Nonparametric Analysis of Time Series with Complex Features

by

Cong Feng

(Under the direction of Lily Wang and Lynne Seymour)

Abstract

Time series with complex features, such as non-linearity, high-dimensionality and functional structures, have inspired many interests in the statistics community recently. In this dissertation, novel non/semi-parametric methods are investigated to model and/or forecast such time series based on spline estimation. We first consider a class of semi-parametric GARCH models with additive autoregressive components linked together by a dynamic coefficient. We propose estimators which are computationally efficient and theoretically reliable for the additive components and the dynamic coefficient. We also propose a framework to model and forecast the functional time series via functional principal component analysis. For comparing the functional derivatives of regression functions from two groups, we develop a novel method to construct simultaneous confidence bands for the difference of derivatives. The performance of the proposed methods is evaluated by various simulated processes and real datasets in finance and climate.

INDEX WORDS: nonparametric, non-linearity, high-dimensionality, functional time series, functional principal component analysis, B-splines, simultaneous confidence bands

Nonparametric Analysis of Time Series with Complex Features

by

Cong Feng

B.S., Zhejiang University, China 2005

M.S., University of Georgia 2008

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Department of Statistics

Athens, Georgia

2012

Nonparametric Analysis of Time Series with Complex Features

by

Cong Feng

Approved:

Major Professors:	Lily Wang
	Lynne Seymour
Committee:	Jeongyoun Ahn
	Jaxk Reeves
	T.N. Sriram
	Xiangrong Yin

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012

Dedication to My Parents

Acknowledgments

First and foremost, I would like to thank my PhD advisors, Dr. Lily Wang and Dr. Lynne Seymour, for their long-term strong support and tremendous patience in mentoring my PhD research, revising my manuscripts, providing me with numerous opportunities to engage with the statistical research community, such as attending workshops and conferences, peer-reviewing scientific articles, etc. Under their guidance and influence, I am not only growing to be a great and enthusiastic statistician, but also striving to be a person with great personality and confidence to embrace everyday life with broader vision, bigger heart and grateful attitude.

I would like to thank my committee members Dr. Jeongyoun Ahn, Dr. Jaxk Reeves, Dr. T.N. Sriram and Dr. Xiangrong Yin, for contributing their time, efforts and scientific insights in supporting my PhD studies at UGA. I must acknowledge my research collaborators: Dr. Lijian Yang, Dr. Qiongxia Song, Ms. Guanqun Cao and Dr. David Stooksbury for their stimulating discussion and input into my research work. I would like to extend my special thanks to Dr. Jeongyoun Ahn, Dr. Yehua Li and Dr. John A. Knox who wrote very supportive recommendation letters to bring me into this great

statistics department in 2007. I also want to thank Dr. Jaxk Reeves for recruiting me as a graduate consultant in our Statistical Consulting Center; this is an invaluable experience for me. There is a long list of professors I am grateful for due to their excellent teaching: Dr. T.N. Sriram, Dr. Gauri Datta, Dr. Daniel Hall, Dr. William McCormick, Dr. Xiangrong Yin, Dr. Jeongyoun Ahn, Dr. Jaxk Reeves and many others. They helped me to lay a solid foundation in methodology and theory in statistics. I also want to express my gratitude to Dr. John Stufken, Dr. T.N. Sriram, Dr. Xiangrong Yin and my two advisors for giving me constructive career advice and support. I am thankful for the tremendous assistance provided by administrative staff in the Statistics Department, with special thanks to Tim Cheek, Daphney Smith and Julie Davis.

My time in US was made enjoyable in large part due to many talented colleagues and close friends I met here, including Ben Gentry, Lina Wang, Xiaoling Yang, Qianqiu Zhu, Kun Xu, Jia Xu, Ashley Askew, Yufei Liu, Tianle Hu, Jun Wang, Ziming Zhao, Hai Pan, Qimei Ran and many others.

Last but not least, I would like to thank my beloved parents for their perpetual and absolute love and trust.

Contents

List of Tables

List of Figures

Chapter 1

Introduction and Literature Review

1.1 Time Series with Complex Features

Time series is a long-standing research area in traditional statistics, which is closely related with many natural and man-made phenomena in real life, such as climate evolution, economic dynamics and financial returns, etc. One problem facing the statistics community today is modeling of time series data with complex structures collected by financial companies, federal agencies, research institutes and other organizations. Such series pose formidable challenges in both applications and theory because they exhibit features such as non-linearity, high-dimensionality and complex functional structures.

As an example, consider $\{\mathbf{X}_t^T, Y_t\}_{t=1}^n = \{X_{t1}, ..., X_{td}, Y_t\}_{t=1}^n$, a $d+1$ dimensional time series where d can be very large. Suppose we model Y_t as:

$$Y_t = m(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\,\varepsilon_t, m(\mathbf{X}_t) = E(Y_t|\mathbf{X}_t),\ t = 1, ..., n,$$

where $E(\varepsilon_t|\mathbf{X}_t) = 0$, $E(\varepsilon_t^2|\mathbf{X}_t) = 1$, and m, σ are unknown d-dimensional functions we want to estimate nonoparametrically. In the above model, $\mathbf{X}_t$ may usually consist of lagged values of Y_t or other covariates. It is possible that such time series exhibit non-linearity and high-dimensionality. There are many good examples of such time series: for example, the weekly 3-month Treasury bill secondary market rate from 1970 to 1997 of 1460 observations shown in Fig 1.1 (refer to Example 4.6 in Tsay (2005)). This data was obtained from Federal Reserve Bank of St. Louis. The series of rates follow a nonlinear instead of a simple linear time series model. Local linear regression methods have been applied to address the complexities of modeling this time series.

Another example is a functional time series. Suppose $\{Y_t(u_j)\}_{t=1}^n$ curves with temporal dependence on some equally or unequally spaced grid $\{u_j\}_{j=1}^m$ at time points $t = 1, ..., n$. At the t-th time point, its sample path $\{u_j, Y_{tj}\}$ is a noisy realization of a smooth function $X_t(u_j)$ and the series is defined by:

$$Y_{tj} = X_t(u_j) + \sigma_{tj}\varepsilon_{tj},\ j = 1, ..., m,\ t = 1, ..., n$$

with independent errors ε_{tj} satisfying $E(\varepsilon_{tj}) = 0, E(\varepsilon_{tj}^2) = 1$. The main interest is to forecast the h-ahead curves $\{Y_t(u_j)\}_{t=n+1}^{n+h}$. Hyndman and Ullah (2007) showed a

motivational example of a functional time series. Fig 1.2 is the plot of age-specific log mortality rate curves from 1816 to 2006. In this example, Y_{tj} is the observed log death rate at age u_j in t-th year. Forecasting the future mortality curves is very important for the insurance and pension industry, and even the government. But the literatures related to dealing with functional time series is sparse due to certain inherent complexities.

