. For example, given data $(y_i, \mathbf{x}_i)$, where $i = 1, 2, 3, \ldots, n$, a common regression problem is to investigate a relationship between y_i and $\mathbf{x}_i$, written, $y_i \sim \mathbf{x}_i$, or more formally,

$$y_i = f(\boldsymbol{\beta}_0, \mathbf{x}_i) + \epsilon_i,$$

with $n = 1, 2, 3, \ldots$, that is, estimating parameters $\boldsymbol{\beta}_0$ that relate y_i and $\mathbf{x}_i$ via a function f, which may be linear or nonlinear and ϵ_i is an error which follows some distribution, usually unknown. A fascinating approach to this problem is by finding $\boldsymbol{\beta}$ so that the residuals become as small as possible.

Classical methods like the Least Squares (LS), which are easy to interpret and computationally convenient, minimize the squared errors, $\sum_{i=1}^{n}(y_i - f(\boldsymbol{\beta}_0, \mathbf{x}_i))^2$, resulting in parameters that are sensitive, if there are outliers in the data and so is not a reliable method. The simplest form is given by the model $y_i = \boldsymbol{\beta}_0 + \epsilon_i$, where the LS parameter estimate

$\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}_0$ is $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$, the mean response, and it is common knowledge that the mean, unlike the median, is very sensitive to outliers. Methods like the M-Estimators, an example being the Least Absolute Deviation, which minimizes the deviation in absolute values, that is, $\sum_{i=1}^{n} |y_i - f(\boldsymbol{\beta}_0, \mathbf{x}_i)|$, and R-Estimates have been introduced to overcome this setback encountered with least squares. Even at that, they are still found to be affected by heavy-tailed distributions.

Huber and others went ahead to develop the theory of M estimators to mitigate the setbacks encountered when using LS, and rank-based (R) estimation techniques were only used for simple problems like location comparisons for two-sample problems. Later Jaeckel (1972), Hettmansperger & McKean (1998) and others showed that R estimators, also called Wlicoxon estimators can be obtained by minimizing

$$\sum_{i=1}^{n} \varphi\bigg(R(e_i(\beta))\bigg) e_i(\beta), \tag{2.4}$$

where $R(e_i(\beta))$ is the rank of $e_i(\beta)$ among $i = 1, 2, 3, \ldots, n$, and $\varphi(\cdot)$ is a score function. Although the classical M and R estimation techniques are not robust, weights are introduced in these methods to help improve the results by making them robust.

To understand more about the rank method, from model (2.2), we get

$$\epsilon_i = Y_i - \mathbf{X}_i^T \boldsymbol{\alpha}(t_i) - \mathbf{Z}_i^T \boldsymbol{\beta},$$

and minimize the dispersion of the residuals ϵ_i. We focus particularly on the rank-based dispersion function of Jaeckel (1972) described as follows. Let $\varphi_n(i)$, $i = 1, 2, \cdots, n$ be a non-decreasing set of scores, not all equal, such that

$$\sum_{i=i}^{n} \varphi_n(i) = 0.$$

The rank dispersion function is defined as

$$D(\mathbf{e}) = \sum_{i=1}^{n} \varphi_n\big(R(e_i)\big)e_i. \tag{2.5}$$

The function D is called a dispersion function which acts like the Euclidean norm in terms of least squares. Notice that we can write (2.2) as $Y_i = W_i^T \curlyvee + \epsilon$, with $\curlyvee = (\alpha(t_i), \beta)$ and $W_i = (\mathbf{X}_i, Z_i)^T$. The way in which we estimate the unknown functions makes it so that we can work with a linear model and since this is so, we can use the properties obtained by Jaeckel (1972) and studied in detail in the monograph by Hettmansperger and McKean (2011). With this set-up, a lot of work has been done on different model setups. Often the difficulty in using rank-based approaches is the computational burden but there are several approximate approaches in the linear models. For example, Sievers and Abebe (2004) used the rank method to estimate the linear regression coefficients using an iterated reweighted least squares algorithm. Using arguments from fixed-point theory, they showed that the iteration converged to the rank estimate regardless of sample size. Abebe, McKean, Kloke and Bilgic (2016) extended this approach to longitudinal data using an iterated reweighted generalized estimating equation (GEE) approach.

There has also been some recent work on the rank estimation of models with functional regression coefficients. Bindele, Abebe and Meyer (2018) used the rank method to estimate the index and functional parameters of single-index models. In particular, they considered the model

$$Y = g_0(\boldsymbol{\beta}_0^T \mathbf{x}) + \varepsilon$$

where both $\boldsymbol{\beta}_0$ and g_0 are unknown. Their approach used a leave-one-out Nadaraya–Watson estimator of g_0 with a resulting non-convex optimization for estimating $\boldsymbol{\beta}_0$ and g_0. They established asymptotic optimality results for the estimators and also studied the finite sample

performance of the proposed approach. Bindele, Abebe, and Zeng (2019) where they considered the single-index model and developed a penalized generalized signed-rank procedure for estimation and variable selection following a local linear approximation.

Sun et al. (2019) considered rank-based estimation of the index coefficient and the functional regression coefficients for the single-index varying coefficient model (SIVCM)

$$Y = g_0(\boldsymbol{\beta}_0^T\mathbf{x}) + g_1(\boldsymbol{\beta}_0^T\mathbf{x})z_1 + \cdots + g_p(\boldsymbol{\beta}_0^T\mathbf{x})z_p + \varepsilon$$

Note that the single index model is just the intercept model of SIVCM; that is, when $g_1 = \cdots = g_p \equiv 0$. They established the consistency and asymptotic normality of the rank estimator and performed an extensive Monte-Carlo simulation study to demonstrate the robustness and efficiency of the rank estimator compared to the least squares estimators. Their approach also used a local linear approximation.

Motivated by problems in deep-water fish ecology, Sun et al. (2021) again considered the single-index varying coefficient model (SIVCM)

$$Y = g_0(\boldsymbol{\beta}_0^T\mathbf{x}) + g_1(\boldsymbol{\beta}_0^T\mathbf{x})z_1 + \cdots + g_p(\boldsymbol{\beta}_0^T\mathbf{x})z_p + \varepsilon$$

and proposed a rank-based procedure for the estimation and selection of the functional regression coefficients. The estimation of the functional regression coefficients together with the selection procedure using a back-fitting type computational algorithm to minimize a rank-based objective function. In their approach, they employed a group LASSO penalty for selecting functional regression coefficients following a spline approximation. Since only functional coefficients are being selected, they demonstrate that this same approach can be used for the varying-coefficient models

$$Y = g_0(U) + g_1(U)z_1 + \cdots + g_p(U)z_p + \varepsilon$$

by simply taking $\boldsymbol{\beta}_0^T\mathbf{x}$ to be a known variable U and skipping the estimation of $\boldsymbol{\beta}_0$.

Bindele and Abebe (2015) considered a rank estimation of the partial linear regression model

$$Y = \boldsymbol{\beta}_0^T\mathbf{x} + g(t) + \varepsilon$$

in the context of response missing at random. They used kernel approximation to estimate g as a known function of t but unknown with respect to $\boldsymbol{\beta}_0$.

The work proposed in this dissertation extends and generalizes these to semi-parametric models that have fixed and varying components - semi-varying coefficient models. The first part of the model given in (2.3)

$$Y_i = \mathbf{X}_i^T\boldsymbol{\alpha}(t_i) + \mathbf{Z}_i^T\boldsymbol{\beta} + \epsilon_i$$

is the varying coefficient model and its second component is a linear model. We are interested in developing a rank procedure for estimating and selecting both the functional coefficients $\boldsymbol{\alpha}$ and the fixed coefficients $\boldsymbol{\beta}$ using an objective function that contains both group LASSO and regular LASSO terms.

In Chapter 3, we propose a general R estimation procedure for the semi-varying coefficient model that is a robust and more efficient alternative to the classical least squares method when data are contaminated with heavy-tailed error distributions, or when data contain outliers. We propose a backfitting algorithm by iterating between the fixed and varying coefficient functions following a B-spline approximation. We show that the resulting estimator is robust and asymptotically efficient when compared to the least squares estimator. We propose a robust procedure for estimating and selecting coefficient functions using group LASSO (GRPLASSO) and estimating and selecting the fixed parameters using LASSO. We demonstrate via a Monte Carlo experiment (Chapter 4) that the proposed procedure is more efficient than LS when the error distribution is not normal, and it performs as well as LS

under normal error distribution. It also provides better selection consistency that the LS for functional coefficients across all simulation scenarios.

Chapter 3

Semi-Varying Coefficient Model

3.1 Introduction

Throughout the years, considerable work has been done on robust procedures for linear models. Several classes of robust estimates have been proposed for these models. One is the rank class of estimates. This class uses an objective function which depends on the choice of a score function, φ. Generally this robust analysis is highly efficient relative to the LS analysis; see Hettmansperger and McKean (1998). Many interesting problems, in the real world are nonlinear in nature and traditional procedures based on LS estimates have been in use for many years. As these LS methods of non-linear models use the Euclidean norm, they are as easy to interpret as linear models. The asymptotic theory for nonlinear LS has been developed by Jennrich (1969) and Wu (1981), among others. In this chapter, we propose a nonlinear analysis based on the rank objective function. In section 2.2, we give a brief introduction to semi-varying coefficient models. In Section 2.3 we present a family of estimates for nonlinear models. In Section 3.4, we show that these estimates are strongly consistent under certain assumptions.

3.2 Semivarying Coefficient Model Setup

Let $i \in \mathbb{N}$ and let $(Y_i, \mathbf{X}_i, \mathbf{Z}_i, t_i)$ be such that $Y_i \in \mathbb{R}^1$, $\mathbf{X}_i \in \mathbb{R}^p$, $\mathbf{Z}_i \in \mathbb{R}^q$ and $t_i \in [t_0, t_1]$. We shall consider a semivarying coefficient model of the form

$$Y_i = \mathbf{X}_i^T \boldsymbol{\alpha}(t_i) + \mathbf{Z}_i^T \boldsymbol{\beta} + \epsilon_i, \tag{3.1}$$

where $\epsilon_i \in \mathbb{R}$ are random errors, $\boldsymbol{\alpha}(\cdot) = (\alpha_1(\cdot), \alpha_2(\cdot), ..., \alpha_p(\cdot))^T \in \mathbb{R}^p$ is an unknown smooth function vector, $\boldsymbol{\beta} = (\beta_1, ..., \beta_q)^T$ the constant coefficient vector, with true values $\boldsymbol{\alpha}_0(\cdot)$ and $\boldsymbol{\beta}_0$, respectively.

3.3 Rank Estimation of Semivarying Coefficient Model

We start by approximating $\alpha_k(t)$ using B-spline approximation. Let $\mathbb{S}^r_{C_n}(t)$ be the set of spline functions of order $r+1$ having knots $\mathbb{K} = \{t_0 = \tau_0 < \tau_1 < ... < \tau_{C_n} < \tau_{C_n+1} = t_1\}$. We say that $B(t) \in \mathbb{S}^r_{C_n}(t)$ if and only if $B(t) \in C^{r-1}[t_0, t_1]$, where it's restriction to $[\tau_k, \tau_{k+1}]$ is a polynomial of at most degree r. A piecewise constant function, linear spline and quadratic spline corresponds to $r = 0, 1$ and 2, respectively.

Let

$$B_k(t) = (\tau_k - \tau_{k-r-1})[\tau_{k-r-1}, ..., \tau_k](z-t)^r_+, \qquad k = 1, ..., q_n,$$

where $q_n = C_n + r + 1$, $[\tau_{k-r-1}, ..., \tau_k](z-t)^r_+$ denotes the $(r+1)^{th}$ order divided difference of the function $(z-t)^r_+$, $\tau_k = t_0$, when $k = -r, ..., -1$, and $\tau_k = t_1$, when $k = C_{n+2}, ..., C_{n+r+1}$. Then $\boldsymbol{\pi}(t) = (B_1(t), B_2(t), ..., B_{q_n}(t))^T$ forms a basis for $\mathbb{S}^r_{C_n}(t)$. For more about spline functions, see Schumaker (1981).

So we can approximate $\alpha_k(t)$ as follows:

$$\alpha_k(t) \approx \sum_{j=1}^{q_n} B_j(t)\theta_{k,j} = \boldsymbol{\pi}(t)^T \boldsymbol{\theta}_k,$$

where $B_j(t)$ are spline basis and $\theta_{k,j}$ are spline coefficients.

Now

$$Y_i \approx \sum_{k=1}^{p}\sum_{j=1}^{q_n} X_{i,k}B_j(t_i)\theta_{k,j} + \mathbf{Z}_i^T\boldsymbol{\beta} + \epsilon_i = \boldsymbol{\Pi}_i^T\boldsymbol{\Theta} + \mathbf{Z}_i^T\boldsymbol{\beta} + \epsilon_i,$$

with $\boldsymbol{\Pi}_i = (X_{i,1}\boldsymbol{\pi}_i^T, ..., X_{i,p}\boldsymbol{\pi}_i^T)^T \in \mathbb{R}^{pq_n}$, $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^T, ..., \boldsymbol{\theta}_p^T)^T$ and $\boldsymbol{\pi}_i = \boldsymbol{\pi}(t_i)$.

Given general rank scores of the form $a_\varphi(i) = \varphi(\frac{i}{n+1})$, we define the general rank estimator of the semivarying coefficient model parameters $(\boldsymbol{\Theta}, \boldsymbol{\beta})$, say $(\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\beta}})$, as a minimizer

of the rank dispersion function

$$D_n(\boldsymbol{\Theta}, \boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R(e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}))}{n+1}\Big) e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}), \tag{3.2}$$

where $e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}) = Y_i - \mathbf{\Pi_i}^T \boldsymbol{\Theta} - \mathbf{Z}_i^T \boldsymbol{\beta}$ are the model residuals, $R(e) \equiv \sum_{i=1}^{n} I\{e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}) \leqslant e\}$ is the rank of the model residuals, $\varphi : \mathbb{R} \to (0, 1)$ is a continuously differentiable and non-decreasing score function.

3.4 Partially Adaptive Group Penalized Rank Estimation

In the context of the approximate model

$$Y_i \approx \sum_{k=1}^{p} \sum_{j=1}^{q_n} X_{i,k} B_j(t_i) \theta_{k,j} + \mathbf{Z}_i^T \boldsymbol{\beta} + \epsilon_i = \mathbf{\Pi}_i^T \boldsymbol{\Theta} + \mathbf{Z}_i^T \boldsymbol{\beta} + \epsilon_i$$

with $\mathbf{\Pi}_i = (X_{i,1}\boldsymbol{\pi}_i^T, ..., X_{i,p}\boldsymbol{\pi}_i^T)^T \in \mathbb{R}^{pq_n}$, $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1^T, ..., \boldsymbol{\theta}_p^T)^T$ and $\boldsymbol{\pi}_i = \boldsymbol{\pi}(t_i)$, our goal is to select important variables in the parametric and non parametric parts. For the parametric part, we shall refer to Z_k as important if and only if $\beta_k \neq 0$. For the nonparametric part, we refer to $\mathbf{X}_k$ as important if and only if $\boldsymbol{\theta}_k \neq \mathbf{0}$. Let us suppose, without loss of generality that $\{\beta_{0k} \neq 0\}_{k=1}^{c}$ and $\{\beta_{0k} = 0\}_{k=c+1}^{q}$. Let $\boldsymbol{\beta}^* = (\beta_1, ..., \beta_c)^T$ with true value $\boldsymbol{\beta}_0^*$. Also, let $\{\alpha_{0l}\}_{l=1}^{v}$ be nonzero components of $\boldsymbol{\alpha}_0(t)$ with $\{\alpha_{0l}(t) \equiv 0\}_{l=v+1}^{p}$.

Again, given a general rank scores of the form $a_\varphi(i) = \varphi(\frac{i}{n+1})$, the partially adaptive group $L_r (r \geqslant 1)$ norm penalized estimator $(\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\beta}})$ can be obtained by minimizing

$$\begin{aligned} Q(\boldsymbol{\Theta}, \boldsymbol{\beta}) &= D_n(\boldsymbol{\Theta}, \boldsymbol{\beta}) + \sum_{k=1}^{p} \lambda_{1nk} \|\boldsymbol{\theta}_k\|_r + \sum_{k=1}^{q} \lambda_{2nk} |\beta_k| \\ &= \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R(e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}))}{n+1}\Big) e_i(\boldsymbol{\Theta}, \boldsymbol{\beta}) + \sum_{k=1}^{p} \lambda_{1nk} \|\boldsymbol{\theta}_k\|_r + \sum_{k=1}^{q} \lambda_{2nk} |\beta_k|, \end{aligned} \tag{3.3}$$

where λ_{1nk} and λ_{2nk} are tuning parameters that control model complexity of nonparametric and parametric parts, respectively.

3.5 Definitions and Assumptions

Before establishing the theoretical properties of the rank estimators, we give some definitions and assumptions needed to establish the results.

Define $\mathcal{H}_\gamma$ as the collection of all functions on $[T_0, T_1]$ whose d^{th} order derivative satisfies the Holder condition of order ν with $\gamma \equiv d + \nu$. That is, for any $h \in \mathcal{H}_\gamma$, there is a constant $c \in (0, \infty)$ such that for each $h \in \mathcal{H}_\gamma$, $|h^{(d)}(s) - h^{(d)}(t)| \leqslant c|s - t|^\nu$ for any $T_0 \leqslant s, t, \leqslant T_1$.

To facilitate theoretical discussion, we will write the model of interest as

$$Y_i = \mathbf{W}_i^T \curlyvee_0 + \epsilon_i,$$

where $\curlyvee_0 = (\boldsymbol{\Theta}_0, \boldsymbol{\beta}_0)$ is the true parameter and $\mathbf{W}_i = (\boldsymbol{\Pi}_i, \boldsymbol{Z}_i)^T$. We will write $D_n(\curlyvee)$ to represent the rank dispersion function and $\mathbf{S}_n(\curlyvee) = \nabla_\curlyvee D_n(\curlyvee)$ to represent the gradient of D_n. Moreover, we will let $\Phi = \varphi'$.