Motivated by these challenging questions, researchers have developed various cutting-edge nonparametric techniques to model and forecast time series with such unique features; see Fan and Gijbels (1996), Bosq (1998), Fan and Yao (2003).

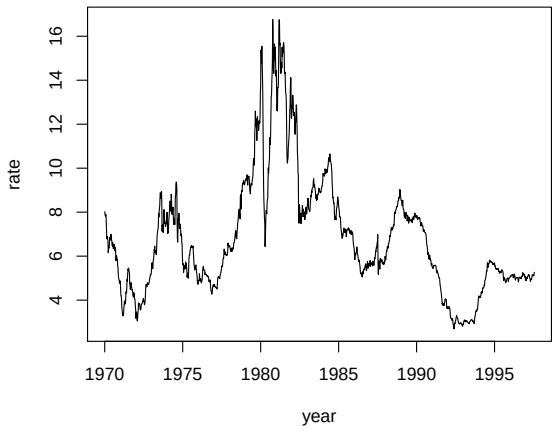


Figure 1.1: Time plot of U.S. weekly 3-month Treasury bill rate in the secondary market from 1970 to 1997.

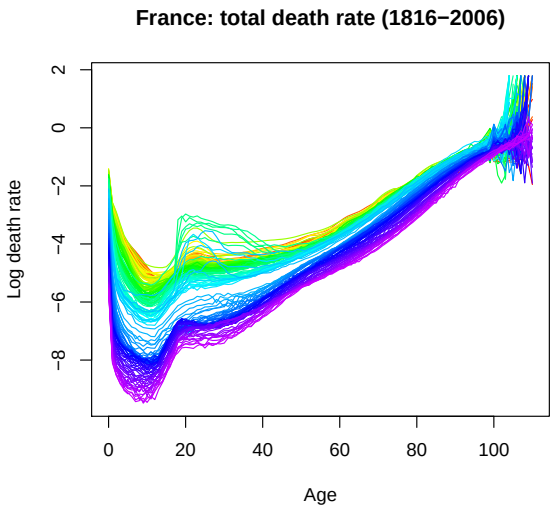


Figure 1.2: Total death rates in France from 1816-2006. The colors indicate the time ordering of the curves in the same order as the colors in a rainbow (the oldest curves are red and the most recent curves are purple).

1.2 Literature Review

1.2.1 Volatility Model and News Impact Curve

Since the introduction of ARCH model by Engle (1982), there has been an explosion of papers analyzing conditional volatility; see GARCH model (Bollerslev 1986), exponential GARCH model (Nelson 1991), integrated GARCH model (Bollerslev and Engle 1993), and threshold GARCH models (Glosten *et al.* 1993). These models have been used in many financial applications of interest rate data, stock return data, foreign exchange data, etc. All of these new and existing models regress the conditional volatility

parametrically on lagged values of shocks and lagged values of volatility. These empirical studies also reveal that there is a long term persistence in the effects of shocks in the period t onto the conditional volatility in the future period $t + s$.

In the economic world, it is well known that news drives the market: good news lifts the market, while bad news depresses the market. Nelson (1991) emphasized the interesting "leverage effect" which indicates that negative shocks to asset markets increase predictable volatility more than positive shocks. Engle and Ng (1993) proposed a standard measure of how news (shocks) influences stock volatility - a news impact curve for these conditional volatility models. The news impact curve describes the relationship between surprises in conditional volatility and shocks. The asymmetric effects of good news and bad news onto volatility suggests the asymmetric shape of the news impact curve; see the discussion in Engle and Ng (1993). The variants of GARCH model, such as GJR model (Glosten *et al.* 1993) and EGARCH model (Nelson 1991), allow such leverage effects. In order to increase the flexibility of models, researchers have introduced the non/semi-parametric (G)ARCH models; see for example, Pagan and Schwert (1990), Engle and Ng (1993), Masry and Tjøstheim (1995), Härdle and Tsybakov (1997), Hafner (1998), Härdle, *et al.* (1998), Bühlmann and McNeil (2002), Linton and Mammen (2005) and Yang (2006). These models have achieved great success compared with the parametric GARCH models when applied to real data with many lags. However their applications have been hampered by the curse of dimensionality when smoothing the high-dimensional and strongly correlated time series.

1.2.2 Functional Data and Functional Time Series

Functional data analysis (FDA) is an emerging and promising research field developed in the last two decades. It provides a new perspective for the traditional statistical analysis. Most statistical analyses consider one or more observations taken from a sample of subjects. These observations can be a number or a vector of numbers. In FDA context, the observations are curves or surfaces which are in nature examples of functions. Therefore, curves and surfaces are called "functional data" and statistical methods to analyze such data are called "functional data analysis"; see Ramsay and Silverman (2005).

Functional data always arises from two scenarios: 1) functional observations are independent, such as longitudinal trajectories collected from a sample of subjects in randomized designs; 2) functional observations are dependent, such as temporally-dependent daily financial transaction curves and spatially-dependent geophysical patterns. The former scenario is the focus of functional data analysis nowadays. Functional principal component analysis (FPCA) is the most popular dimension reduction tool in FDA. A central issue in this case is estimation and inference of various components obtained by FPCA, such as mean functions, covariance functions and eigenfunctions; see Yao, *et al.* (2005a,b), Ferraty and Vieu (2006), Li and Hsing (2010) and Cao, *et al.* (2012a). In order to study the variability of mean functions, estimation and inference of functional derivatives also attract attention from many statisticians recently; see Liu and Müller (2009) and Hall, *et al.* (2009) and Cao, *et al.* (2012b). In the latter scenario, such

dependent functional data hasn't received as much attention as has independent functional data. Only a few papers can be found on this topic, including Hyndman and Ullah (2007), Shen (2009), and Hörmann and Kokoszka (2010). We consider a functional time series $\{X_t, t \in \mathbb{Z}\}$ as a dependent sequence of random functions $\{X_t(u), u \in [a, b]\}$ in terms of t. A key issue in analyses of such data is to take into account the temporal dependence of functional data.

1.3 Summary

As discussed above, the development of nonparametric statistics and functional data analysis in theory and methodology has provided us powerful tools to address the challenging problems arising in the time series with complex features. Many applications have been seen in financial time series, fertility/mortality dynamics and climate time series, etc. The contributions in this dissertation are to study the various non/semiparametric methods to model and/or forecast such complex time series based on spline estimation.

In Chapter 2, we consider a class of semiparametric GARCH models with additive autoregressive components linked together by a dynamic coefficient. We proposed estimators for the additive components and the dynamic coefficient based on spine smoothing. The estimation procedure involves only a small number of least squares operations; thus it is computationally efficient. Under regularity conditions, the proposed estimator of the parameter is root-n consistent and asymptotically normal. A simultaneous confi-

dence band for the nonparametric component is proposed by an efficient one-step spline backfitting. For the empirical financial return series, we find further statistical evidence of the asymmetric news impact function.

In Chapter 3, a novel method is proposed for forecasting time series of smooth curves, using functional principal component (FPC) analysis in combination with time series modeling and FPC scores forecasting. We achieve the smoothing, dimension reduction and prediction at the same time via expedient computation. The work is motivated by the demand to forecast the time series of economic functions, such as Treasury bond yield curves. Extensive simulation studies have been carried out to compare the prediction accuracy of our method with other existing methods.

In Chapter 4, we develop a new procedure to construct simultaneous confidence bands for the difference of derivatives of regression functions from two groups. We show that the proposed spline confidence bands are asymptotically efficient, as if there is no measurement error. Simulation experiments have provided strong evidence that corroborates with the asymptotic theory. The application of the proposed methodology to a temperature data collected in Athens, GA in US has shown us how the temperature transitions change in response to the major global atmospheric pressure oscillations.

1.4 References

Bollerslev, T. and Engle, R. (1993). Common persistence in conditional variances. *Econometrica* **61**, 167C186.

Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. New York: Springer-Verlag.

Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *Computational Statistics & Data Analysis* **40**, 665-683.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

Engle, R. F. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* **48**, 1749-1778.

Cao, G., Yang, L. and Todem, D. (2012a). Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics,* in press.

Cao, G., Wang, J., Wang, L. and Todem, D. (2012b). Spline confidence bands for functional derivatives. *Journal of Statistical Planning and Inference,* forthcoming.

Glosten, L. R., Jaganathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779-1801.

Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall: London.

Fan, J. and Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. New York: Springer.

Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer: Berlin.

Hall, P., Müller, H. G. and Yao, F. (2009). Estimation of functional derivatives. *Annals of Statistics* **37**, 3307-3329.

Härdle, W. and Tsybakov, A. B. (1997). Locally polynomial estimators of the volatility function. *Journal of Econometrics* **81**, 223-242.

Härdle, W., Tsybakov, A. B. and Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* **68**, 221-245.

Hörmann, S. and Kokoszka, P.(2010). Weakly dependent functional data. *The Annals of Statistics*, **38**, 1845-1884.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis* **51**, 4942C4956.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics* **38**, 3321-3351.

Linton, O. B. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* **73**, 771-836.

Liu, B. and Müller, H. G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association* **104**, 704-717.

Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econometric Theory* **11**, 258-289.

Nelson, D. (1991). Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica* **59**, 347-370.

Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* **45**, 267-290.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Second Edition. Springer Series in Statistics. Springer: New York.

Shen, H.(2009). On modeling and forecasting time series of curves. *Technometrics* **51**, 227-238.

Tsay, R. S. (2005). Analysis of financial time series. Wiley, New York.

Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica* **12**, 801-821.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873-2903.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590.

Chapter 2

Efficient Semiparametric GARCH Modeling of Financial Volatility[1]

[1]Wang, L., Feng, C., Song, Q., Yang, L. (2012). Efficient semiparametric GARCH modeling of financial volatility. *Statistica Sinica* **22**, 249-270. Reprinted here with permission of publisher.

Abstract

We consider a class of semiparametric GARCH models with additive autoregressive components linked together by a dynamic coefficient. We propose estimators for the additive components and the dynamic coefficient based on spline smoothing. The estimation procedure involves only a small number of least squares operations, thus it is computationally efficient. Under regularity conditions, the proposed estimator of the parameter is root-n consistent and asymptotically normal. A simultaneous confidence band for the nonparametric component is proposed by an efficient one-step spline backfitting. The performance of our method is evaluated by various simulated processes and a financial return series. For the empirical financial return series, we find further statistical evidence of the asymmetric news impact function.

KEYWORDS: B-spline, confidence band, knots, news impact curve, volatility.

2.1 Introduction

Forecasting financial market volatility is important in many applications such as portfolio selection, asset management, pricing of primary and derivative assets. Consider a time series $\{Y_t\}_{t=1}^{\infty}$ of the form $Y_t = \sigma_t \xi_t$, where the $\{\xi_t\}_{t=1}^{\infty}$'s are i.i.d with mean 0 and variance 1, and $\{\sigma_t^2\}_{t=1}^{\infty}$ denotes the conditional volatility series. Engle (1982) introduced the autoregressive heteroskedastic (ARCH) models for conditional volatility as a quadratic

function of past observations. For example, an ARCH model of order q is defined as

$$\sigma_t^2 = \gamma + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_q Y_{t-q}^2, \quad \gamma > 0,\ \alpha_i \geq 0,\ i = 1, ..., q.$$

Research on financial volatility models has grown tremendously since then; for example, the generalized autoregressive conditional heteroscedasticity (GARCH) models. The most popular version of the GARCH models is the GARCH$(1,1)$ model of Bollerslev (1986):

$$\sigma_t^2 = \gamma_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2, \quad \gamma_0 > 0,\ \alpha_0,\ \beta_0 \geq 0,$$

or equivalently $\sigma_t^2 = \beta_0 \sigma_{t-1}^2 + m_0(Y_{t-1})$, where $m_0(y) \equiv \alpha_0 y^2 + \gamma$ is the "news impact curve".

The quadratic form of the function $m_0(\cdot)$ had been questioned by many. For example, Glosten *et al.* (1993) proposed the following GJR model

$$\sigma_t^2 = \gamma_0 + \alpha_0 Y_{t-1}^2 + \delta_0 Y_{t-1}^2 I(Y_{t-1} < 0) + \beta_0 \sigma_{t-1}^2$$

with $m_0(y) \equiv \gamma + \alpha y^2 + \delta y^2 I\,(y < 0)$, allowing different "leverages" of good and bad news on m_0. For this reason, recent studies have introduced the non/semi-parametric (G)ARCH models to increase the flexibility of the class of models; see for example, Pagan and Schwert (1990), Engle and Ng (1993), Masry and Tjøstheim (1995), Härdle and Tsybakov (1997), Hafner (1998), Härdle, *et al.* (1998), Bühlmann and McNeil (2002), Linton and Mammen (2005) and Yang (2006). These models generalize and

outperform the parametric GARCH models when applied to real data with many lagged variables. However, smoothing high dimensional and strongly correlated time series data still presents great challenges in both computation and theory.

As an alternative, additive models (Stone (1985)) overcome these difficulties while keeping the flexibility of the models. Yang, *et al.* (1999) analyzed a multiplicative form of volatility using nonparametric smoothing. Carroll *et al.* (2002) and Yang (2002) proposed a truncated version of the nonparametric GARCH model with a finite number of lags J

$$\sigma_t^2 = \sum_{j=1}^{J} \beta_0^{j-1} m_0(Y_{t-j}), \quad \beta_0 \in [\beta_1, \beta_2]\,. \tag{2.1.1}$$

However, for small J, it may not capture the persistence of volatility for many time series; see Linton and Mammen (2005) and Yang (2006).