We will make the following assumptions:

A_1. $\{\alpha_k(t) \in \mathcal{H}_\gamma\}_{k=1}^v$ for some $\gamma > N + \frac{1}{2}$.

A_2. The random field $\{(Y_i, X_i, Z_i, t_i), i \in \mathbb{N}\}$ is strictly stationary. For all $i \neq j$ in $\mathbb{N}$, t_i and t_j admit a joint density $f_{i,j}$ satisfying $|f_{i,j}(t_1, t_2) - f(t_1)f(t_2)| \leqslant c_1$ for all $t_1, t_2 \in [T_0, T_1]$, where $c_1 > 0$ is a constant and f denotes the marginal density of T_i.

A_3. For all $i \in \mathbb{N}$, the random design vectors X_i and Z_i are bounded in probability, and the eigen values of $E(X_i X_i^T | T_i = t)$, $t \in [T_0, T_1]$ are bounded away from 0 and ∞ uniformly.

A_4. $\varphi\Big(\frac{R(\epsilon_i)}{n+1}\Big)\epsilon_i$ is convex and $E(\Phi(\epsilon_i)|X_i, Z_i, T_i) = 0$, for all $i \in \mathbb{N}$. In addition, for some $\delta > 0$, $\sup_{i \in \mathbb{N}} E(|\Phi(\epsilon_i)|^{2+\delta} | X_i, Z_i, T_i) < \infty$, and there exist positive numbers b_i with $0 < \inf b_i < \sup b_i < \infty$ such that $\sup_{i \in \mathbb{Z}} |E(\Phi(\epsilon_i + s))|X_i, Z_i, T_i) - b_i s| = O_p(s^2)$ as $s \to 0$, where Φ is the derivative of φ.

A_5. There exist constants $0 < c_2, c_3 < \infty$ such that $\sup_{i\in\mathbb{N}} E\{[\Phi(\epsilon_i+s)-\Phi(\epsilon_i)]^2|X_i, Z_i, U_i\} \leqslant c_3|s|$, as $s \to 0$ and $|\Phi(v+s)-\Phi(v)| \leqslant c_3$ for any $|s| \leqslant c_2$ and $v \in \mathbb{R}^1$.

A_6. Let $a_{1n} = \sup\{\lambda_{1nk} : k = 1, ..., v\}$, $a_{2n} = \sup\{\lambda_{2nk} : k = 1, ..., c\}$, $b_{1n} = \inf\{\lambda_{1nk} : k = v+1, ..., p\}$ and $b_{2n} = \inf\{\lambda_{2nk} : k = c+1, ..., q\}$ with $\max\{a_{1n}, a_{2n}\}/\sqrt{n} \to 0$, $\max\{a_{1n}, a_{2n}\}q_n^{1/2}/\sqrt{n} \to 0$, $n^{-\frac{1}{2}}b_{1n} \to \infty$ and $n^{-\frac{1}{2}}b_{2n} \to \infty$.

A_7. Suppose the error distribution has finite Fisher information; that is, $0 < I(f) < \infty$.

A_8. We will assume $\boldsymbol{\Theta} \in int(\mathbb{A}_1)$ and $\boldsymbol{\beta} \in int(\mathbb{A}_2)$, where $\mathbb{A}_1$ and $\mathbb{A}_2$ are compact subspaces of $\mathbb{R}^{pq_n}$ and $\mathbb{R}^q$, respectively.

For the conditions given above, A_1 is a condition of smoothness that helps to track the rate of convergence of the spline estimator with respect to the variable coefficients. A_2 is a standard condition used, for example, by Tran (1990), Hallin et al. (2004) and Tang (2014) in the spatial setting. Although our setting is just a sub-class of the spatial setting. A_3 was used in Huang et al. (2002) and is a technical condition to derive the optimal convergence rate of the estimators in variable coefficient setting. Conditions A_4 and A_5 are only on score function $\varphi(u)$ and the random error ϵ_i; same conditions were also used in Tang (2014) and Lu et al. (2014). A_6 is used for the convergence rate of tuning parameters. A_7 is used to put the model into a linear setting so as to establish some asymptotics. In particular, as seen in Hettemanspeger & McKean (2011), we see that

$$\frac{1}{\sqrt{n}}\mathbf{S}_n(\curlyvee) = \frac{1}{\sqrt{n}}\mathbf{S}_n(\curlyvee_0) - \tau_\varphi^{-1}\Sigma\sqrt{n}(\curlyvee_n - \curlyvee_0) + o_p(1),$$

with $\sqrt{n}(\curlyvee_n - \curlyvee_0) = O_p(1)$ and $\Sigma = \lim_{n\to\infty} n^{-1}\sum \mathbf{W}_i^T\mathbf{W}_i$. A_8 is a condition imposed on the parameters from which we get boundedness.

We will denote the nonzero parts of $\curlyvee_0$ and $\mathbf{W}_i$ as $\curlyvee_0^*$ and $\mathbf{W}_i^*$, respectively. We will comformably define $\Sigma^* = \lim_{n\to\infty} n^{-1}\sum \mathbf{W}_i^{*T}\mathbf{W}_i^*$.

3.6 Asymptotic Properties

Before giving the theorems on the asymptotic properties of the minimizer of $Q(\boldsymbol{\Theta}, \boldsymbol{\beta})$, we can assume that there exists a vector $\boldsymbol{\beta}^0 = (\beta_1^0, ..., \beta_q^0)^T$ such that $|\beta_k^0| \neq 0,\ k \in \{1, ..., c\}$; $|\beta_k^0|_1 = 0,\ k \in \{c+1, ..., q\}$. Moreover, under our assumptions, Wang (2016) shows that there exists a vector $\boldsymbol{\Theta}^0 = (\boldsymbol{\theta}_1^{0T}, ..., \boldsymbol{\theta}_p^{0T})^T$ such that $||\boldsymbol{\theta}_k^0||_1 \neq 0,\ k \in \{1, ..., v\}$; $||\boldsymbol{\theta}_k^0||_1 = 0$, $k \in \{v+1, ..., p\}$ and that

$$\sup_{t \in [T_0, T_1]} |\alpha_{0k}(t) - \boldsymbol{\pi}(t)^T \boldsymbol{\theta}_k^0| = O(q_n^{-\gamma}) \ ,$$

for $k = 1, \ldots, p$.

The theorem below establishes the estimation sparsity of the rank estimation and selection procedure for semi-varying coefficient models.

Theorem 3.1 *Assume the regularity conditions A_1 - A_7 hold and that $q_n = O(n^{1/(2\gamma+1)})$. Then, for $r = 1, 2$, $\hat{\beta}$ and $\hat{\alpha}(t)$ satisfy*

(a) $\hat{\alpha}_k(t) \equiv 0,\ k = v+1, ..., p$ holds with probability tending to 1;

(b) $\hat{\beta}_k = 0,\ k = c+1, ..., q$ holds with probability tending to 1.

Theorem 3.2 *(Estimation Efficiency)*

Suppose A_1. - A_4. hold. Then, the estimator $\hat{\gamma}$ is asymptotically consistent and

$$\sqrt{n}(\hat{\gamma}_n^* - \gamma_0^*) \xrightarrow{D} N(0, \tau_\varphi^2 \Sigma^{*-1}).$$

Proof of Theorem 3.1

We first establish the consistency of the rank estimator. By Lemma 1 of Wu (1981), the estimator is consistent if there exists a $C > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(\inf_{n^{1/2}||\curlyvee-\curlyvee_0||_2=C} \Big(Q(\curlyvee) - Q(\curlyvee_0)\Big) > 0 \right) = 1.$$

Equivalently,

$$\lim_{n\to\infty} \mathbf{P}\left(\inf_{n^{1/2}||\curlyvee-\curlyvee_0||_2=C} \Phi_n(\curlyvee) + \sum_{k=1}^{p} \lambda_{1nk}(||\boldsymbol{\theta}_k||_r - ||\boldsymbol{\theta}_k^0||_r) + \sum_{k=1}^{q} \lambda_{2nk}(|\beta_k| - |\beta_k^0|) > 0 \right) = 1,$$

where

$$\Phi_n(\curlyvee) = D_n(\curlyvee) - D_n(\curlyvee_0).$$

It suffices to prove that

$$\lim_{n\to\infty} \inf_{\curlyvee: n^{1/2}||\curlyvee-\curlyvee_0||_2=C} (Q(\curlyvee) - Q(\curlyvee_0)) > 0,$$

a.s, by Lemma 1 of Wu (1981). Observe that for any $r \geqslant 1$,

$$\Big| ||\theta_k||_r - ||\theta_k^0||_r \Big| \leqslant ||\theta_k - \theta_k^0||_r \leqslant ||\theta_k - \theta_k^0||_1 \leqslant q_n^{\frac{1}{2}} ||\Theta - \Theta^0||_2.$$

Let $A_1 = \{\Theta : \Theta = \Theta^0 + n^{-1/2}u_1, ||u_1||_2 < C_1\}$ and $A_2 = \{\beta : \beta = \beta^0 + n^{-1/2}u_2, ||u_2||_2 < C_2\}$. Observe that A_1 and A_2 are open sets containing Θ^0 and β_0, respectively. Since A_1 and A_2 are open in $\mathbb{A}_1$ and $\mathbb{A}_2$, respectively, their complements are closed in their respective supersets and for $\Theta \notin A_1$, we have that $||u_1||_2 \geqslant C_1$ and because A_1^c is a closed subset of a compact space $\mathbb{A}_1$ it is compact and therefore closed and bounded, so we have $\sum_{k=1}^{v} ||u_{1k}||_2 \leqslant ||u_1||_2 \leqslant M_1$. A similar argument can be used to show that $\sum_{k=1}^{v} ||u_{2k}||_2 \leqslant ||u_2||_2 \leqslant M_2$.

Now,

$$
\begin{aligned}
Q(\curlyvee) - Q(\curlyvee_0) &= \Phi_n(\curlyvee) + \sum_{k=1}^{p} \lambda_{1nk}(||\theta_k||_r - ||\theta_k^0||_r) + \sum_{k=1}^{q} \lambda_{2nk}(|\beta_k| - |\beta_{0k}|) \\
&= \Phi_n(\curlyvee) + \sum_{k=1}^{v} \lambda_{1nk}(||\theta_k||_r - ||\theta_k^0||_r) + \sum_{k=v+1}^{p} \lambda_{1nk}(||\theta_k||_r - ||\theta_k^0||_r) \\
&+ \sum_{k=1}^{c} \lambda_{2nk}(|\beta_k| - |\beta_k^0|) + \sum_{k=c+1}^{p} \lambda_{2nk}(|\beta_k| - |\beta_k^0|) \\
&\geqslant \Phi_n(\curlyvee) + \sum_{k=1}^{v} \lambda_{1nk}(||\theta_k||_r - ||\theta_k^0||_r) + \sum_{k=1}^{c} \lambda_{2nk}(|\beta_k| - |\beta_k^0|) \\
&\geqslant \Phi_n(\curlyvee) - a_{1n} \sum_{k=1}^{v} (||\theta_k||_r - ||\theta_k^0||_r) - a_{2n} \sum_{k=1}^{c} (|\beta_k| - |\beta_k^0|) \\
&\geqslant \Phi_n(\curlyvee) - a_{1n} n^{-1/2} q_n^{1/2} \sum_{k=1}^{v} ||\theta_k - \theta_k^0||_2 - a_{2n} n^{-1/2} \sum_{k=1}^{c} |\beta_k - \beta_k^0| \\
&\geqslant \Phi_n(\curlyvee) - a_{1n} n^{-1/2} q_n^{1/2} M_1 - a_{2n} n^{-1/2} M_2, \\
&\geqslant \Phi_n(\curlyvee) - \max\{a_{1n}, a_{2n}\} n^{-1/2} q_n^{1/2} M_1 - \max\{a_{1n}, a_{2n}\} n^{-1/2} M_2,
\end{aligned}
$$

Since $\lim_{n\to\infty} \max\{a_{1n}, a_{2n}\} n^{-1/2} q_n^{1/2} = 0$, to complete the proof it suffices to show that

$$
\lim_{n\to\infty} \inf_{\curlyvee : n^{1/2}||\curlyvee - \curlyvee_0||_2 = C} (D_n(\curlyvee) - D_n(\curlyvee_0)) > 0.
$$

This is shown to be true (using assumptions covered under $A_1 - A_8$) for general rank estimators of nonlinear models with compact parameter spaces in Abebe, McKean, and Bindele (2012). Therefore, we have

$$
\lim_{n\to\infty} P\left(\inf_{n^{1/2}||\curlyvee - \curlyvee_0||_2 = C} \Phi_n(\curlyvee) + \sum_{k=1}^{p} \lambda_{1nk}(||\theta_k||_r - ||\theta_k^0||_r) + \sum_{k=1}^{q} \lambda_{2nk}(|\beta_k| - |\beta_{0k}|) > 0 \right) = 1.
$$

Now, we show that:

(i) $\hat{\theta}_k = 0$, $k = v+1, ..., p$, holds with probability tending to 1 and

(ii) $\hat{\beta}_k = 0$, $k = c+1, ..., q$, holds with probability tending to 1.

To this end, it is enough to prove that

$\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,s}} < 0$, $\theta_{k,s} < 0$ and

$\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,s}} > 0$, $\theta_{k,s} > 0$.

We write the objective function Q as

$$\begin{aligned} Q(\Theta,\beta) &= \frac{1}{n}\sum_{i=1}^{n}\varphi\Big(\frac{R(Y_i - \Pi_i^T\Theta - \mathbf{Z}_i^T\beta)}{\hat{n}+1}\Big)(Y_i - \Pi_i^T\Theta - Z_i^T\beta) + \sum_{k=1}^{p}\lambda_{1nk}||\theta_k||_r + \sum_{k=1}^{q}\lambda_{2nk}|\beta_k| \\ &= D_n(\curlyvee) + \sum_{k=1}^{p}\lambda_{1nk}||\theta_k||_r + \sum_{k=1}^{q}\lambda_{2nk}|\beta_k|. \end{aligned}$$

Now let $r = 2$, and observe that $n^{-\frac{1}{2}}\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,j}}$,

$$n^{-\frac{1}{2}}\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,j}} = -n^{-\frac{1}{2}}S_{n,k,j}(\curlyvee) + n^{-\frac{1}{2}}\lambda_{1nk}\frac{\theta_{k,j}}{||\theta_k||_2}$$

Since $\sqrt{n}(\curlyvee_s - \curlyvee_{0s}) = O(1)$, it follows from the asymptotic linearity result in Hettmansperger and McKean (2011) that

$$\frac{1}{\sqrt{n}}\mathbf{S}_n(\curlyvee) = \frac{1}{\sqrt{n}}\mathbf{S}_n(\curlyvee_0) - \tau_\varphi^{-1}\Sigma\sqrt{n}(\curlyvee - \curlyvee_0) + o_p(1).$$

Thus

$$\begin{aligned} n^{-\frac{1}{2}}\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,j}} =& \frac{1}{\sqrt{n}}S_{k,j}(\curlyvee_0) - \tau_\varphi^{-1}\Sigma\sqrt{n}(\curlyvee - \curlyvee_0) + o_p(1) + n^{-\frac{1}{2}}\lambda_{1nk}\frac{\theta_{k,j}}{||\theta_k||_2} \\ =& O_p(1) + n^{-\frac{1}{2}}\lambda_{1nk}\frac{\theta_{k,j}}{||\theta_k||_2}, \end{aligned}$$

Notice that $\frac{O_p(1)}{n^{-\frac{1}{2}}b_{1n}} \to 0$, and $b_{1n} = \inf\{\lambda_{1nk} : k = v+1, \cdots, p\}$ so that the sign of $\frac{\partial Q(\Theta,\beta)}{\partial \theta_{k,j}}$ is completely determined by the sign of $\theta_{k,j}$ and therefore, they have the same sign since $n^{-\frac{1}{2}}\lambda_{1nk} \geqslant n^{-\frac{1}{2}}b_{1n} \to \infty$.

Similarly, for

$\frac{\partial Q(\Theta,\beta)}{\partial \beta_k} < 0$, $\beta_k < 0$ and
$\frac{\partial Q(\Theta,\beta)}{\partial \beta_k} > 0$, $\beta_k > 0$

we have by a similar argument as before, that

$$\begin{aligned} n^{-\frac{1}{2}}\frac{\partial Q(\Theta,\beta)}{\partial \beta_k} =& -n^{-\frac{1}{2}}S_{n,j}(\curlyvee) + n^{-\frac{1}{2}}\lambda_{2nk}\mathrm{sgn}(\beta_k), \\ =& O_p(1) + n^{-\frac{1}{2}}\lambda_{2nk}\mathrm{sgn}(\beta_k), \end{aligned}$$

since $\sqrt{n}(\curlyvee_s - \curlyvee_{0s}) = O_p(1)$,

Notice that $\frac{O_p(1)}{n^{-\frac{1}{2}}b_{2n}} \to 0$, so that the sign of $\frac{\partial Q(\Theta,\beta)}{\partial \beta_k}$ is completely detremined by $\mathrm{sgn}(\beta_k)$ because $n^{-\frac{1}{2}}\lambda_{2nk} \geqslant n^{-\frac{1}{2}}b_{2n} \to \infty$,

So for $\hat{\curlyvee}^{**} = (\hat{\Theta}^{**}, \hat{\beta^{**}})$ with $\Theta^{**} = (\theta_{v+1}, ..., \theta_p)$ and $\beta^{**} = (\beta_{c+1}, ..., \beta_q)$, we have that

$$\lim_{n\to\infty} P(\hat{\curlyvee}^{**} = \mathbf{0}) = 1.$$

Proof of Theorem 3.2

To show the asymptotic normality of the estimated parameters, from A_7, we have

$$n^{-\frac{1}{2}} S_{n,j}(\curlyvee_0^*) - \tau_\varphi^{-1} \Sigma \sqrt{n}(\hat{\curlyvee}_n^* - \curlyvee_0^*) + n^{-\frac{1}{2}} \lambda_n \mathrm{G}(\curlyvee_{k,s}) = o_p(1) \ ,$$

with $\lambda_{k,s} = \theta_{k,s}$ when the partial derivative is with respect to θ, $\lambda_{k,s} = \beta_k$ when the derivative is with respect to β, G is the derivative of the penalty term and $\lambda_n = \lambda_{1nk}$ or λ_{2nk}, depending on whether or not we are differentiating with respect to θ or β.
Now, observe that $n^{-\frac{1}{2}} \lambda_n \mathrm{sgn}(\curlyvee_{k,s}) \leqslant n^{-\frac{1}{2}} \max\{a_{1n}, a_{2n}\} \mathrm{sgn}(\curlyvee_{k,s}) \to 0$, so that $n^{-\frac{1}{2}} \lambda_n \mathrm{sgn}(\curlyvee_{k,s}) \to 0$ either way. Thus

$$n^{-\frac{1}{2}} S_{n,j}(\curlyvee_0^*) = \tau_\varphi^{-1} \Sigma^* \sqrt{n}(\hat{\curlyvee}_n^* - \curlyvee_0^*) + o_p(1).$$

Since, $n^{-\frac{1}{2}} S_n(\curlyvee_0^*) \xrightarrow{D} N(0, \tau_\varphi^2 \Sigma^{*-1})$, where $\Sigma^* = \lim_{n\to\infty} \frac{1}{n} W^{*T} W^*$ for $\curlyvee^* = (\Theta^*, \beta^*)$ with $\Theta^* = (\theta_1, ..., \theta_v)$ and $\beta^* = (\beta_1, ..., \beta_c)$ we have that

$$\sqrt{n}(\hat{\curlyvee}_n^* - \curlyvee_0^*) \xrightarrow{D} N(0, \tau_\varphi^2 \Sigma^{*-1}).$$

3.7 Estimation Algorithm

In this section, we discuss a backfitting algorithm to implement the rank estimation and selection procedures for the coefficients of semi-varying coefficient models.

Step 1: Set the initial input for $\hat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}$ and input data $\{(\mathbf{X}_i, t_i, Y_i), i = 1, \cdots, n\}$. Set $m = 1$. Set two tolerance limits $\mathtt{tol}_\beta$ and $\mathtt{tol}_\theta$.

Step 2: Set $\tilde{Y}_i^{(m)} = Y_i - \mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(m-1)}$ and fit the expanded form of the varying coefficient model

$$\tilde{Y}_i^{(m)} = \mathbf{\Pi_i}^T \mathbf{\Theta} + e_i$$

to estimate $\hat{\mathbf{\Theta}}^{(m)}$ of $\mathbf{\Theta}$ as

$$\hat{\mathbf{\Theta}}^{(m)} = \text{argmin}_{\mathbf{\Theta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R(e_i(\mathbf{\Theta}))}{n+1} \Big) e_i(\mathbf{\Theta}) + \sum_{k=1}^{p} \lambda_{1nk} ||\theta_k||_2 \right\},$$

where $e_i(\mathbf{\Theta}) = \tilde{Y}_i^{(m)} - \mathbf{\Pi_i}^T \mathbf{\Theta}$.

Step 3: If $\alpha_k(\cdot)$, $k = 1, \ldots, p$, is not selected, delete the column of $\mathbf{X}$ associated with $\alpha_k(\cdot)$. Organize the remaining columns of $\mathbf{X}$ as $\mathbf{X}^{\texttt{lasso}}$.

Step 4: Use input data $\{(\mathbf{Z}_i, Y_i), i = 1, \cdots, n\}$ and $\hat{\mathbf{\Theta}}^{(m)}$, fit a linear model with the response $Y_i - \mathbf{\Pi_i}^T \hat{\mathbf{\Theta}}^{(m)}$ on $\mathbf{Z}_i$ to obtain an estimate $\hat{\boldsymbol{\beta}}^{(m)}$ of $\boldsymbol{\beta}$ defined by

$$\hat{\boldsymbol{\beta}}^{(m)} = \text{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \varphi\Big(\frac{R(e_i(\boldsymbol{\beta}))}{n+1} \Big) e_i(\boldsymbol{\beta}) + \sum_{k=1}^{q} \lambda_{2nk} |\beta_k| \right\},$$

where $e_i(\boldsymbol{\beta}) = Y_i - \mathbf{\Pi_i}^T \hat{\mathbf{\Theta}}^{(m)} - \mathbf{Z}_i^T \boldsymbol{\beta}$.

Step 5: If β_k is not selected, delete the corresponding column of $\mathbf{Z}$ associated with β_k and let the remaining columns be $\mathbf{Z}^{\texttt{lasso}}$.

Step 6: Set $m \leftarrow m + 1$. If $m < 2$, then go back to **Step 2**. If $m \geqslant 2$ and $(\|\hat{\boldsymbol{\beta}}^{(m)} - \hat{\boldsymbol{\beta}}^{(m-1)}\|) > \texttt{tol}_\beta \|\hat{\boldsymbol{\beta}}^{(m-1)}\|$ and $(\|\hat{\mathbf{\Theta}}^{(m)} - \hat{\mathbf{\Theta}}^{(m-1)}\|) > \texttt{tol}_\theta \|\hat{\mathbf{\Theta}}^{(m-1)}\|$, then go back to **Step 2**. Otherwise, STOP.

Chapter 4

Simulation

To assess the finite sample performance of the proposed methodology, we performed Monte Carlo simulation experiment. This chapter presents the results of this analysis.

4.1 Simulation Setup

We consider the semi-varying coefficient model:

$$\begin{aligned} Y_i &= \alpha_0(t_i) + \sum_{k=1}^{6} \alpha_k(t_i) x_{ki} + \sum_{k=1}^{5} \beta_k z_{ki} + \varepsilon_i \\ &= (1 + 3t_i^2) + 3\exp(-t_i^2)x_{1i} + 1.5\sin(\pi t_i)x_{2i} + 0.8t_i x_{3i} + 1.5z_{1i} + z_{4i} + 3z_{5i} + \varepsilon_i \end{aligned}$$

In this setup, we have taken the true varying coefficients to be

$$\alpha_0(t_i) = 1 + 3t_i^2,\ \alpha_1(t_i) = 3\exp(-t_i^2),\ \alpha_2(t_i) = 1.5\sin(\pi t_i),\ \alpha_3(t_i) = 0.8t_i$$
$$\alpha_4(\cdot) = \alpha_5(\cdot) = \alpha_6(\cdot) = 0$$

The true fixed parameters are

$$\beta_1 = 1.5,\ \beta_2 = 0,\ \beta_3 = 0,\ \beta_4 = 1,\ \beta_5 = 3$$

Thus the true model contains three zero functional coefficients and two zero fixed coefficients. The times that serve as the index for the varying coefficients are generated as $t_i \sim U(-1, 1)$. The covariates $\mathbf{Z}$ are generated from the multivariate uniform distribution on $(-1, 1)$ and $\mathbf{X}$ are generated from the multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\text{cov}(x_k, x_l) = 0.5^{|k-l|}$. We considered three different error distributions, the standard normal distribution, $N(0, 1)$, the $t-$distribution with three degrees of

freedom and the contaminated normal distribution with a 5% rate of contamination, i.e, $\mathcal{CN}(0.95) = 0.95N(0,1) + 0.05N(0,10^2)$. The t distributed errors allow us to evaluate the effect of heavy tails on the proposed on the proposed estimator while the contaminated normal distribution allows us to evaluate the effect of outliers on the estimator. Two different sample sizes ($n = 200$, $n = 400$) are considered.

Both penalized rank (R) and penalized least squares (LS) estimators of the model coefficients were calculated. The assessment of the performance of the estimation and variable selection procedures for functional coefficients was done by using the following measures:

- True positive rate (TPR): The proportion of non-zero coefficients estimated as non-zero
- False positive rate (FPR): The proportion of zero coefficients estimated as non-zero
- Mean Squared Error (MSE): The mean squared error for the non-zero coefficients

These were computed on $B = 400$ replications. For each non-zero functional coefficient $\alpha_k(\cdot)$, the MSE of $\alpha_k(\cdot)$ was computed as

$$\text{MSE}(\alpha_k) = \frac{1}{nB}\sum_{j=1}^{B}\sum_{i=1}^{n}(\hat{\alpha}_k(t_i^j) - \alpha_k(t_i^j))^2 \ ,$$

where t_i^j is the ith time index generated in the jth iteration. The MSE of the non-zero parametric coefficient β_k was calculated as

$$\text{MSE}(\beta_k) = \frac{1}{B}\sum_{j=1}^{B}(\hat{\beta}_k^j - \beta_k)^2 \ ,$$

where $\hat{\beta}_k^j$ is the estimate of β_k obtained in the jth iteration.

Graphs of the 400 estimated coefficient functions along with the true function are plotted to show the performance of the selected functions, including the zero functions. Moreover, graphs containing the 95% pointwise confidence interval for the functional coefficients are provided.

Table 4.1 contains the TPR and FPR results. Table 4.2 and Table 4.3 contain the results of the estimated MSEs for the functional coefficients and fixed coefficients, respectively.

4.2 Simulation Results

TPR and FPR

The TPR and FPR results contained in Table 4.1 show that LS has a slight edge over R when the underlying distribution is Gaussian. However, R shows superior performance when the error distribution is heavier tailed (t_3) or there are outliers in the data (error distribution $\mathcal{CN}(.95)$). This is in line with our expectations and similar to what is observed for linear models. For the smaller sample size cases ($n = 200$), LS shows low TPR and high FPR, a very undesirable combination. On the other hand, R appears to maintain its high TPR and low FPR rates even at the $n = 200$ sample size. Outliers seem to have an unduly large effect on the LS estimator with FPR $\approx 5\%$ for $n = 200$ which is double the FPR of the R estimator and TPR $\approx 65\%$ which is around 6% lower than the TPR of R. For $n = 400$, the FPRs of LS and R are comparable but the TPR of R is around 13% higher than the TPR of LS (82% versus 69%). The observation is analogous for heavy tails with the FPR of LS over three times that of R for $n = 200$. An interesting observation is that when the sample size is large ($n = 400$), LS appears to catch up with R in its performance for the heavier tailed distribution (t_3).

TPR and FPR				
Error	Size	Method	TPR	FPR
$N(0,1)$	200	LS	0.9275	0.0000
		R	0.8950	0.0050
	400	LS	0.9875	0.0000
		R	0.9888	0.0000
t_3	200	LS	0.7263	0.0208
		R	0.8200	0.0058
	400	LS	0.8013	0.0000
		R	0.8100	0.0000
$\mathcal{CN}(0.95)$	200	LS	0.6538	0.0467
		R	0.7100	0.0225
	400	LS	0.6888	0.0025
		R	0.8236	0.0025
		Oracle	1.0000	0.0000

Table 4.1: True and False Positive Rates of the Parametric Part

MSEs of $\alpha(\cdot)$

Table 4.2 gives the results of the MSEs of the estimated coefficient functions of the non-zero functional coefficients. In the estimation of the quadratic intercept function $\alpha_0(t) = 1 + 3t^2$, LS and R give comparable performance in the Gaussian error case. However for heavy tailed (t_3) and contaminated ($\mathcal{CN}(.95)$) data, R gives a vastly superior performance where in some cases MSE(α_0) for the R estimator is a third of that of the LS estimator. In the case of $n = 200$, for the estimation of $\alpha_1(t) = 3\exp(-t^2)$ and $\alpha_2(t) = 1.5\sin(\pi t)$, an exponential and a sine function, respectively, the LS gives a vastly superior performance for Gaussian data and R gives a vastly superior performance for heavy tailed and contaminated

data. The differences in MSE(α_1) are particularly stark. Although the pattern is similar for $n = 400$, larger sample sizes seem to ameliorate the situation. Although the pattern in superiority of performance remains the same for the linear $\alpha_3(t) = 0.8t$, rather surprisingly, the differences in the performance of LS and R are somewhat mediated. This suggests a potential link between performance of the estimators and the underlying nonlinearity of the coefficient function.

MSEs of Non-Zero functions						
Error	Size	Method	α_0	α_1	α_2	α_3
$N(0,1)$	200	LS	0.0493	0.0386	0.1051	0.0467
		R	0.0472	0.0976	0.1910	0.0909
	400	LS	0.0259	0.0188	0.0640	0.0152
		R	0.0226	0.0259	0.0965	0.0175
t_3	200	LS	0.1753	1.7672	0.6520	0.1884
		R	0.0958	0.1397	0.2404	0.1419
	400	LS	0.0482	0.0421	0.1031	0.0616
		R	0.0293	0.0328	0.1087	0.0520
$\mathcal{CN}(0.95)$	200	LS	0.2998	2.4375	0.7543	0.1950
		R	0.1636	0.5268	0.5431	0.1917
	400	LS	0.0839	0.0807	0.1777	0.1369
		R	0.0294	0.0341	0.1219	0.0400

Table 4.2: MSEs of Non-Zero Functions

MSEs of β

The results from the MSE(β) calculations mirror those of MSE(α) with LS giving superior performance than R for Gaussian data and the reverse when the data are heavy tailed or contaminated. However, the differences in MSE(β) were not as stark as those between

MSE(α) values. Relative efficiencies of the R estimator calculated as

$$\text{RE} = \frac{\text{MSE}_{LS}(\beta)}{\text{MSE}_{R}(\beta)}$$

ranged roughly between 70% and 95% for Gaussian data. The RE values were all over 100% in some cases up to 140% ($t_3, n = 200$) indicating a large gain in efficiency from using the R estimator instead of the LS estimator.

MSEs of Non-Zero Parameters					
Error	Size	Method	β_1	β_4	β_5
$N(0,1)$	200	LS	0.1213	2.4682	1.5886
		R	0.1612	2.2575	1.9741
	400	LS	0.0209	0.0175	0.0429
		R	0.0268	0.0252	0.0473
t_3	200	LS	0.2200	2.6083	1.9776
		R	0.1564	2.5304	1.6706
	400	LS	0.0446	0.0406	0.0730
		R	0.0432	0.0388	0.0653
$\mathcal{CN}(0.95)$	200	LS	0.3124	2.7037	1.8971
		R	0.2190	2.6638	1.7588
	400	LS	0.0766	0.0710	0.1030
		R	0.0757	0.0688	0.0842

Table 4.3: Mean Squared Error of the Non-Zero Parameters

In the following pages, we will provide Figures 4.1 – 4.6 which contain plots of the estimated functional coefficients from both LS and R estimation methods. A discussion of the results is given following the Figures.

Estimated $\alpha(\cdot)$**,** $N(0,1)$**,** $n = 200$

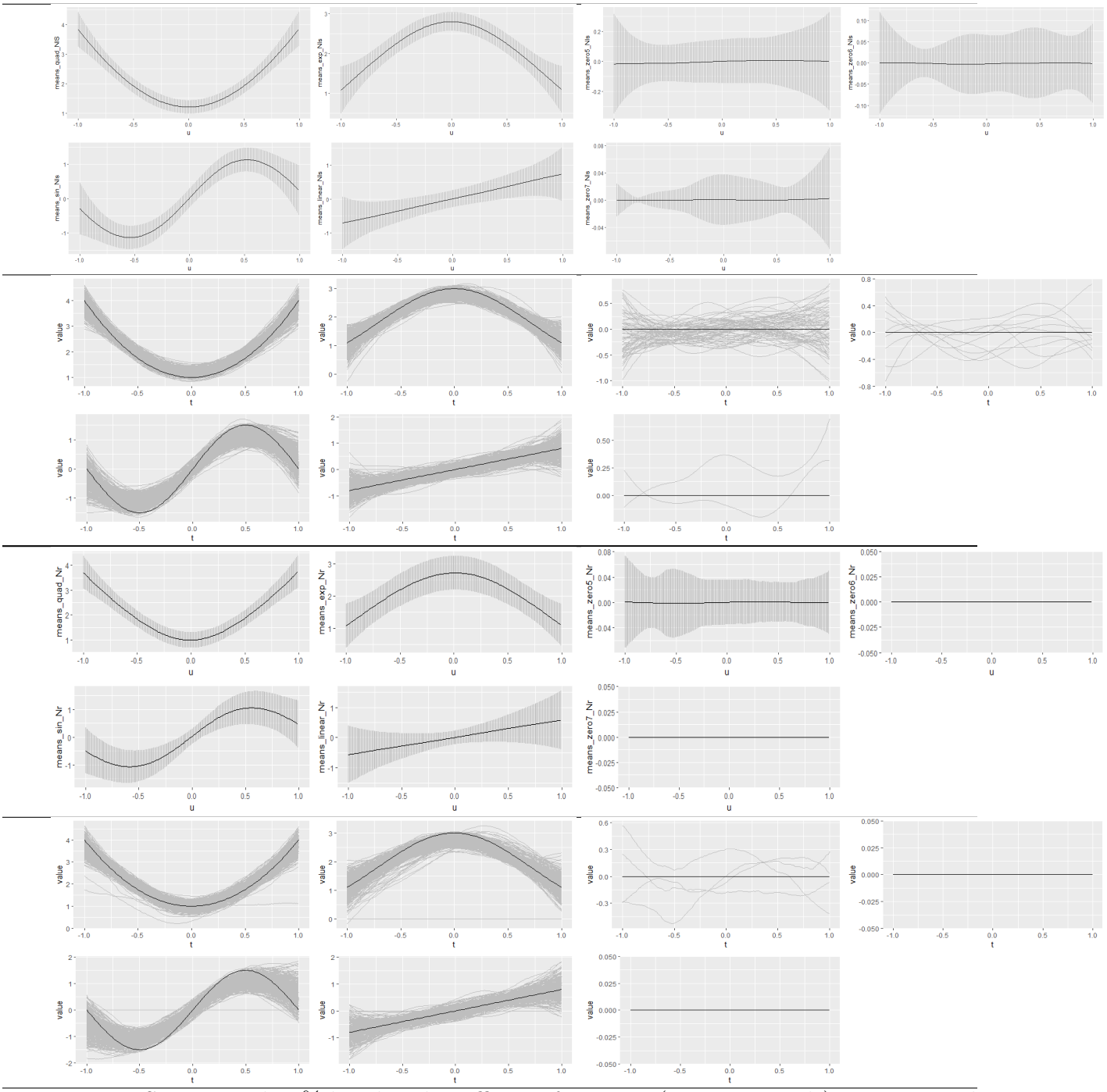

Figure 4.1: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $N(0,1), n = 200$.