In this paper, we re-examine model (2.1.1) based on a data-driven lag selection procedure. Most of the existing methods rely on marginal integration kernel smoothing (Linton and Nielsen, 1995) or iterative approaches such as backfitting algorithm (Hastie and Tibshirani, 1990). The marginal integration can be computationally expensive if the selected number of lags J or sample size n is large, and it requires $O(n^3)$ operations (Hengartner and Sperlich, 2005). Moreover, n is required to be larger than $10,000$ for convergence when smoothing 10-dimensional data, so it is not routinely used in practice despite good theoretical properties. Widely used R/Splus packages `gam` and `mgcv`, based on backfitting with splines, provide convenient implementation in practice but lack theoretical justifications except some special cases in Opsomer and Ruppert (1997).

Our goal is to develop a simple but flexible semiparametric method with a well-justified theory and a fast algorithm to implement the method in practice. This is done by approximating the nonparametric components with polynomial splines. The use of spline smoothing traces back to Stone (1985), who first obtained the rate of convergence of the polynomial spline estimates for the generalized additive model. In volatility studies, Engle and Ng (1993) employed linear spline smoothing to estimate the news impact function, without pursuing asymptotic results.

Our approach allows for formal derivation of the asymptotic properties of the proposed estimators. We establish the $\sqrt{n}$-consistency and the asymptotic normality for the parameter estimator and L^2 convergence rate for the functional component. To examine the validity of certain forms of the volatility models, we provide a simultaneous confidence band for the news impact curve using the one-step spline-backfitted spline estimator in Song and Yang (2009b).

The rest of the paper is organized as follows. Section 2.2 gives details of the model specification, proposed methods of estimation and presents the asymptotic results. In addition, we discuss some alternative methods and the practical issue of lag selection. In Section 2.3, we describe a spline confidence band for the news impact curve. In Section 2.4, we report our findings in an extensive simulation study. An application to a real financial return data set is given in Section 2.5. Most of the technical proofs are contained in Section 2.7.

2.2 Methodology

2.2.1 Semiparametric GARCH with additive autoregressive structure

Consider a stationary time series $\{Y_t\}_{t=1}^T$, with $Y_t = \sigma_t \xi_t$, $t = 1, 2, ..., T$. We rewrite model (2.1.1) as the following additive autoregressive model,

$$Y_t^2 = c + \sum_{j=1}^{J} m_j(Y_{t-j}) + \epsilon_t, \quad \epsilon_t = \sigma_t^2 \left(\xi_t^2 - 1\right), \tag{2.2.1}$$

where the component functions $m_1(\cdot), ..., m_J(\cdot)$ are linked by a scalar parameter β_0 such that $m_j(y) = \beta_0^{j-1} m_1(y)$ for $j \geq 2$. Define the least squares risk function $R(\beta)$ over $\left[\beta_{1,}\beta_2\right]$ as,

$$R(\beta) = E\left[\sum_{j=1}^{J} \left\{m_j(Y_t) - \beta^{j-1} m_1(Y_t)\right\}^2\right]. \tag{2.2.2}$$

Since $R(\beta) = \sum_{j=1}^{J} \left\{\left(\beta_0^{j-1} - \beta^{j-1}\right)^2\right\} E\left\{m_1(Y_t)^2\right\}$ is a convex function with respect to β, β_0 is the unique minimizer of $R(\beta)$ over $\left[\beta_{1,}\beta_2\right]$. For identifiability, the component functions in (2.2.1) satisfy $E\left\{m_j\left(Y_t\right)\right\} = 0$, $j = 1, ..., J$.

Our interest is to estimate the news impact function m_1 and dynamic coefficient parameter β_0. To reach this goal, first we employ the polynomial spline smoothing to obtain the estimates $\hat{m}_j(\cdot)$ of the additive components $m_j(\cdot)$ without taking into account the parametric link of the components; Then we estimate the dynamic coefficient β_0 by using the link restriction between the additive components $\hat{m}_j(\cdot)$ $(j = 1, .., J)$. For simplicity

of notation, let us call the above approach the spline additive GARCH (GARCH-ADD) approach.

We now only consider the estimation of $m_j(\cdot)$ based on all bounded measurable function on compact interval $[a, b]$, where a, b are some fixed constants. When applying to real data, one can use fixed truncation to satisfy this condition. Let $\mathcal{S}_n$ be the space of polynomial splines on $[a, b]$ of degree $p \geq 1$. We introduce a knot sequence with N interior knots

$$u_{-p} = ... = u_{-1} = u_0 = a < u_1 < ... < u_N < b = u_{N+1} = ... = u_{N+p+1},$$

where $N \equiv N_n$ increases when sample size n increases, whose precise order is given in Assumption (A5). The spline of degree p for the jth variable is denoted as $\{b_{j,k}\}_{k=-p}^{N}$ (de Boor (2001)). Then, $\mathcal{S}_n$ consists of functions $g(\cdot)$ satisfying (i) $g(\cdot)$ is a polynomial of degree p on each of the subintervals $I_k = [u_k, u_{k+1})$, $k = 0, ..., N-1$, $I_N = [u_N, b]$; and (ii) for $p \geq 2$, $g(\cdot)$ is $p-1$ time continuously differentiable on $[a, b]$.

Equally-spaced knots are used here for simplicity of proof, while adaptively choosing the locations of the knots could have been done for real data analysis. Let $h = (b - a)/(N+1)$ be the distance between neighboring knots. Define next the space $G = G[a, b]$ of additive splines as the linear space spanned by the following sequence defined by: $\{1, b_{j,k}(y_j), j = 1, ..., J, k = -p, ..., N\}$.

Let $\left(\hat{\lambda}_0', \hat{\lambda}_{1,-p}', ..., \hat{\lambda}_{J,N}'\right)^{\mathrm{T}}$ be the solutions of the least squares problem

$$\left(\hat{\lambda}_0', \hat{\lambda}_{1,-p}', ..., \hat{\lambda}_{J,N}'\right)^{\mathrm{T}} = \underset{R^{1+J(N+p)}}{\operatorname{argmin}} \sum_{t=J+1}^{T} \left\{ Y_t^2 - \lambda_0 - \sum_{j=1}^{J} \sum_{k=-p}^{N} \lambda_{j,k} b_{j,k}\left(Y_{t-j}\right) \right\}^2 .$$

Denote $n = T - J$. Let $\hat{c} = n^{-1} \sum_{t=J+1}^{T} Y_t^2$, which is a $\sqrt{n}$-consistent estimator of c by the Central Limit Theorem. The centered spline estimator of each component function is

$$\hat{m}_j(y) = \sum_{k=-p}^{N} \hat{\lambda}_{j,k} b_{j,k}(y) - \frac{1}{n} \sum_{t=J+1}^{T} \sum_{k=-p}^{N} \hat{\lambda}_{j,k} b_{j,k}\left(Y_{t-j}\right), \quad 1 \le j \le J. \tag{2.2.3}$$