Estimated $\alpha(\cdot)$**,** $N(0,1)$**,** $n = 400$

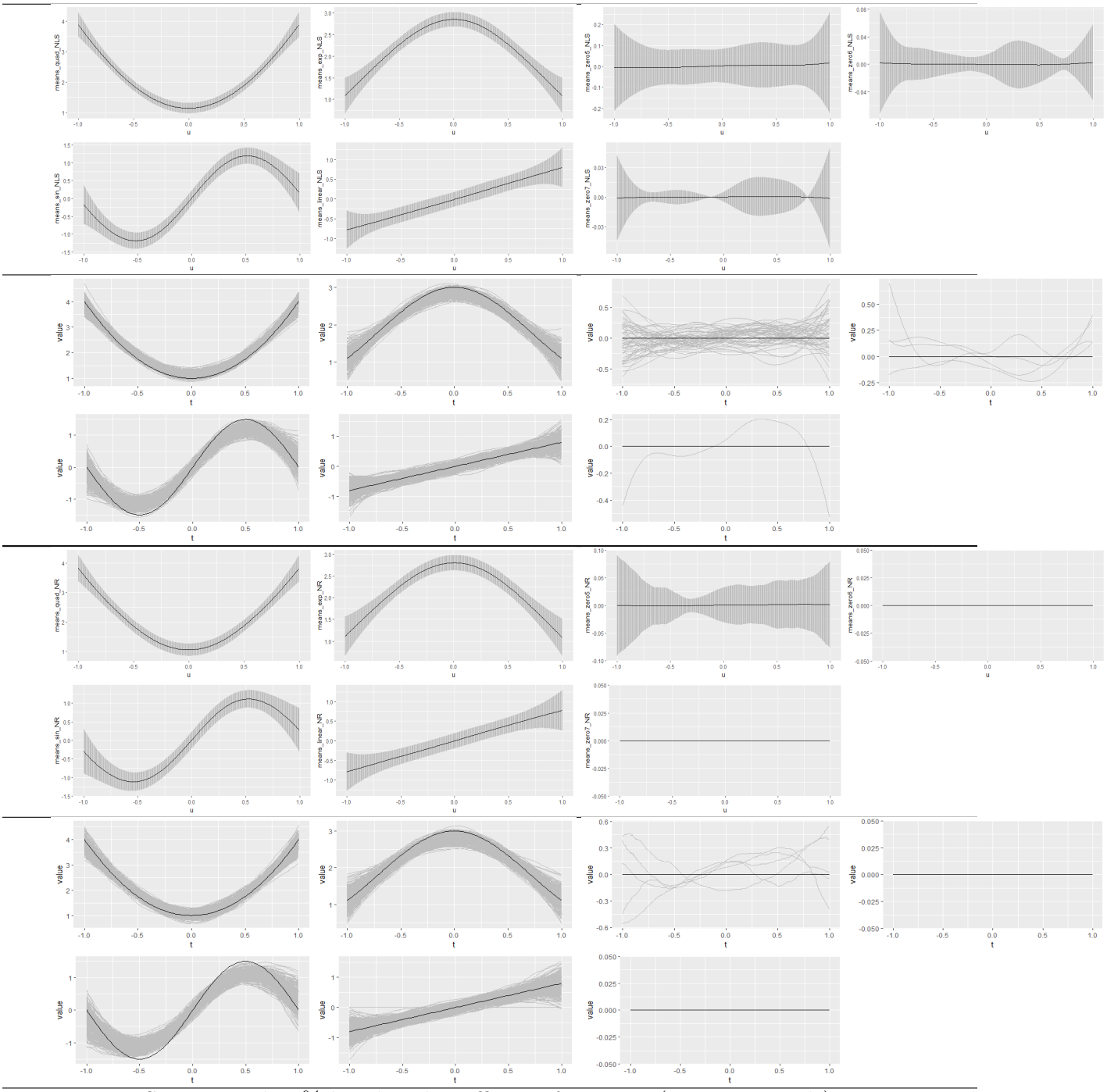

Figure 4.2: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $N(0,1), n = 400$.

Estimated $\alpha(\cdot)$, t_3, $n = 200$

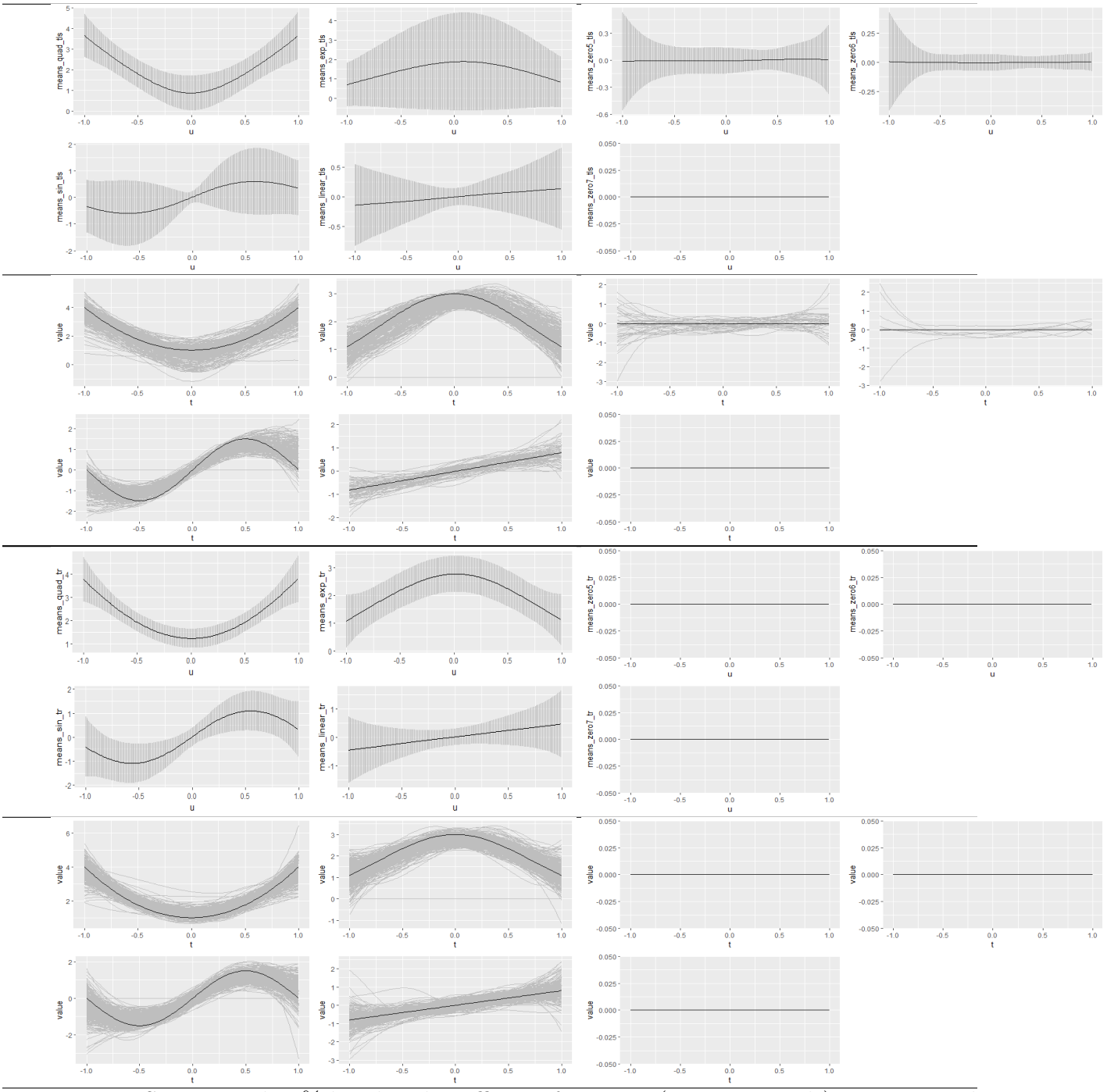

Figure 4.3: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $t_3, n = 200$.

Estimated $\alpha(\cdot)$, t_3, $n = 400$

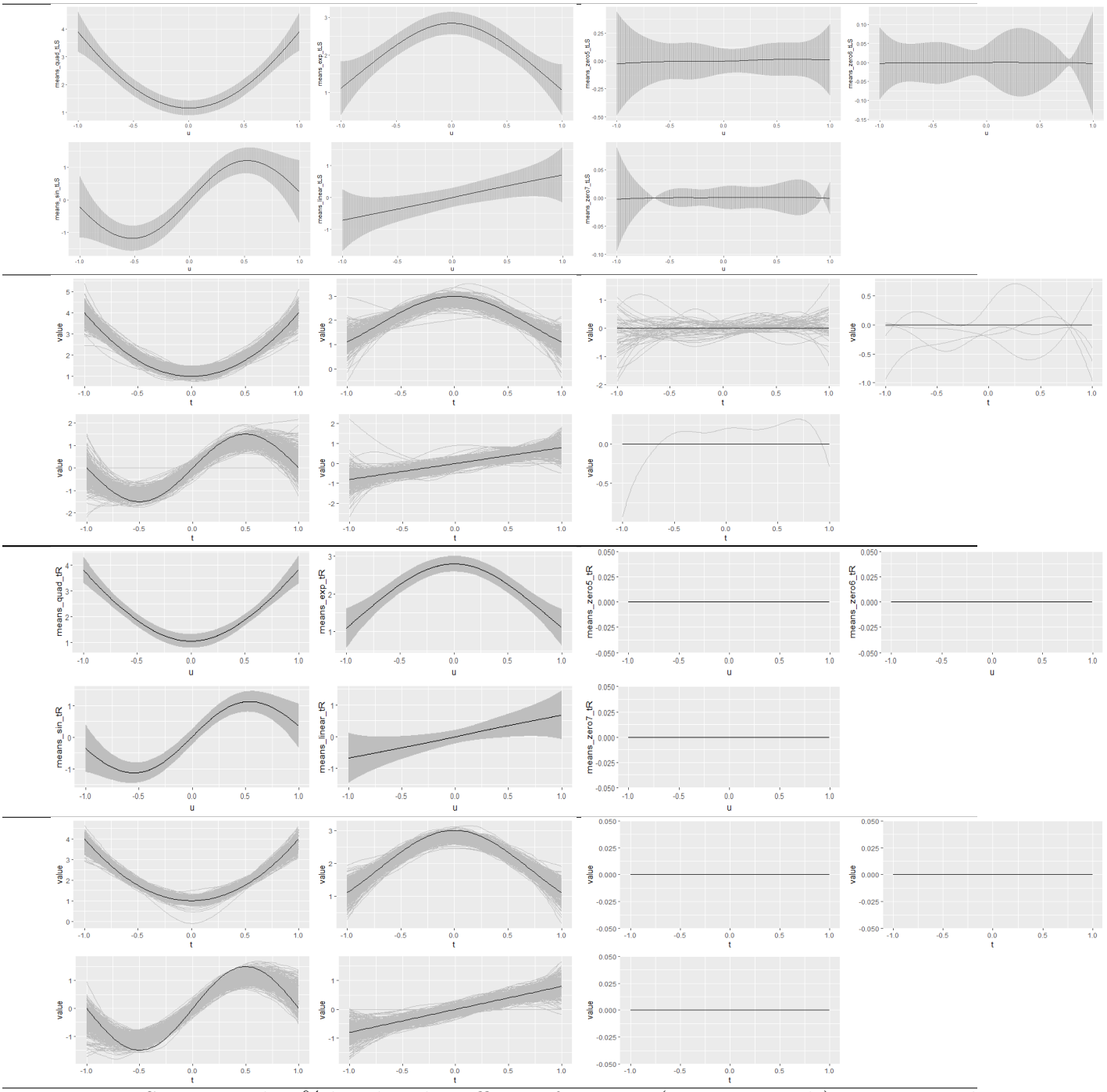

Figure 4.4: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $t_3, n = 400$.

Estimated $\alpha(\cdot)$, $CN(.95)$, $n = 200$

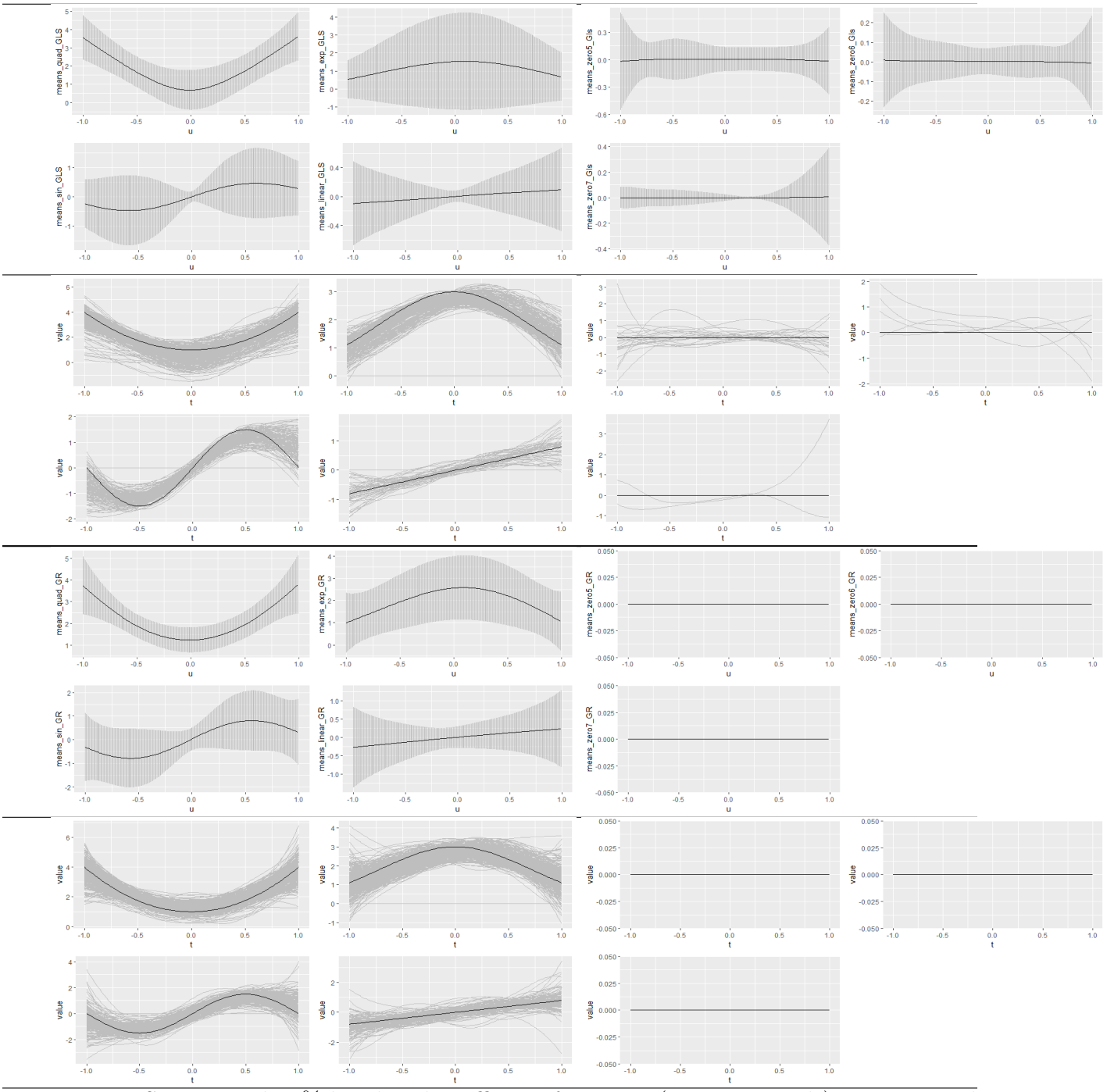

Figure 4.5: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $CN(.95), n = 200$.

Estimated $\alpha(\cdot)$, $CN(.95)$, $n = 400$

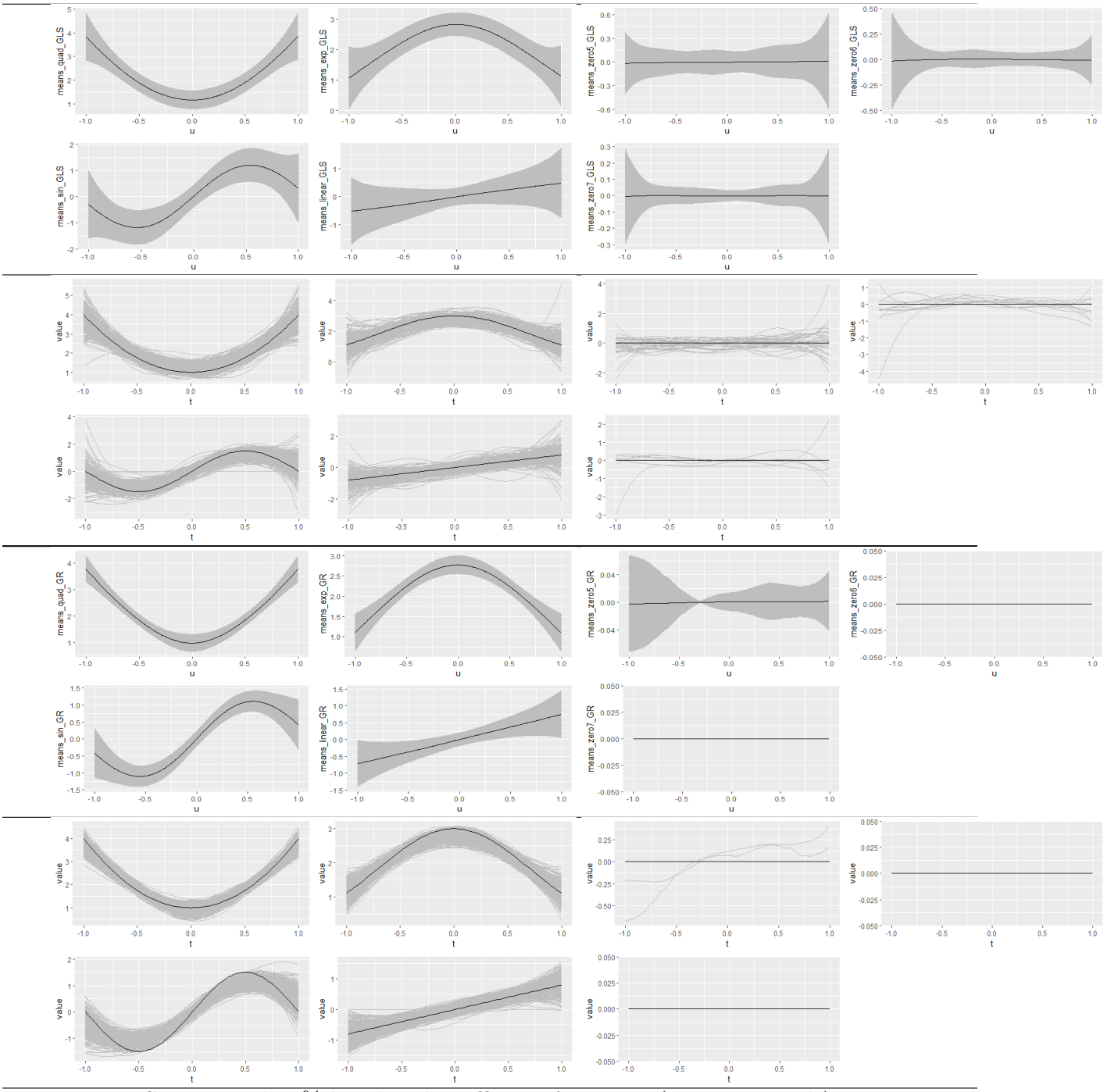

Figure 4.6: LS estimated 95% band and coefficient functions (top two panels); R estimated 95% band and coefficient functions (bottom two panels); $CN(.95), n = 400$.

Discussion of the results of the estimation of $\alpha(\cdot)$

The efficiency of the estimation of the non-zero functional coefficients will be in line with the MSE(α) results given in Table 4.2. The graphs, however, will allow us to make comparisons of the performances of LS and R estimation of the zero functional coefficient estimation.

For the normal error case, while the efficiency of the LS estimator in the estimation of the non-zero functional coefficients is superior to that of R, it appears that it does worse in estimating the zero functional coefficients. R on the other hand does a better job in estimating the zero functional coefficients without substantial loss in the efficiency of the estimation of the non-zero functional coefficients.

For the heavier tailed data (t_3), the rank estimator gives better efficiency in the estimation of the non-zero functional coefficients as expected. Moreover, the rank estimator gives better performance in the estimation of the zero functional coefficients while the LS misfires a few times.

The observation for the contaminated data case is similar to that of the heavy tailed data. The rank estimator gives excellent performance in efficiently estimating the non-zero functions as well as in correctly estimating the zero functions.

Generally, the rank estimator does an excellent job with the correct estimation of the zero functional coefficients even in the case of Gaussian data, which is remarkable. It is also efficient in the estimation of the non-zero functional coefficients as shown in Figures 4.1 – 4.6 and Table 4.2.

The "oracleness" of rank estimation for functional coefficients in the sense of efficiency and selection consistency is notable. A similar observation was made by Sun et al. (2022) which gave a rank-based estimation and selection procedure of the functional regression coefficients for the single-index varying coefficient regression model.

Although we have not included any confidence interval techniques for the regression parameters even with small sample sizes. This can be constructed via an empirical likelihood

(EL) for the regression parameters based on the rank estimating equation. The limiting distribution of log-empirical likelihood ratio is the χ^2 distribution. This has been studied by Bindele and Zhao (2015).

Chapter 5

Factors Affecting Excess Mortality from the COVID-19 Pandemic

5.1 Background

Besides being important input in COVID prediction models, correctly estimated excess mortality from COVID-19 is a measure of the magnitude of impact of the pandemic on countries across the globe (Wang et al., 2022). Because excess mortality provides a metric of the overall mortality impact of COVID-19, it can be "useful to monitor trends within and between countries and inform international, national and local public health policies" (Beaney et al., 2020). Since 2020, there have been a number of studies that examined excess mortality on a regional or country basis. These include Chen et al. (2021) that studied excess mortality in California, Stang et al. (2020) in Germany, Ballin et al. (2021) in Sweden, Stokes et al. (2021) in the US at the county level, Haklai et al. (2021) in Israel, Aron and Muellbauer (2020) in England, among others.

When it comes to factors affecting COVID-19 excess mortality, Kapitsinis (2020) studied the effect of pre-pandemic healthcare conditions on cross-country COVID-19 excess mortality in 2020. Rossen et al. (2021) studied factors to lead to disparities in COVID-19 excess mortality in the US using time-series SARIMA models.

Our interest in this dissertation is to examine fixed and time-varying factors that affect cross-country COVID-19 excess mortality on the global scale using semi-varying coefficient models, to provide a practical estimation algorithm for estimation and variable selection, and to make a recommendation on the estimation approach that provides optimal selection of the important factors.

5.2 Data and Variables

Before applying to a real data, we will provide a brief discussion of the data.

To examine the impact of varying coefficients non-varying coefficients on COVID-19, we employ an unbalanced panel data of daily COVID-19 information on confirmed cases, confirmed death, stringency index, population, human development index, median age, among other variables from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (Our World Data). The dataset contains monthly/daily observations, for the period February 24, 2020 to November 29, 2021. The variable definitions are explained below.

Mortality Rate:

We define mortality rate as the ratio of confirmed number of deaths to the number of confirmed cases of COVID-19 virus. This definition has been employed by the CDC (2019) and the study by Akesson et al. (2021). On the one hand, confirmed number of deaths indicates the cumulative number of confirmed COVID-19 deaths per day. In reporting confirmed deaths, countries follow the World Health Organization's International Classification of Disease guidelines (WHO, 2016) as well as individual country guideline on recording COVID-19 deaths. Total confirmed deaths include 55,808 observations with a mean of 11,163 and a standard deviation of 9,003. On the other hand, confirmed cases of COVID-19 virus entails the cumulative number of infected people daily.

Stringency index: The stringency index is a composite score drawn from nine metrics on school closures, workplace closures, public event cancellation, restrictions on public gathering, public transport closures, stay-at-home requirements, public information campaigns, internal movement restrictions and controls on international travels. These metrics are generated using the Oxford COVID-19 Government Response Tracker. The index on any given

day ranges between 0 and 100. A higher score indicates a stricter response with 100 representing the strictest response. The data entails daily stringency index for 58,649 with a mean of 59 and standard deviation of 32.

Other explanatory variables: More explanatory variables employed in this model include population density, median age, human development index and people aged 65 years and older. Population density is defined as the number of people per square mile/kilometer. Median age gives us an important single point for the age distribution of population of different countries of the world. In this context, we define it as the age 'midpoint' of a population such that there the same number of people that are older than the median age as there are that are younger than it. As the name indicates, people aged 65 year and older is a variable that captures the population shares of adults who are 65 years and above. We include this variable because given the fact that COVID-19 virus affects older as well as vulnerable population, we wanted to see if this will be captured in the results. Human development index was also employed as an explanatory variable. The index is defined as a measure of key dimension of human development such as life expectancy, access to education, and gross national income per capita adjusted for the price level of each country. To bring this pandemic to an end, a large share of the world needs to be immune to the virus. The safest way to achieve this is with a vaccine. Vaccines are a technology that humanity has often relied on in the past to bring down the death toll of infectious diseases. We use people vaccinated to see how this affects the mortality rate.

The attribution of deaths to specific causes can be challenging under any circumstances. Health problems are often connected, and multiplicative, meaning an underlying condition can often lead to complications which ultimately result in death. Both guidelines state that if the practitioner suspects that COVID-19 played a role in an individual's death it should be specified on the death certificate. In some cases, COVID-19 may be the underlying cause of death, having led to complications such as pneumonia or ARDS. Even when it's the

underlying and not the direct cause, COVID-19 should be listed.

For the variable Cases, a person must have a positive result from laboratory tests to be classified under confirmed cases,regardless of whether they have shown symptoms of COVID-19 or not.

Hospital patients are people who are in hospital due to COVID-19 at a given time and ICU patients are people are in ICU due to COVID-19 at a given time.

5.3 Data Cleaning

At the time this data was collected, it contained 136,687 observations and 68 variables. The number of observations are from 237 locations, which includes countries. This data had a lot of NA values for most locations, leaving only a hand full of locations with data to work with. We started by selecting the explanatory variables that change with time, like People Vaccinated, as well as variables that do not change with time like Median Age. Because our work does not consider missing values at this time, we decided to filter out countries with data in the chosen variables. Although many variables have been ignored for this data set, most of them are products of other variables and so are highly correlated and we decided to move forward with the following variables:

Time Dependent Variables: Time dependent variables are "new cases per million", "new deaths per million", "ICU patients per million", "hospital patients per million", "stringency index", and "people vaccinated".

Fixed Variables: Fixed variables are "population density", "median age", "aged 65 or older", "aged 70 or older", "GDP per capita", "life expectancy" and "human development index" and the "mortality rate" is the dependent variable.

The data were collected over a period of 90 weeks and we want to see which variables have an effect on mortality rate.

5.4 A Mixed Backfitting Algorithm

Before we get into the algorithm, we need to discuss a little more about how we handle this high-frequency data. Since this is a cross-sectional data, we regard each of the 23 selected countries as an observation, so in all, we have 23 observations, each with its own sample size. With this, we want to see what time varying variables affect the mortality rate for each country and what variables have been selected per country. The varying variables will make up the matrix $\mathbf{X}$ and the non-varying variables will make up the matrix $\mathbf{Z}$. We let $c = 23$ represent the number of countries and n_s the number of time points for country s, $s = 1, \ldots, c$. We let $n = n_1 + \cdots + n_c$ denote the total number of observations across all countries.

The mixed backfitting algorithm is given below:

Step 1: Set the initial input for $\hat{\boldsymbol{\beta}}^{(0)} = \mathbf{0}$ and input data $\{(\mathbf{X}_{si}, \mathbf{Z}_{si}, t_i, Y_{si}), i = 1, \cdots, n_s;\ s = 1, \ldots, c\}$. Set $m = 1$. Set two tolerance limits $\texttt{tol}_\beta$ and $\texttt{tol}_\theta$.

Step 2: Set $\tilde{Y}_{si}^{(m)} = Y_{si} - \mathbf{Z}_{si}^T \hat{\boldsymbol{\beta}}^{(m-1)}$ and fit the expanded form of the varying coefficient model for each $s = 1, \ldots, c$

$$\tilde{Y}_{si}^{(m)} = \mathbf{\Pi}_{si}^T \mathbf{\Theta}_s + e_{si}$$

to estimate $\hat{\mathbf{\Theta}}_s^{(m)}$ of $\mathbf{\Theta}_s$ as

$$\hat{\mathbf{\Theta}}_s^{(m)} = \text{argmin}_{\mathbf{\Theta}_s} \left\{ \frac{1}{n_s} \sum_{i=1}^{n_s} \varphi\Big(\frac{R(e_{si}(\mathbf{\Theta}_s))}{n+1}\Big) e_{si}(\mathbf{\Theta}_s) + \sum_{k=1}^{p} \lambda_{1n_s k} ||\theta_{ks}||_2 \right\},$$

where $e_{is}(\mathbf{\Theta}_s) = \tilde{Y}_{si}^{(m)} - \mathbf{\Pi}_{si}^T \mathbf{\Theta}_s$.

Step 3: Stack the selected functions in the same format as the matrix $\mathbf{X}$ and use it to compute the residual. If $\alpha_{sk}(\cdot)$, $k = 1, \ldots, p; s = 1, \ldots, c$, is not selected, delete the column of $\mathbf{X}$ associated with $\alpha_{sk}(\cdot)$. Organize the remaining columns of $\mathbf{X}$ as $\mathbf{X}^{\texttt{lasso}}$.

Step 4: Use input data $\{(\mathbf{Z}_{si}, Y_{si}), i = 1, \cdots, n_s; s = 1, \ldots, c\}$ and $\hat{\mathbf{\Theta}}_s^{(m)}$, fit a linear model with the response $Y_{si} - \mathbf{\Pi}_{si}^T \hat{\mathbf{\Theta}}_s^{(m)}$ on $\mathbf{Z}_{si}$ to obtain an estimate $\hat{\boldsymbol{\beta}}^{(m)}$ of $\boldsymbol{\beta}$ defined by

$$\hat{\boldsymbol{\beta}}^{(m)} = \text{argmin}_{\boldsymbol{\beta}} \left\{ \frac{1}{n} \sum_{s=1}^{c} \sum_{i=1}^{n_s} \varphi\Big(\frac{R(e_{si}(\boldsymbol{\beta}))}{n+1}\Big) e_{si}(\boldsymbol{\beta}) + \sum_{k=1}^{q} \lambda_{2nk} |\beta_k| \right\} ,$$

where $e_{si}(\boldsymbol{\beta}) = Y_{si} - \mathbf{\Pi}_{si}^T \hat{\mathbf{\Theta}}_s^{(m)} - \mathbf{Z}_{si}^T \boldsymbol{\beta}$.

Step 5: If β_k is not selected, delete the corresponding column of $\mathbf{Z}$ associated with β_k and let the remaining columns be $\mathbf{Z}^{\texttt{lasso}}$.

Step 6: Set $m \leftarrow m+1$. If $m < 2$, then go back to **Step 2**. If $m \geqslant 2$ and $\|\hat{\boldsymbol{\beta}}^{(m)} - \hat{\boldsymbol{\beta}}^{(m-1)}\| > \texttt{tol}_\beta \|\hat{\boldsymbol{\beta}}^{(m-1)}\|$ and $\sum_{s=1}^{c} \|\hat{\mathbf{\Theta}}_s^{(m)} - \hat{\mathbf{\Theta}}_s^{(m-1)}\| > \texttt{tol}_\theta \sum_{s=1}^{c} \|\hat{\mathbf{\Theta}}_s^{(m-1)}\|$, then go back to **Step 2**. Otherwise, STOP.

5.5 Results

We first present the results of the effects of the time-varying variables on excess mortality. A brief interpretation of the estimated coefficient functions is provided below the estimates of each variable. Since time-varying coefficients were estimated for each individual country in the cleaned data, the results are also reported in country specific plots.

5.5.1 Least Squares Estimation and Selection

The estimated functional coefficients using the least squares (LS) approach are given in the following pages.

Figure 5.1 gives the results of the coefficients for the variable "New Deaths per Million". According to the LS estimates, it appears that there is no discernible relationship between

new deaths per million and excess mortality in Cyprus, Estonia, Finland, Luxembourg, Malta, and Spain. With the exception of the UK, all other countries finish the time period showing a decreasing relationship between new deaths per million and excess mortality. The downward trend following an inflection on the LS estimated coefficient functions gives the impression that the relationship between COVID mortality and new deaths became increasingly negative over time.

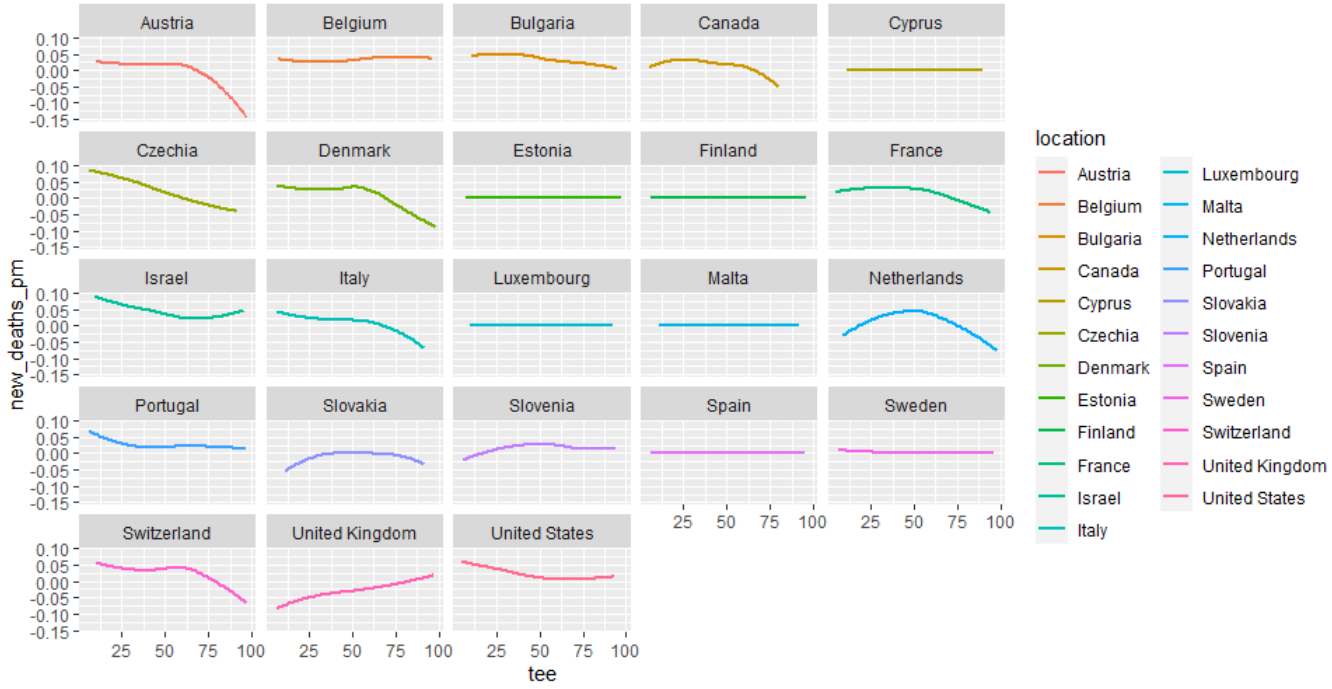

Figure 5.1: LS Estimated Coefficient Functions for New Deaths per Million.