To estimate the parameter β_0, we regress $\{\hat{m}_2(Y_t)\}_{t=J+1}^{T}$ on $\{\hat{m}_1(Y_t)\}_{t=J+1}^{T}$ and solve for the least squares solution of $\sum_{t=J+1}^{T} \{\hat{m}_2(Y_t) - \beta \hat{m}_1(Y_t)\}^2$. The performance is improved by averaging over all the components, so we define the sample least squares criterion,

$$\hat{R}(\beta) = \frac{1}{n} \sum_{t=J+1}^{T} \sum_{j=1}^{J} \left\{ \hat{m}_j(Y_t) - \beta^{j-1} \hat{m}_1(Y_t) \right\}^2, \tag{2.2.4}$$

and the minimizer of (2.2.4) $\hat{\beta}$ is the GARCH-ADD estimator of the dynamic coefficient.

2.2.2 Asymptotic properties of the GARCH-ADD estimators

For our theoretical results, we enforce the following technical assumptions.

(A1) *The data-generating process* $\{Y_t, t > 0\}$ *is strictly stationary and* α*-mixing with exponentially decaying mixing coefficients* $\alpha(k) \le K_0 e^{-\lambda_0 k}$ *for some positive con-*

stants K_0 *and* λ_0. *The* α*-mixing coefficients for* $\{Y_t\}_{t=1}^{T}$ *is defined as*

$$\alpha(k) = \sup_{B\in\sigma\{Y_s, s\leq t\}, C\in\sigma\{Y_s, s\geq t+k\}} |P(B\cap C) - P(B)P(C)|, \ k \geq 1.$$

(A2) *Function* m_1 *is a pth degree continuously differentiable function on interval* $[a,b]$.

(A3) *For any* $t, t' = 1, 2, ..., T, t \neq t'$, *the joint density* $f(y_t, y_{t'})$ *of* $(Y_t, Y_{t'})$, *is continuous and* $0 < c_f \leq \inf_{(y_t,y_{t'})\in[a,b]^2} f(y_t, y_{t'}) \leq \sup_{(y_t,y_{t'})\in[a,b]^2} f(y_t, y_{t'}) \leq C_f < \infty$.

(A4) *The noise* ξ_t *satisfies* $E(\xi_t|\mathcal{F}_{t-1}) = 0, E\left(\xi_t^2|\mathcal{F}_{t-1}\right) = 1$, and $E\left(|\xi_t|^{5+\delta}|\mathcal{F}_{t-1}\right) < M_\delta$ *for some* $\delta > 0$ *and a finite positive* M_δ.

(A5) *The number of interior knots of the spline basis functions with degree* $p > 1$ *satisfies*: $c_N n^{1/(2p)} \log n \leq N \leq C_N n^{1/2}/\log^3 n$, *for some positive constants* c_N *and* C_N.

Remark 1. Assumption (A1) is a standard assumption in time series literature; see Linton and Mammen (2005), Wang and Yang (2007). Assumption (A2) is very relaxed in our paper compared with marginal integration method; see Linton and Nielsen (1995). Assumption (A3) only requires that the pairwise joint density is bounded away from 0 and ∞. So it is a much weaker assumption compared with Assumption (iv) in Carroll *et al.* (2002) and Assumption (c) of Huang and Yang (2004) which require the boundedness of the joint density of the J variables. Assumption (A4) is comparable with Assumption

(vi) in Carroll *et al.* (2002). Assumption (A5) gives the order of the number of interior knots.

We now describe our asymptotic results for the parameter in Theorems 2.1 and 2.2, and the consistency result for the nonparametric news impact curve is given in the Appendix.

Theorem 2.1. *Under Assumptions (A1)-(A5), as* $n \to \infty$, $\hat{\beta} \longrightarrow \beta_0$, *a.s.*

Theorem 2.2. *Under Assumptions (A1)-(A5), as* $n \to \infty$, $\sqrt{n}\left(\hat{\beta} - \beta_0\right)$ *has an asymptotic normal distribution with mean* 0 *and variance* $D^{-2}\sum_t \mathrm{Cov}(V_0, V_t)$, *where* $V_t = \varepsilon_t H(\beta_0, m_1(Y_t))$, *and* $H(\beta_0, m_1(Y_t))$ *is given in (2.7.9) in Appendix, and* $D = \sum_{j=2}^{J}(j-1)^2\beta_0^{2j-4}E\left[m_1^2(Y_t)\right]$.

As an added refinement, considering that the additive components are linked, we define

$$\hat{m}_1^*(y) = \sum_{j=1}^{J}\hat{\beta}^{(j-1)}\hat{m}_j(y) \Bigg/ \sum_{j=1}^{J}\hat{\beta}^{2(j-1)}. \tag{2.2.5}$$

As discussed in Carroll *et al.* (2002), the asymptotic variance of $\{\hat{m}_1^*(y) - m_1(y)\}$ is smaller than that of $\left\{\hat{\beta}^{-(j-1)}\hat{m}_j(y) - m_1(y)\right\}$ for all j. We show, in the Section 2.7, that $\hat{m}_1^*(y)$ has the same convergence rate as $\hat{m}_1(y)$.

2.2.3 The alternatives

There is a host of possible alternative methods for estimating the GARCH models nonparametrically, for example, a referee has suggested that we can improve the efficiency of the estimators by taking the advantage of the structure of model (2.1.1). Define

$\sigma_t^2(\beta, m) = \sum_{j=1}^{J} \beta^{j-1} m\left(Y_{t-j}\right)$, and let β_0 and m_0 be defined as the minimizers of the population least squares (LS) criterion function $E\left\{Y_t^2 - \sigma_t^2(\beta, m)\right\}^2$, or be the minimizers of the negative likelihood (NL) criterion function $E\left[\log\left(\sigma_t^2(\beta, m)\right) + \frac{Y_t^2}{\sigma_t^2(\beta, m)}\right]^2$. Similar to the method in Section 2.2.1, we approximate $m(\cdot)$ by polynomial splines. Thus, the empirical version of the LS or NL problem is $\sum_{t=J+1}^{T} \left\{Y_t^2 - \hat{\sigma}_t^2(\beta, \boldsymbol{\lambda})\right\}^2$ or $\sum_{t=J+1}^{T} \left[\log\left\{\hat{\sigma}_t^2(\beta, \boldsymbol{\lambda})\right\} + \frac{Y_t^2}{\hat{\sigma}_t^2(\beta, \boldsymbol{\lambda})}\right]$, where $\boldsymbol{\lambda} = \{\lambda_{1-p}, ..., \lambda_N\}$ and $\hat{\sigma}_t^2(\beta, \boldsymbol{\lambda})$
$= \sum_{j=1}^{J} \sum_{k=1-p}^{N} \beta^{j-1} \lambda_k b_k(Y_{t-j})$.