Figure 5.2 contains the LS estimated coefficient functions relating new cases to excess mortality over time. Except for the US and Italy, this coefficient is not selected for most countries. The increasing and then flattening out trend in Italy and the US captures a relationship that initially became increasingly positive over time. However, this relationship remained more or less flat in the later half or the time period under consideration.

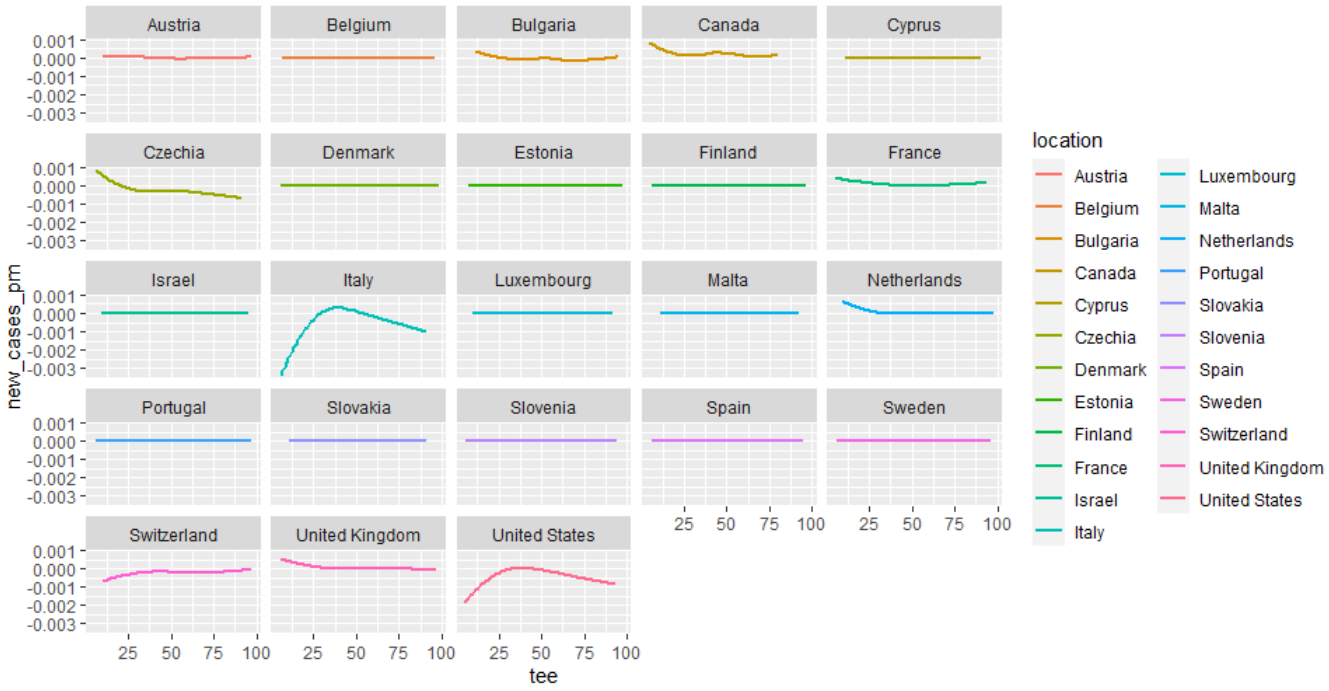

Figure 5.2: New Cases per Million.

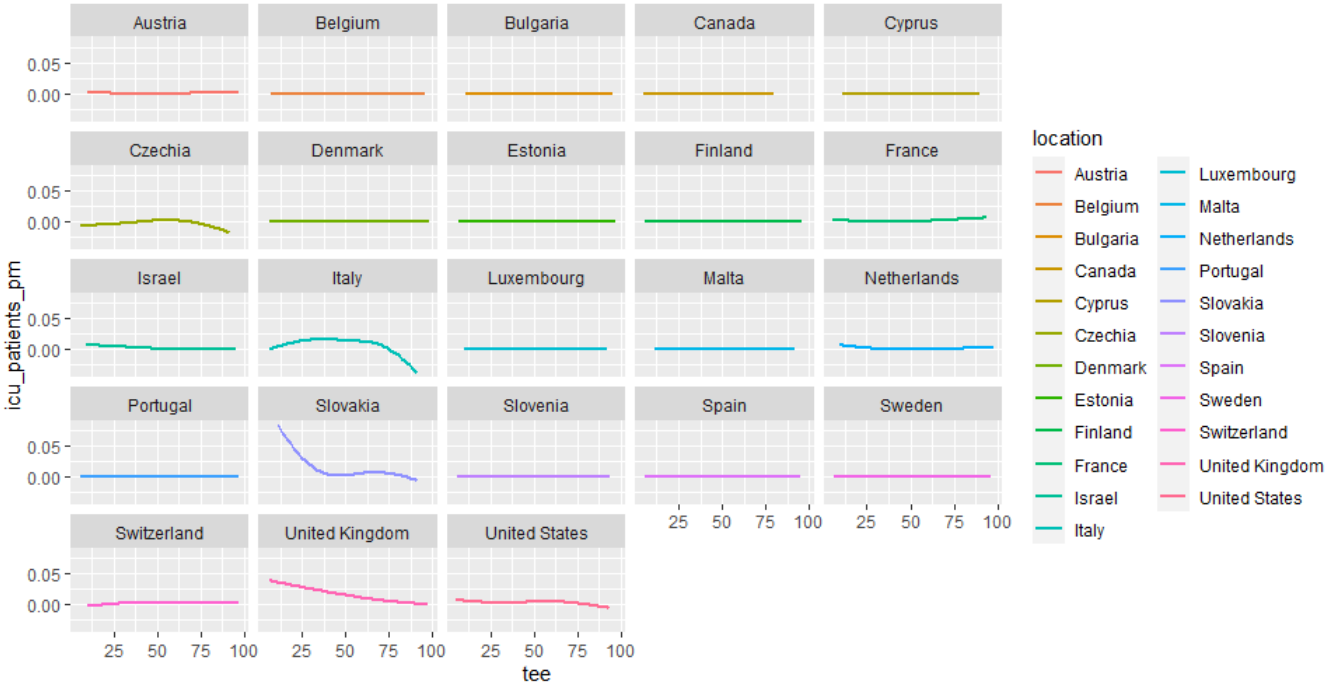

Figure 5.3: ICU Patients per Million.

In Figure 5.3, we see the relationship between the number of ICU patients and excess mortality, where no pattern has been shown for most of the countries except for Italy, Slovakia, UK and the US. These all show a relationship between excess mortality and the number of ICU patients that became increasingly negative over time.

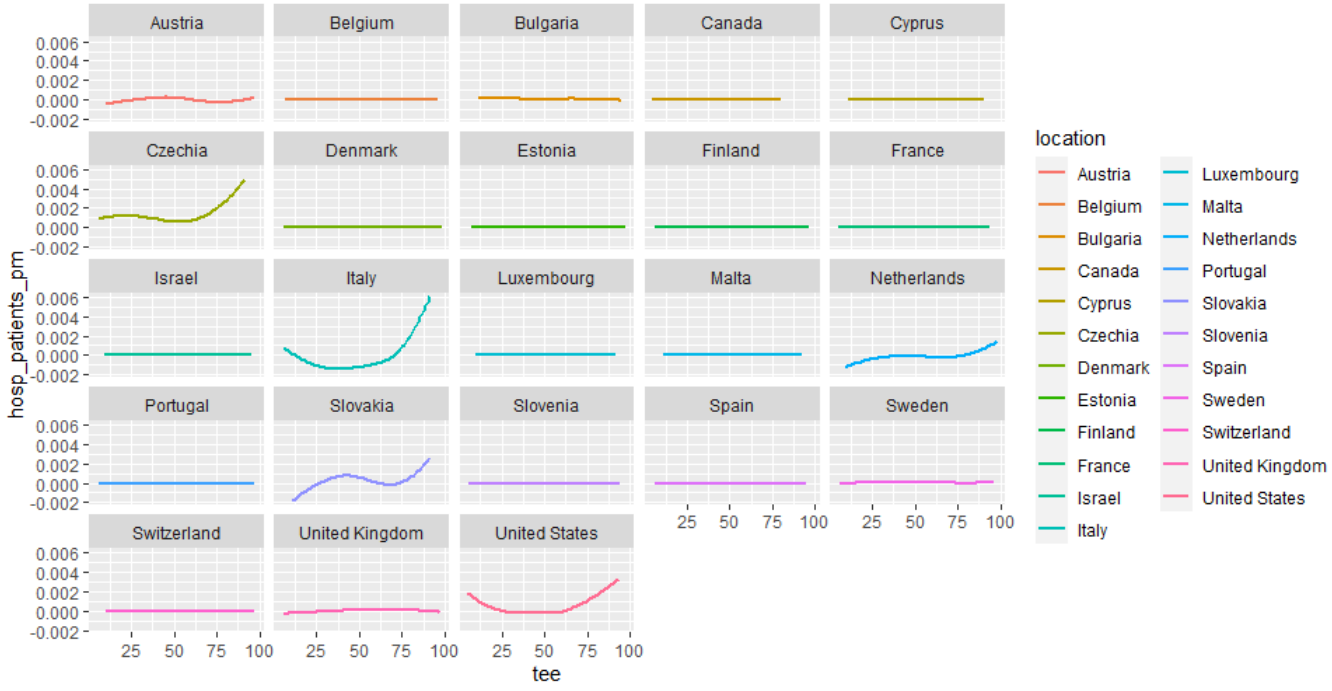

Figure 5.4: Hospital Patients per Million.

Figure 5.4, shows how the relationship between excess mortality and the number of hospital patients became increasingly positive for Czech, Italy, Netherlands, Slovakia and the United States. Some of these relationships are quadratic within the range of time while others are either constant or show relationships of a higher degree polynomials but all ended the time period with an increasing trend.

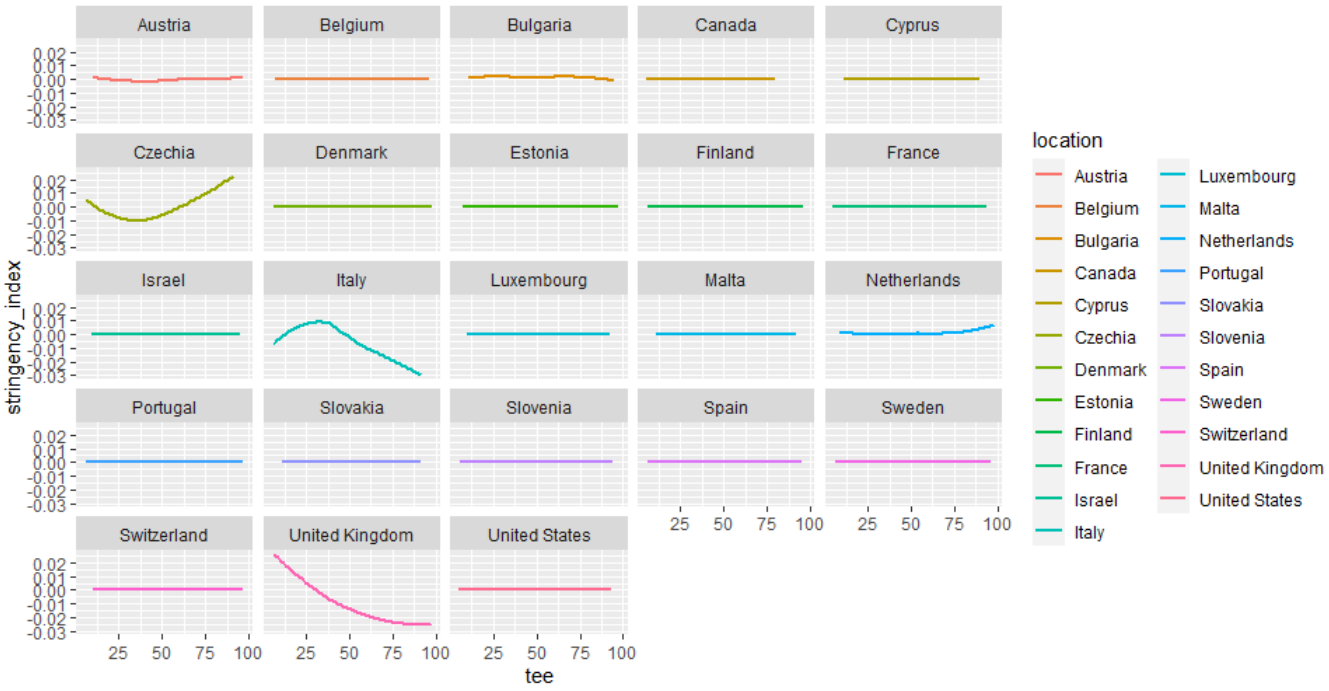

Figure 5.5: Stringency Index.

Stringency Index shows no effect on excess mortality for most countries in Figure **??**, except for the Czech Republic, Italy and the US. Although we believe Stringency Index played a huge role in curbing the impact of this disease, there may have been other factors responsible for this flattening or reverse effect as we see for the Czech republic even though the government declared a state of emergency and curfew as early as March. For Italy and the US, excess mortality appears to decline over time as stricter lockdown measures were carried out.

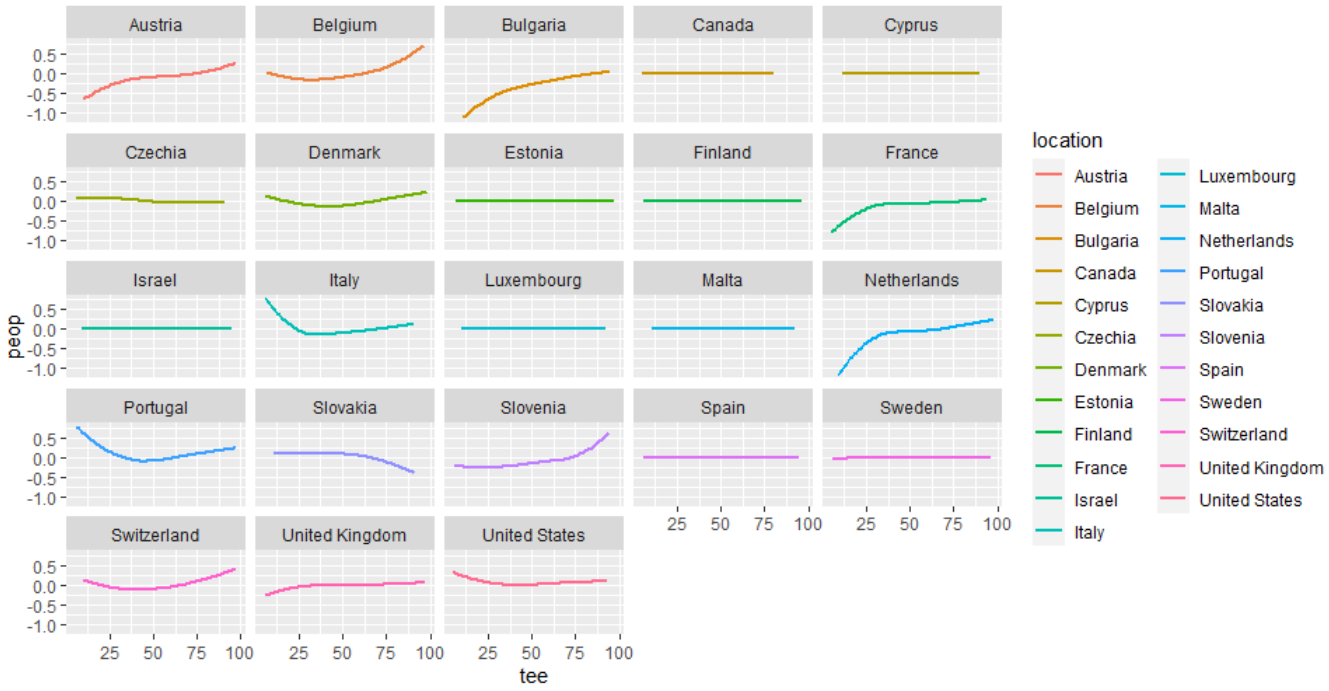

Figure 5.6: People Vaccinated.

Figure 5.6 shows how the number of people vaccinated affects excess mortality. This seems to have affected different countries in different ways and this may be due to time lags and vaccination periods since there were no vaccinations until later in the year. Only in Slovakia does excess mortality and the number of people vaccinated appear to have an increasingly decreasing relationship over time.