The minimizer of β based on the above LS or NL criterion is the estimator of β, denoted by GARCH-LS and GARCH-NL, respectively. We have not investigated their asymptotic properties due to some technical challenges. But the numerical performance of these two estimators have been studied in a comprehensive Monte Carlo study; see Section 2.4.

2.2.4 Selection of knots and lags

An important aspect for regression splines is the choice of the knots. Splines with few knots are generally smoother than splines with many knots; however, increasing the knots usually can improve the fit of the spline function to the data. The number of knots used in our simulation is $N = [c_1 n^{1/(2p)} \log(n)] + c_2$, where $[a]$ denotes the integer part of a, and c_1 and c_2 are positive integers. As pointed out in Wang and Yang (2007), there is no optimal method to select (c_1, c_2). In our simulation, the simple choice $c_1 = c_2 = 1$ works well, so these are set as default values.

For all the above modeling approaches, we need to determine the number of lags J. For the GARCH-ADD approach, we adopt the consistent BIC lag selection method for nonlinear additive autoregressive models (Huang and Yang, 2004) and the BIC is defined as

$$BIC(J) = \log\left[\frac{1}{n}\sum_{t=J+1}^{T}\left\{Y_t^2 - \hat{c} - \sum_{j=1}^{J}\hat{m}_j\left(Y_{t-j}\right)\right\}^2\right] + \frac{\log\log(n)}{n}\left\{1 + J(N+p+1)\right\}.$$

Numerical results of knots and lags selection in a simulation study are reported in Section 2.4.

2.3 Confidence Band for the News Impact Curve

In this section, we introduce a simultaneous confidence band for the news impact curve. For nonlinear additive autoregressive model, Song and Yang (2009b) proposed a two-step spline smoothing method to estimate each additive component: the first step spline smoothing does a quick initial estimation of all additive components and removes all except the ones of interest; the second smoothing is then applied to the cleaned univariate data to refine the estimator of each component with the asymptotically oracle efficiency. They also established an asymptotic $100(1-\alpha)\%$ conservative confidence band

$$\hat{m}_j(y) \pm 2\hat{\sigma}_j(y)\left\{\log(N+1)\right\}^{1/2} Q_N(\alpha), \tag{2.3.1}$$

where $\hat{m}_j$ is the spline-backfitted spline estimator, $\hat{\sigma}_j$ is the estimator of the standard deviation function of $\hat{m}_j$, and $Q_N(\alpha)$ is an inflation factor; see Song and Yang (2009b).

When constructing the confidence band in (2.3.1), one needs additional smoothing steps to estimate the functions $\hat{\sigma}_j$ in (2.3.1), which may cause the results less accurate; see Song and Yang (2009a). In this article we propose a bootstrap version of (2.3.1) similar to Song and Yang (2009a). The following are the detailed procedure of constructing the simultaneous confidence band. Denote a predetermined large integer by n_B. By default n_B is 500.

Step 1. Pre-estimate m_j by its centered pilot estimator $\hat{m}_j$, $j = 1, ..., J$, through an under-smoothed spline smoothing procedure with N_1 knots.

Step 2. Construct the pseudo-response $\hat{W}_t = Y_t^2 - \hat{c} - \sum_{j=2}^{J} \hat{m}_j(Y_{t-j})$ and approximate m_1 by linear spline smoothing with N_2 knots based on $\left\{\hat{W}_t, Y_{t-1}\right\}_{t=J+1}^{T}$. Define the estimator $\breve{m}_1(\cdot) = \arg\min_{g(\cdot)\in\mathcal{S}_n} \sum_{t=J+1}^{T} \left\{\hat{W}_t - g(Y_{t-1})\right\}^2$, and denote residual $\hat{\varepsilon}_t = \hat{W}_t - \breve{m}_1(Y_{t-1})$.

Step 3. Let $\{\delta_{t,b}\}_{J+1\le t\le T}^{1\le b\le n_B}$ be i.i.d. mean 0 and variance 1 samples of the following discrete distribution $\delta_{t,b} = \frac{1\pm\sqrt{5}}{2}$ with probability $\frac{5\pm\sqrt{5}}{10}$.

Step 4. For any $1 \le b \le n_B$, define the b-th wild bootstrap sample $\hat{W}_{t,b}^* = \breve{m}_1(Y_{t-1}) + \delta_{t,b}\hat{\varepsilon}_t$, $J+1 \le t \le T$. Then the bootstrap estimator of $m_1(y)$ is $\breve{m}_1{}^{(b)}(y) = \sum_{k=-1}^{N_2} \hat{\varphi}_k^{(b)} B_k(y)$, where $\left(\hat{\varphi}_{-1}^{(b)}, \hat{\varphi}_2^{(b)}, ..., \hat{\varphi}_{N_2}^{(b)}\right)^{\mathrm{T}}$ are the estimated spline coefficients.

Step 5. Denote by $L_{\alpha/2}(y)$ and $U_{\alpha/2}(y)$ respectively the lower and upper $100(1-\alpha/2)\%$ quantiles of the set $\left\{\breve{m}_1^{(b)}(y)\right\}_{b=1}^{n_B}$. The wild bootstrap $100(1-\alpha)\%$ pointwise confidence interval for function value $m_1(y)$ at one point y is $\left\{L_{\alpha/2}(y), U_{\alpha/2}(y)\right\}$.

Step 6. According to Song and Yang (2009b), when localized at any point y, the uniform confidence band in (2.3.1) is wider than the pointwise confidence interval in Huang (2003) by a common factor $F_\alpha = 2z_{1-\alpha/2}^{-1}\left\{\log(N_2+1)\right\}^{1/2} Q_N(\alpha)$. We define the $(1-\alpha)\%$ bootstrap confidence band for $m_1(y)$ as $\breve{m}_1(y) + \left\{L_{\alpha/2}(y) - \breve{m}_1(y)\right\} F_\alpha, \breve{m}_1(y) + \left\{U_{\alpha/2}(y) - \breve{m}_1(y)\right\} F_\alpha$.

Remark 3. Song and Yang (2009b) proposed to use $N_1 \sim n^{2/5}\log n$ knots for the initial spline estimation in step 1 and $N_2 \sim n^{1/5}$ knots for the backfitting spline in estimation step 2. In our simulation, N_1 and N_2 for the spline estimation are calculated as $N_1 = \min\left\{[c_1 n^{2/5}\log(n)] + c_2, [n/4-1]/J\right\}$ and $N_2 = \left[c_3 n^{1/5}\log(n)\right] + c_4$ and tuning constants $c_1 = 1$, $c_2 = 1$, $c_3 = 0.5$, $c_4 = 1$ by default.

2.4 Simulation

We carried out some simulations to illustrate the finite-sample behavior of the proposed estimators defined in Section 2.2. We compared the performance of the GARCH-ADD, GARCH-LS and GARCH-NL estimators with the GARCH(1,1) and GJR(1,1) estimators.

We generated time series $Y_t = \sigma_t \xi_t$ with the noise sequence $\{\xi_t\}_{t=1}^T$ i.i.d standard normal random variables. The volatility $\{\sigma_t^2\}_{t=1}^T$ was from the following models:

$$
\begin{aligned}
A &: \sigma_t^2 = 0.10 + 0.20 Y_{t-1}^2 + 0.75 \sigma_{t-1}^2, \\
B &: \sigma_t^2 = 0.05 + 0.20 Y_{t-1}^2 + 0.05 Y_{t-1}^2 I\left(Y_{t-1} < 0\right) + 0.75 \sigma_{t-1}^2, \\
C &: \sigma_t^2 = 1 - 0.90 \exp(-2 Y_{t-1}^2) + 0.70 \sigma_{t-1}^2,
\end{aligned}
$$

where the news impact curve in model A is symmetric, and a switching asymmetry has been built into model B. Model C involves exponential curves and a similar model has been studied by Carroll, *et al.* (2002) and Bühlmann and McNeil (2002).

We first considered time series from models A, B and C with $J_{\text{model}} = 5$. For $T = 500$, 1000, 2000 and 3000, we generated 200 replications for the above three processes of size $T + 1000$. Then the first 1000 observations were discarded to make sure the time series behave like strictly stationary. We truncated each time series according to its 2.5th and 97.5th percentile. For these truncated time series, we estimated the parameter β_0 and the news impact curve m_1 by cubic splines. The number of lags, J, was selected according to the BIC described in Section 2.2.4. The minimization of $\hat{R}(\beta)$ was based on a grid search of 100 points around the true value.

(Insert Table 2.1 about here)

The 3rd to the 5th columns in Table 2.1 provide the sample mean (MEAN), standard deviation (STD) and mean squared errors (MSE) of $\hat{\beta}$ based on the GARCH-ADD, GARCH-LS and GARCH-NL methods. As we expected, when the sample size increases,

the parameter β_0 is more accurately estimated, with smaller MSE, confirmative of the conclusions of Theorem 2.1. As one referee expected, the GARCH-LS and GARCH-NL estimators provide more accurate estimation in some cases, especially for Model C. We did not see obvious advantage using these model structures for Models A and B. The mean and median of selected number of lags J_{fit} were reported in the last column of Table 2.1, and one sees that J_{fit} is close to $J_{\text{model}} = 5$ for moderately large sample size. For the news impact curve estimation, in our simulation we tried both $\hat{m}_1$ in (2.2.3) and $\hat{m}_1^*$ in (2.2.5), and the refined $\hat{m}_1^*$ performed slightly better as we expected. The 6th column in Table 2.1 shows the average MSEs (AMSE) in $[-2.0, 2.0]^{J_{\text{fit}}}$ for $\hat{m}_1^*$.

Next, to illustrate the finite-sample behavior of our confidence bands, we calculated the percentage of coverage of the true news impact function by the confidence bands for three different models above. Two nominal confidence levels 0.99 and 0.95 were considered. We carried out 500 replications, and for each replication, 500 bootstrap samples were generated for the bootstrap band. Table 2.2 contains the Monte Carlo coverage probabilities of the proposed bands. One can see the coverage rate gets close to the nominal level for all three models as sample size increases.

(Insert Table 2.2 about here)

We also carried out simulation considering model misspecification, and we generated time series from models A and B with $J_{\text{model}} = \infty$. Remember that for $J_{\text{model}} = \infty$, process A is a GARCH(1,1) process, so clearly GARCH(1,1) is the preferred estimator in this case. For process B, a GJR(1,1) is the desired model. It is thus interesting to see how much efficiency, if any, is lost by using the proposed nonparametric methods with

selected finite number of lags; see results in Table 2.3. For $T = 1000$, the nonparametric methods lost a small amount of efficiency relative to the parametric ones. But that effect decreases as the sample size increases for both processes A and B. Overall, we find that the GARCH-ADD works quite robust though β_0 is not the true parameter anymore. One explanation is that the selected number of lags based on our method is usually also large when $J_{\text{model}} = \infty$.

(Insert Table 2.3 about here)

In all our simulation experiments, our proposed GARCH-ADD method worked very fast, and we provide the time in seconds for all the methods in the last column in Table 2.3. The proposed GARCH-ADD method only needs to solve a moderate number of linear least squares and a simple univariate nonlinear optimization. So in most cases one can see that the GARCH-ADD works much faster compared to its competitors which involve high-dimensional nonlinear optimization.

2.5 Application

In this section, we investigate the news impact curve on BMW daily stock return series to discover the relationship between past return shocks and conditional volatility. We collected the samples of daily percentage returns on the BMW share price from June 1st 1986 to January 30th 1994. There were a total of 2000 observations. We truncated Y_t by its 0.01 and 0.99 quantiles.

For comparison, we also fitted the classical GARCH(1,1) and GJR(1,1) models. We compared the goodness-of-fit of our model with these two models in terms of volatility prediction error $\frac{1}{n}\sum_{t=J+1}^{T}\left(\hat{\sigma}_t^2 - Y_t^2\right)^2$ and the log-likelihood $-\sum_{t=J+1}^{T}\log\left\{\hat{\sigma}_t^{-1}\varphi\left(\frac{Y_t}{\hat{\sigma}_t}\right)\right\}$ with $J_{\text{fit}} = 50$. Clearly, the semiparametric method had an edge over the two parametric models in terms of prediction error and log-likelihood. One can see from Table 2.4 that the leverage effects of the GJR model can be further enhanced by a nonlinear link to yield a much better volatility fit.

(Insert Table 2.4 about here)

To examine the validity of the GARCH and GJR models, we constructed the spline bootstrap confidence band. Figure 2.2 plots the GARCH, GJR, GARCH-ADD fits with the 95% confidence band. From Figure 2.2, we find that the spline estimated news impact curve stands obvious contrast to the GARCH(1,1) fit, which shows strong evidence of the asymmetry of the news impact curve. But it seems that all three models can be fully covered by the bootstrap band.

(Insert Figure 2.1 about here)

For diagnostic purpose, we show the estimated autocorrelation function (ACF) of the daily standardized residuals $\hat{\varepsilon}^2$ with the 95% Bartlett intervals, and one sees that the autocorrelation in the daily returns series is very small.