The LS estimated fixed coefficients are as follows: $\beta_1 = -3.7010 \times 10^{-4}, \beta_2 = 4.700 \times 10^{-3}, \beta_3 = 0.1225, \beta_4 = -0.2246, \beta_5 = -1.0223 \times 10^{-6}, \beta_6 = 5.9415 \times 10^{-2}, \beta_7 = -11.7921$. Here β_1 is the coefficient of "population density", β_2 is the coefficient of "median age", β_3 is the coefficient of "aged 65 or older", β_4 is the coefficient of "aged 70 or older", β_5 is the coefficient of "GDP per capita", β_6 for "life expectancy" and β_7 the coefficient of "human development index". These results are difficult to interpret since several of the variables are strongly correlated. For instance, the number of people aged 65 or older is likely correlated with the number of people aged 70 or older, thus the coefficients β_3 and β_4 may be difficult

to interpret. This is also likely the case for β_5, β_6, and β_7 since countries with higher life expectancy and higher GDP per capita will have higher development index.

5.5.2 Rank Estimation and Selection

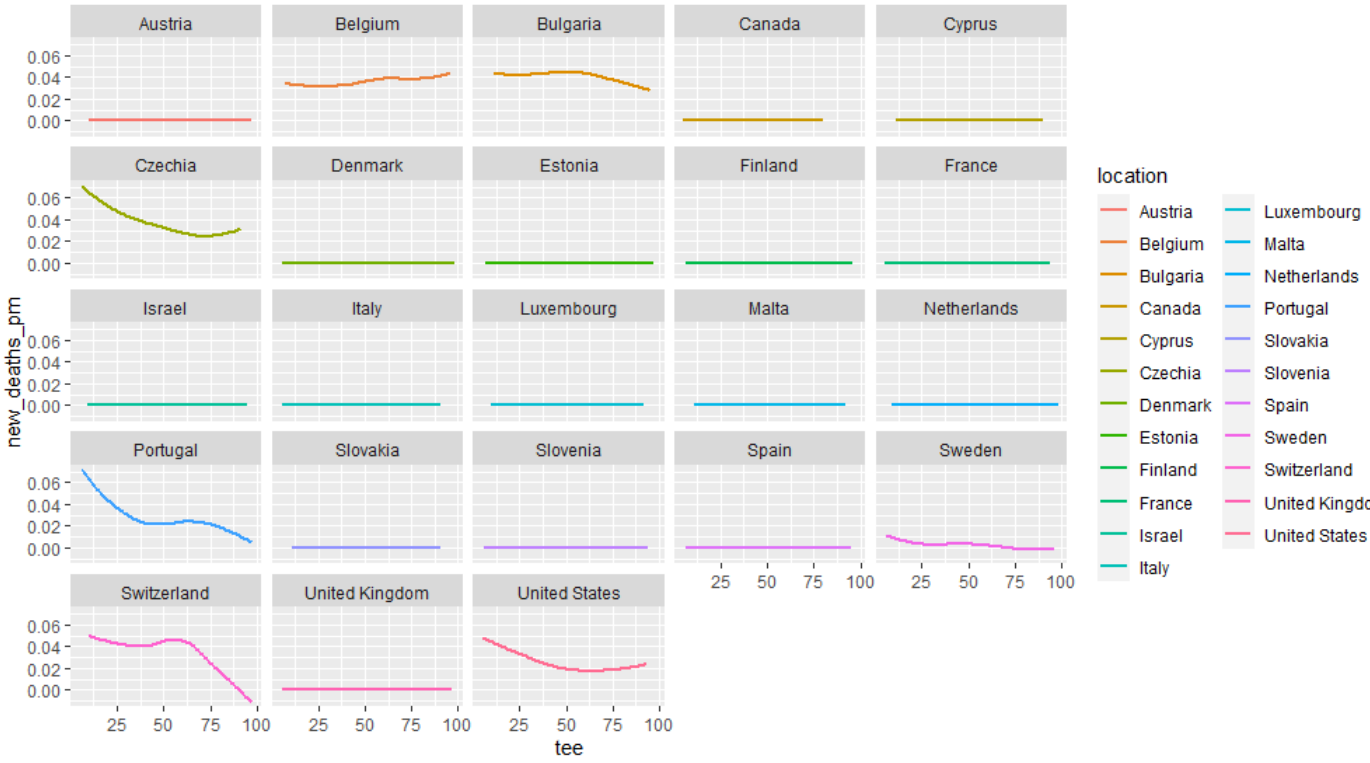

Figure 5.7: New Deaths per Million.

Figure 5.7 is a display of coefficient functions estimated by rank method for the effect of new deaths on excess mortality, the affected countries appear to show a decline in excess mortality as new deaths increases. This may

be a result of time lag which has to do with the reporting of new deaths.

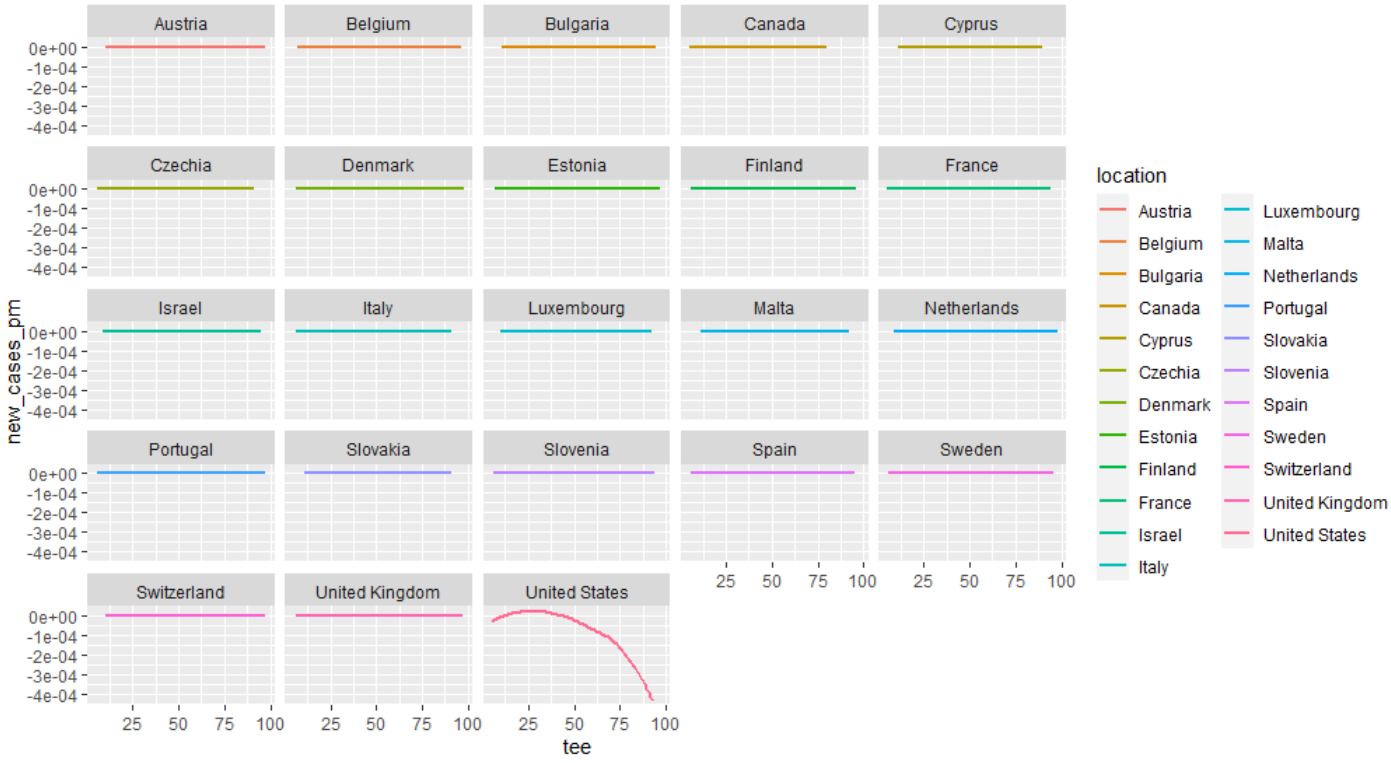

Figure 5.8: New Cases per Million.

In Figure 5.8 only the US shows a decline in excess mortality as new cases increases. Other countries show no relationship between new cases and excess mortality.

Figure 5.9: ICU Patients per Million.

Figure 5.10: Hospital Patients per Million.

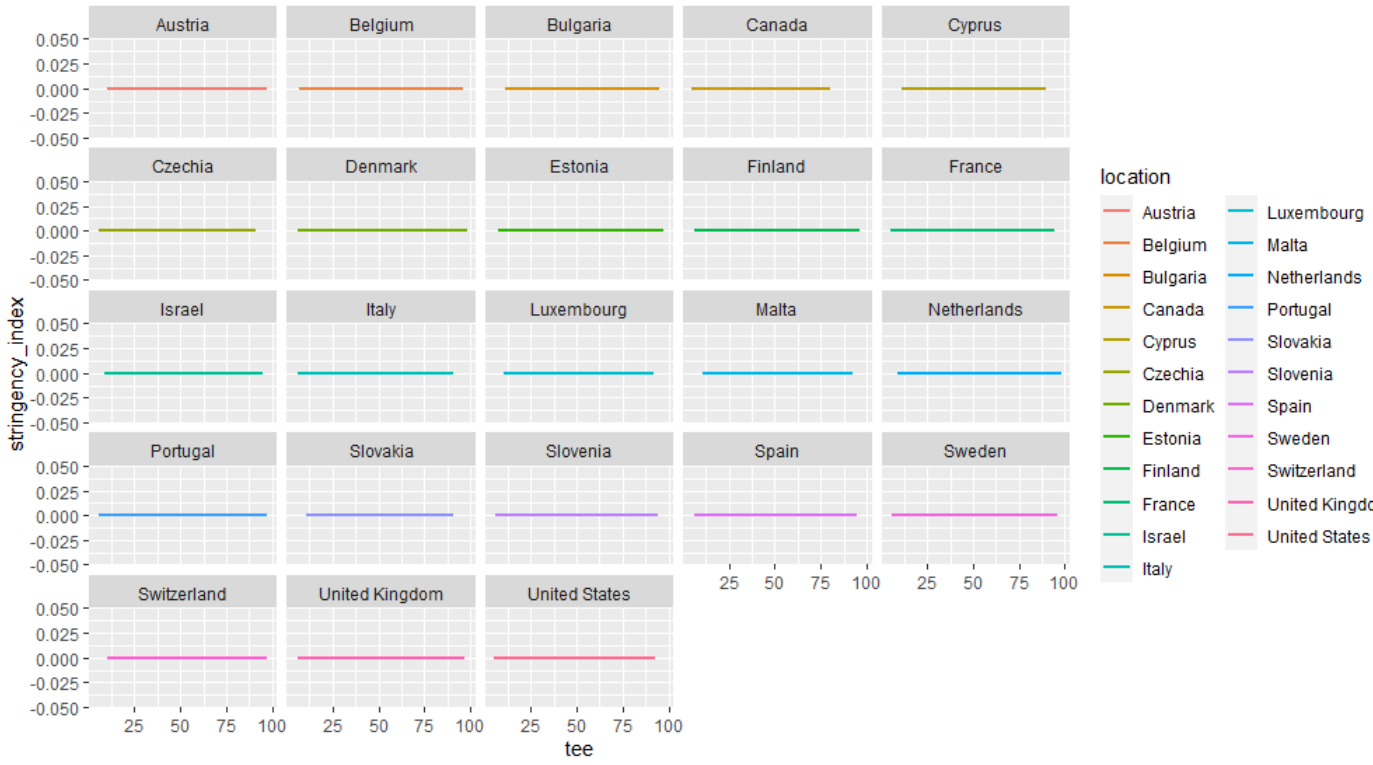

Figure 5.11: Stringency Index.

In Figures 5.9 and 5.10, only In the US does the result show a pattern between ICU patients vs excess mortality and hospital patients vs excess mortality, respectively,. excess mortality decreases with ICU patients in the US but increases with hospital patients but in 5.11, stringency index does not appear to have relationship with excess mortality for any of the countries studied in this work. Stringency index may not be needed to describe excess mortality for the rank method.

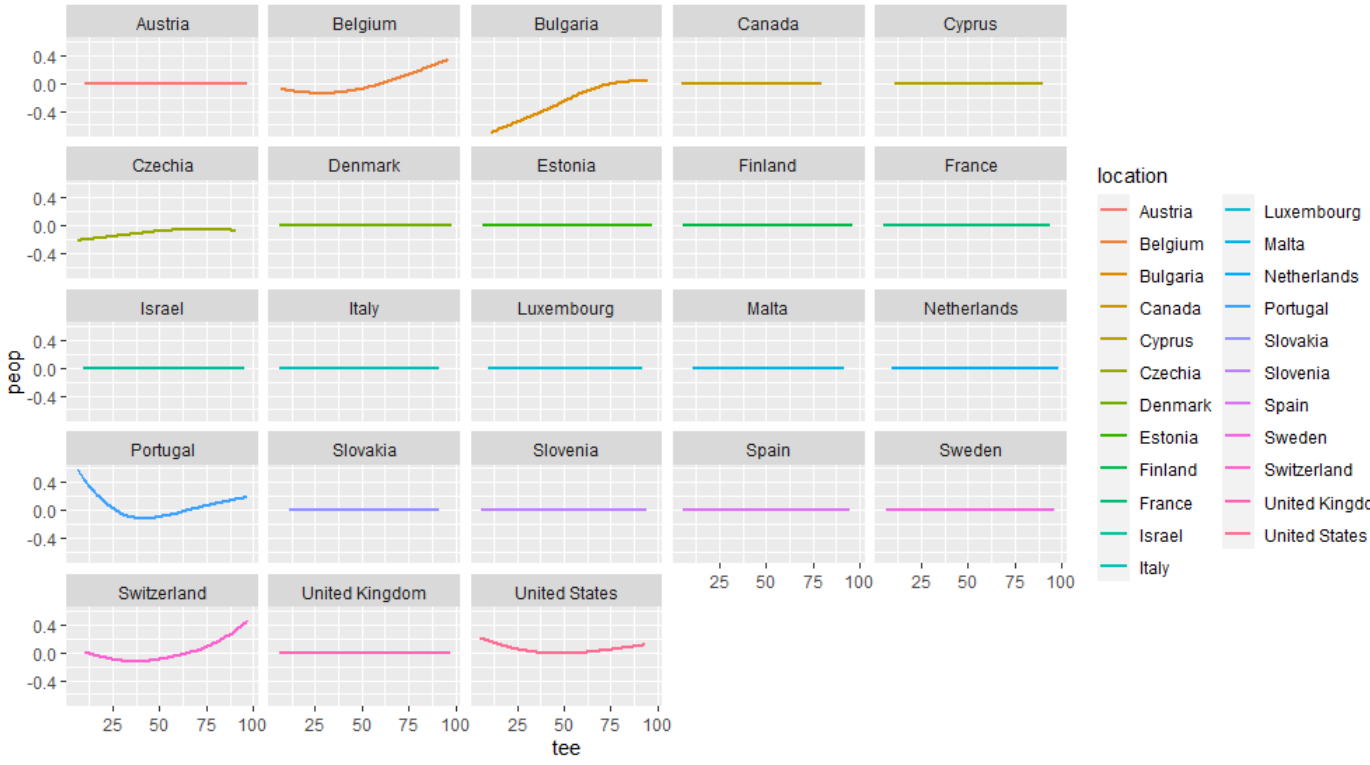

Figure 5.12: People Vaccinated.

Figure 5.9 shows a relationship between people vaccinated and excess mortality for Belgium, Bulgaria, Portugal, Switzerland and the US being obviously not constant. Rising for these countries except for Portugal where it showed a significant drop after which it began to rise within the time range. For the parametric part, the estimated parameters are as follows:

$\beta_1 = -1.3273 \times 10^{-4}, \beta_2 = 1.9908 \times 10^{-2}, \beta_3 = -1.0518 \times 10^{-2}, \beta_4 = -7.8982 \times 10^{-2}, \beta_5 = -9.1290 \times 10^{-6}, \beta_6 = 5.8820 \times 10^{-2}, \beta_7 = -4.0392$. Where β_1 is the coefficient of "population density", β_2 is the coefficient of "median age", β_3 is the coefficient of "aged 65 or older", β_4 is the coefficient of "aged 70 or older", β_5 is the coefficient of "gdp per capita", β_6 for "life expectancy" and β_7 the coefficient of "human development index". In all these, β_5 appears to have the least effect, if any, on excess mortality, while human development index appears to have the most effect, with a negative value, showing a negative relationship with excess mortality.

Mean Square Prediction Error Table	
Rank	6.3272
LS	39.1809

Table 5.1: Mean Square Error for Rank vs LS

5.6 Discussion

Overall, the MSPE from a per-country 5-fold cross-validation table (Table 5.1) shows the Rank method having a better performance than the Least squares method. Though some of the results obtained do not appear to reflect reality, this maybe due to so many factors which may not have been captured within the scope of this work, for example, there were a lot of missing values which lead to one of the reasons 23 locations (including countries) were selected from a total of 237 as others had significantly less or no data to work with for most of the variables like in most underdeveloped countries. Moreover, we may have an issue of reverse causation with some of the variables studied. For instance, increase in stringency occurred following increase in deaths. So, an increasing relationship

here may be more an indicator of this association than the effect of the stringency measures themselves. We believe adjusting the model may be another way to improve the results of our analysis, because as discussed in Hastie and Tibshirani (1993), where they considered different variations of varying coefficient models. Several of the predictors have lagged effects. It is our expectation that the use of a varying coefficient model that includes time-lagged covariates will better capture the drivers of the dynamics of excess mortality. Such models, are however, beyond the scope of this current dissertation.

Chapter 6

Conclusion

This dissertation proposed a robust estimation and selection procedure based on ranks for the semi-varying coefficient model. Such models are useful for the analysis of large spatiotemporal data. Consistency and asymptotic normality of the estimator were established under mild regularity conditions. These may be used to provide Wald-type tests as well as confidence intervals. Moreover, selection consistency and efficiency were established for the proposed procedure to select both functional and fixed regression coefficients. This work is a culmination of recent development in the use of rank estimators for estimation and selection of non- and semi-parametric regression models including single-index models (Bindele, Abebe, and Meyer, 2018; Bindele, Abebe, Zeng, 2019) and single-index varying-coefficient models (Sun et al. 2019, 2021). Another related work is the rank estimation of the partial linear model with responses missing at random (Bindele and Abebe, 2015), but this only considered a univariate nonparametric component which was treated as a nuisance parameter. This dissertation studied a rank-based estimation and selection procedure for semi-varying coefficient models which generalizes many of the models studies previously.

In addition to the theoretical results establishing consistency and efficiency, we also performed a simulation study that provided important insight into the performance of the proposed procedure. For instance, although as expected the efficiency results for non-zero coefficients are comparable to least squares in the case of Gaussian data and far superior to least squares in the case of heavy tailed or contaminated data, the selection consistency results of the proposed rank-based approach were somewhat remarkable. The rank estimator does almost uniformly better than the least squares estimator in correctly estimating of the zero functional coefficients even in the case of Gaussian data. Thus for functional coefficient

estimation and selection, the rank estimation appears to have an "oracleness" in the sense of efficiency and selection consistency. We suspect this is theoretically true as a similar observation was made following an extensive simulation study by Sun et al. (2021) which gave a rank-based estimation and selection procedure of the functional regression coefficients for the single-index varying coefficient regression model.

We also studied the semi-varying coefficient model to study a high-frequency global COVID-19 data. This cross-country dataset had variables that may affect COVID-19 excess mortality that only changed at the level of a country. These included variables like the proportion of the population over 70 years old or the country's GDP per capita, which are unlikely to show changes over a short pandemic. On the other hand, the dataset contained variables that were expected to substantially vary over time for each country. These included variables like the number of new COVID-19 cases or the number of ICU patients. We followed a hierarchical modeling approach where the functional coefficients varied by time and were estimated for each country. These were then fused to estimate the fixed coefficients that only varied across countries. The mixed backfitting computational algorithm iterates between the two scales until convergence.

There are several issues that need to be considered in the future. The first is the computational burden. The current computational approach, while amenable for distributed computing on high-performance computing clusters, is extremely expensive. This is because the approach internally performs iterated reweighted least squares steps within each backfitting step making the computational time significantly more demanding than the least squares approach. There are other avenues that may be considered. For example, there are one-step quadratic approximation procedures for the rank-estimator in the case of linear model estimation. It would be interesting to see if analogous procedures can be developed for the estimation of the more complex semiparametric models.

Another potential future direction is how to impute missing values in the context of the rank estimation of semi-varying coefficient models. For the high frequency global COVID-19 dataset studied in this dissertation, we had to eliminate several countries since they were missing parts of the data. The dataset would have been richer if there are rank-based procedures to impute the missing data, especially given that the missingness is possible not at random (non-ignorable). It would be of interest to explore ideas such as exponential tilting (Kim and Yu, 2011) to explore this direction similar to Bindele and Adekpedjou (2019). Computation

As we noted in the COVID-19 study, some of the variables (eg. stringency index and excess mortality) may have coupled dynamics. These are often described using predator-prey type differential equation models. It is of interest to explore versions of semi-varying coefficient models that account for coupled dynamics using lagged variables. These models are somewhat complex to analyze in the context of semi-varying coefficient models. While estimation appears to be within reach (Sun et al., 2019), model selection in this context will require substantial further research.

Finally, further theoretical investigations of the hierarchical semi-varying coefficient model similar to the one used in the analysis of the COVID-19 data would be of great interest. A rank based model selection and estimation procedure which at the same time quantifies two levels of variation in response would provide means to study data that have complex features but are currently analyzed using linear mixed models as a best approximation. This dissertation has provided an initial point in extending the rank-based procedures for mixed models in Abebe et al. (2016) for longitudinal data in Kloke, McKean, and Rashid (2009) for exchangeable, cluster-correlated data.

Appendix A

Simulation Code

```
##############################################################################
## Multivariate Normal Random Variates
##############################################################################
rmvnorm <- function(n, p, rho){
  X <- matrix(nrow = n, ncol = p)
  sigma <- matrix(nrow = p, ncol = p) #variance-covariance matrix
  for (i in 1:p){
    for (j in 1:p){
      sigma[i,j] = rho^abs(i-j)
    }
  }
  L <- chol(sigma) #Cholesky decomposition
  for (k in 1:n){
    Z <- rnorm(p)
    X[k,] <- L%*%Z
  }
  return(X)
}


##############################################################################
## Mixed Normal Random Variates
##############################################################################
rmixnorm = function(n, p, mu1, sd1, mu2, sd2)
{
  group = rbinom(n, size = 1, prob = p);
  m = sum(group);
  z = numeric(n);
  z[group == 1] = rnorm(m, mu1, sd1);
  z[group == 0] = rnorm(n - m, mu2, sd2);
  return(z)
}
```

```
##############################################################################
## Simulation Code
##############################################################################
n = sample size
rep = replications

for (i in 1:rep){

  x <- rmvnorm(n,p,0.5)

  e <- rmixnorm(n, 0.95, 0, 1, 0, 10)

  z <- matrix(runif(q*n,-1,1),n,q)
  u <- matrix(runif(n,-1,1),n,1)
  f <- (1+3*u^2)+3*exp(-u*u)*x[,1]+1.5*sin(pi*u)*x[,2]+0.8*u*x[,3]
  Z <- 1.5*z[,1] + z[,4] + 3*z[,5]

  y <- f + Z + e

  # run rank variable selection
  fit <- vcm_rvs(x,y,u,9)

  #Set initial alpha and beta
  a.previous <- fit$func
  b.previous <- as.matrix(rep(0, q))

  A <- 10
  B <- 10

  while(A > .5 && B > .01){

    #Estimate new beta using initial alpha
    y2 <- y - (t(fit$w.theta.opt))%*%(fit$basis.coef)

    fit2 <- cv.glmnet(z, y2, alpha = 1, lambda = grid)
    bestlam <- fit2$lambda.min
    fit22 <- glmnet(z, y2, alpha = 1, lambda = grid)
    beta <- as.matrix(predict(fit22, type = "coefficients", s = bestlam))[-1]

    y1 <- y - z%*%beta

    fit <- vcm_rvs(x,y1,u,9)

    a.new <- fit$func
```

```
    b.new <- beta

    A <- norm(as.matrix(a.previous - a.new), "i")/norm(as.matrix(a.previous), "i")
    B <- norm(as.matrix(b.previous - b.new), "i")

    a.previous <- a.new
    b.previous <- b.new

    print(c(A, B))
  }
}
```

Bibliography

Abebe, A., McKean, J. W., and Bindele, H. F. (2012). On the Consistency of a Class of Nonlinear Regression Estimators. *Pakistan Journal of Statistics and Operation Research*, 8(3):543.

Abebe, A., McKean, J. W., Kloke, J. D., and Bilgic, Y. K. (2016). Iterated Reweighted Rank-Based Estimates for GEE Models. In Liu, R. Y. and McKean, J. W., editors, *Robust Rank-Based and Nonparametric Methods*, volume 168, pages 61–79. Springer International Publishing, Cham. Series Title: Springer Proceedings in Mathematics & Statistics.

Aron, J., Muellbauer, J., et al. (2020). Measuring excess mortality: the case of england during the covid-19 pandemic. *Institute for New Economic Thinking, Oxford Working Paper*, 11.

Ballin, M., Bergman, J., Kivipelto, M., Nordström, A., and Nordström, P. (2021). Excess Mortality After COVID-19 in Swedish Long-Term Care Facilities. *Journal of the American Medical Directors Association*, 22(8):1574–1580.e8.

Baltagi, B. H., editor (2003). *Nonstationary panels, panel cointegration, and dynamic panels*. Number 15 in Advances in econometrics. Elsevier, Amsterdam Heidelberg, 1. ed., 2. impr edition.

Beaney, T., Clarke, J. M., Jain, V., Golestaneh, A. K., Lyons, G., Salman, D., and Majeed, A. (2020). Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? *Journal of the Royal Society of Medicine*, 113(9):329–334.

Bindele, H. F. and Abebe, A. (2015). Semi-parametric rank regression with missing responses. *Journal of Multivariate Analysis*, 142:117–132.

Bindele, H. F., Abebe, A., and Meyer, K. N. (2018a). General rank-based estimation for regression single index models. *Annals of the Institute of Statistical Mathematics*, 70(5):1115–1146.

Bindele, H. F., Abebe, A., and Meyer, N. K. (2018b). Robust confidence regions for the semi-parametric regression model with responses missing at random. *Statistics*, 52(4):885–900.

Breiman, L. (2012). Estimating Optimal Transformations for Multiple Regression and Correlation. *Journal of the American Statistical Association*, 80.

Chandran, A. M. M. (2020). Missing Data Not At Random: Characterization of Targeted Interference in Wireless Networks. In *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 262–263, Bologna, Italy. IEEE.

Chen, Y.-H., Glymour, M., Riley, A., Balmes, J., Duchowny, K., Harrison, R., Matthay, E., and Bibbins-Domingo, K. (2021). Excess mortality associated with the COVID-19 pandemic among Californians 18–65 years of age, by occupational sector and occupation: March through November 2020. *PLOS ONE*, 16(6):e0252454.

Chu, T., Zhu, J., and Wang, H. (2011). Penalized maximum likelihood estimation and variable selection in geostatistics. *The Annals of Statistics*, 39(5).

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *The Annals of Statistics*, 27(5).

Fan, J. and Zhang, W. (2000). Simultaneous Confidence Bands and Hypothesis Testing in Varying-coefficient Models. *Scandinavian Journal of Statistics*, 27(4):715–731.

Gilbert, H. F. and O'Leary, M. H. (1975a). Modification of arginine and lysine in proteins with 2,4-pentanedione. *Biochemistry*, 14(23):5194–5199.

Gilbert, H. F. and O'Leary, M. H. (1975b). Modification of arginine and lysine in proteins with 2,4-pentanedione. *Biochemistry*, 14(23):5194–5199.

Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: a roughness penalty approach*. Number 58 in Monographs on statistics and applied probability. Chapman & Hall, London ; New York, 1st ed edition.

Gu, C., Heckman, N., and Wahba, G. (1992). A note on generalized cross-validation with replicates. *Statistics & Probability Letters*, 14(4):283–287.

Haklai, Z., Aburbeh, M., Goldberger, N., and Gordon, E.-S. (2021). Excess mortality during the COVID-19 pandemic in Israel, March–November 2020: when, where, and for whom? *Israel Journal of Health Policy Research*, 10(1):17.

Hallin, M., Lu, Z., and Tran, L. T. (2004). Local linear spatial regression. *The Annals of Statistics*, 32(6).

Hardle, W., Hall, P., and Ichimura, H. (1993). Optimal Smoothing in Single-Index Models. *The Annals of Statistics*, 21(1).

Hastie, T. and Tibshirani, R. (1993a). Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

Hastie, T. and Tibshirani, R. (1993b). Varying-Coefficient Models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779.

Hettmansperger, T. P., McKean, J. W., Hettmansperger, T. P., Hettmansperger, T. P., McKean, J. W., and McKean, J. W. (2011). *Robust nonparametric statistical methods*. CRC Press, Boca Raton. OCLC: 851040536.

Huang, J. Z. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128.

Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *The Annals of Mathematical Statistics*, pages 1449–1458.

Jennrich, R. (1969). Asymptotic Properties of Non-Linear Least Squares Estimators. *The Annals of Mathematical Statistics*, 40.

Kananen, L., Eriksdotter, M., Boström, A., Kivipelto, M., Annetorp, M., Metzner, C., Bäck Jerlardtz, V., Engström, M., Johnson, P., Lundberg, L., Åkesson, E., Sühl Öberg, C., Hägg, S., Religa, D., Jylhävä, J., and Cederholm, T. (2021). Body mass index and Mini Nutritional Assessment-Short Form as predictors of in-geriatric hospital mortality in older adults with COVID-19. *Clinical Nutrition*, page S0261561421003605.

Kapitsinis, N. (2020). The underlying factors of the COVID-19 spatially uneven spread. Initial evidence from regions in nine EU countries. *Regional Science Policy & Practice*, 12(6):1027–1045.

Kim, J. K. and Yu, C. L. (2011). A Semiparametric Estimation of Mean Functionals With Nonignorable Missing Data. *Journal of the American Statistical Association*, 106(493):157–165.

Kloke, J. D. and McKean, J. W. (2012). Rfit: Rank-based Estimation for Linear Models. *The R Journal*, 4(2):57–64.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327.

Lu, Z., Tang, Q., and Cheng, L. (2014). Estimating spatial quantile regression with functional coefficients: A robust semiparametric framework. *Bernoulli*, 20(1).

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., and Roser, M. (2020). Coronavirus pandemic (covid-19). *Our world in data*.

Rossen, L. M., Ahmad, F. B., Anderson, R. N., Branum, A. M., Du, C., Krumholz, H. M., Li, S.-X., Lin, Z., Marshall, A., Sutton, P. D., and Faust, J. S. (2021). Disparities in

Excess Mortality Associated with COVID-19 — United States, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 70(33):1114–1119.

Schimek, M. G. (2000). Estimation and inference in partially linear models with smoothing splines. *Journal of Statistical Planning and Inference*, 91(2):525–540.

Schumaker, L. L. (2007). *Spline functions basic theory*. Cambridge University Press, Cambridge; New York. OCLC: 1105511736.

Share, J. B. (1976). Review of drug treatment for Down's syndrome persons. *American Journal of Mental Deficiency*, 80(4):388–393.

Shi-qin, L., Juan, W., and Gang, Z. (2012). Estimation on Generalized Semi-varying Coefficient Models. *Physics Procedia*, 33:949–953.

Sievers, G. L. and Abebe, A. (2004). Rank estimation of regression coefficients using iterated reweighted least squares. *Journal of Statistical Computation and Simulation*, 74(11):821–831.

Stang, A., Standl, F., Kowall, B., Brune, B., Böttcher, J., Brinkmann, M., Dittmer, U., and Jöckel, K.-H. (2020). Excess mortality due to COVID-19 in Germany. *Journal of Infection*, 81(5):797–801.

Stokes, A. C., Lundberg, D. J., Elo, I. T., Hempstead, K., Bor, J., and Preston, S. H. (2021). COVID-19 and excess mortality in the United States: A county-level analysis. *PLOS Medicine*, 18(5):e1003571.

Sun, W., Bindele, H. F., Abebe, A., and Correia, H. (2019). General local rank estimation for single-index varying coefficient models. *Journal of Statistical Planning and Inference*, 202:57–79.

Sun, W., Bindele, H. F., Abebe, A., and Correia, H. E. (2021). Robust functional coefficient selection for the single-index varying coefficients regression model. *Journal of Statistical Computation and Simulation*, 91(8):1681–1697.

Tang, Q. (2014). Robust estimation for functional coefficient regression models with spatial data. *Statistics*, 48(2):388–404.

Tang, Q. and Cheng, L. (2009). B-spline estimation for varying coefficient regression with spatial data. *Science in China Series A: Mathematics*, 52(11):2321–2340.

Tran, L. T. (1990). Kernel density estimation on random fields. *Journal of Multivariate Analysis*, 34(1):37–53.

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., Comfort, H., Zheng, P., Aravkin, A. Y., Bisignano, C., Barber, R. M., Alam, T., Fuller, J. E., May, E. A., Jones, D. P., Frisch, M. E., Abbafati, C., Adolph, C., Allorant, A., Amlag, J. O., Bang-Jensen, B., Bertolacci, G. J., Bloom, S. S., Carter, A., Castro, E., Chakrabarti, S., Chattopadhyay, J., Cogen, R. M., Collins, J. K., Cooperrider, K., Dai, X., Dangel, W. J., Daoud, F., Dapper, C., Deen, A., Duncan, B. B., Erickson, M., Ewald, S. B., Fedosseeva, T., Ferrari, A. J., Frostad, J. J., Fullman, N., Gallagher, J., Gamkrelidze, A., Guo, G., He, J., Helak, M., Henry, N. J., Hulland, E. N., Huntley, B. M., Kereselidze, M., Lazzar-Atwood, A., LeGrand, K. E., Lindstrom, A., Linebarger, E., Lotufo, P. A., Lozano, R., Magistro, B., Malta, D. C., Månsson, J., Mantilla Herrera, A. M., Marinho, F., Mirkuzie, A. H., Misganaw, A. T., Monasta, L., Naik, P., Nomura, S., O'Brien, E. G., O'Halloran, J. K., Olana, L. T., Ostroff, S. M., Penberthy, L., Reiner Jr, R. C., Reinke, G., Ribeiro, A. L. P., Santomauro, D. F., Schmidt, M. I., Shaw, D. H., Sheena, B. S., Sholokhov, A., Skhvitaridze, N., Sorensen, R. J. D., Spurlock, E. E., Syailendrawati, R., Topor-Madry, R., Troeger, C. E., Walcott, R., Walker, A., Wiysonge, C. S., Worku, N. A., Zigler, B., Pigott, D. M., Naghavi, M., Mokdad, A. H., Lim, S. S., Hay, S. I., Gakidou, E., and Murray, C.

J. L. (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020–21. *The Lancet*, 399(10334):1513–1536.

Wang, K. (2018). Variable selection for spatial semivarying coefficient models. *Annals of the Institute of Statistical Mathematics*, 70(2):323–351.

Wong, H., Ip, W.-c., and Zhang, R. (2008). Varying-coefficient single-index model. *Computational Statistics & Data Analysis*, 52(3):1458–1476.

Wu, C.-F. (1981). Asymptotic Theory of Nonlinear Least Squares Estimation. *The Annals of Statistics*, 9(3).