2.6 Discussion

Non/semi-parametric methods enhance the flexibility of the volatility models that practitioners use. However, due to the limitations in either interpretability, computational complexity or theoretical reliability, most of the nonparametric stochastic volatility models have not been widely used as general tools in volatility analysis. In this chapter, we have advanced semiparametric methods as flexible, computationally efficient and theoretically attractive tools for studying the financial volatility.

We propose approximating the functional component in an additive volatility model by B-splines, which can be done by running OLS operations once the spline basis is chosen. Thus our method is particularly computationally efficient compared to its competitors which have to solve big system equations or optimize high-dimensional nonlinear functions. In addition, we introduced two alternative methods taking into account the model structure. These alternative methods are supposed to be more efficient in principle, but obtaining the asymptotics is likely to be difficult. We leave it as future research work. All the proposed estimators are easily implemented in commonly used software/package such as `lm()` in R.

There is more future work ahead. For example, it is interesting to consider the issue of model misspecification. In this paper, instead of estimating the true dynamic coefficient for $J = \infty$, we estimate a parameter β_0 that approximates the true parameter by using some finite J. If $J = \infty$, β_0 would not be the true dynamic coefficient anymore. The asymptotic results for the misspecified case has to be more fully explored.

2.7 Appendix

Throughout this section, we denote c, C any positive constants, without distinction. Denote $\|\phi\|_2$ the theoretical L^2 norm of a function ϕ on $[a,b]$, $\|\phi\|_2^2 = \int_a^b \phi^2(y) f(y)\, dy$, and define the empirical L^2 norm as $\|\phi\|_{2,n}^2 = n^{-1}\sum_{i=1}^n \phi^2(Y_i)$. The corresponding inner products are defined by $\langle \phi, \varphi \rangle_2 = \int_a^b \phi(y)\varphi(y) f(y)\, dy$ and $\langle \phi, \varphi \rangle_{2,n} = n^{-1}\sum_{i=1}^n \phi(Y_i)\varphi(Y_i)$.

Define the centered version spline basis

$$b_{j,k}^*(y) = b_{j,k}(y) - \frac{E(b_{j,k})}{E(b_{j,k-1})} b_{j,k-1}(y), \quad j = 1, ..., J, \quad k = 1-p, ..., N,$$

with the standardized version given for any $j = 1, ..., J$, $k = 1-p, ..., N$,

$$B_{j,k}(y) = b_{j,k}^*(y) / \left\| b_{j,k}^* \right\|_2. \tag{2.7.1}$$

In practice, basis $\{b_{j,k},\ j = 1, ..., J,\ k = -p, ..., N\}^{\mathrm{T}}$ is used for data analysis, and the mathematically equivalent expression (2.7.1) is convenient for asymptotic analysis. Let $\mathbf{x} = (x_1, ..., x_J)^{\mathrm{T}}$. For a J-dimensional vector $\mathbf{X}_t = (Y_{t-1}, ..., Y_{t-J})^{\mathrm{T}}$, define

$$\mathbf{B}(\mathbf{x}) = \{1, B_{1-p,1}(x_1), ..., B_{J,N}(x_J)\}^{\mathrm{T}}, \ \boldsymbol{B} = \{\mathbf{B}(\mathbf{X}_{J+1}), ..., \mathbf{B}(\mathbf{X}_T)\}^{\mathrm{T}}.$$

Let $m_t = c + \sum_{j=1}^{J} \beta_0^{j-1} m_1(Y_{t-j})$. Define the signal vector $\mathbf{m} = \{m_{J+1}, ..., m_T\}^{\mathrm{T}}$ and the noise vector $\boldsymbol{\epsilon} = \{\epsilon_{J+1}, ..., \epsilon_T\}^{\mathrm{T}}$. Let

$$\boldsymbol{\Lambda}_j = \mathrm{diag}\{0, ..., 0, \underbrace{1, ..., 1}_{\text{from } (N+p)(j-1)+2 \text{ to } (N+p)j+1}, 0, ..., 0\}$$

be a diagonal matrix. Based on the relation $Y_t^2 = m_t + \epsilon_t$, one defines the signal spline smoothers and the noise spline components

$$\begin{aligned} \tilde{m}_j(y) &= \mathbf{B}(y)^{\mathrm{T}} \boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\mathbf{m} - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\mathbf{m}, \\ \tilde{\epsilon}_j(y) &= \mathbf{B}(y)^{\mathrm{T}} \boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\epsilon} - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\boldsymbol{\epsilon}, \end{aligned} \tag{2.7.2}$$

where $\mathbf{1}_n$ is a length n dimensional vector with all elements 1.

Defining $\mathbf{Z} = \left\{Y_{J+1}^2, ..., Y_T^2\right\}$, we can rewrite $\hat{m}_j(y)$ in (2.2.3) using matrix as

$$\hat{m}_j(y) = \mathbf{B}(y)^{\mathrm{T}} \boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\mathbf{Z} - \frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\boldsymbol{\Lambda}_j (\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B})^{-1} \boldsymbol{B}^{\mathrm{T}}\mathbf{Z}.$$

Then one has the following crucial decomposition for proving Theorem 2.1,

$$\hat{m}_j(y) = \tilde{m}_j(y) + \tilde{\epsilon}_j(y), \quad j = 1, ..., J. \tag{2.7.3}$$

To prove Theorems 2.1 and 2.2, we need the following lemma on the L^2 convergence rate of the one-step spline estimator $\hat{m}_1$ to m_1.

Lemma 2.1. *Under Assumptions (A1)-(A5), as* $n \to \infty$,

$$\|\hat{m}_1 - m_1\|_{2,n} + \|\hat{m}_1 - m_1\|_2 = O_{a.s.}\left(h^p + \log n/\sqrt{nh}\right). \tag{2.7.4}$$

Proof. Using the approximation result of polynomial spline on page 149 of de Boor (2001), we have $\|\tilde{m}_1 - m_1\|_2 = O\left(h^p\right)$, and $\|\tilde{m}_1 - m_1\|_{2,n} = O\left(h^p\right)$. According to Lemma A.6 of Wang and Yang (2007), $\|\tilde{\epsilon}\|_2 = O_{a.s.}\left(\log n/\sqrt{nh}\right)$, and $\|\tilde{\epsilon}\|_{2,n} = O_{a.s.}\left(\log n/\sqrt{nh}\right)$. The result in Lemma 2.1 follows from the decomposition in (2.7.3). ∎

The following corollary states the asymptotic property of $\hat{m}_1^*$ given in (2.2.5) to m_1.

Corollary 2.1. *Under Assumptions (A1)-(A5), as* $n \to \infty$,

$$\|\hat{m}_1^* - m_1\|_{2,n} + \|\hat{m}_1^* - m_1\|_2 = O_P\left(h^p + \log n/\sqrt{nh}\right).$$

Proof. The proof is quite straightforward from the above lemma and Theorem 2.2.

$$\|\hat{m}_1^* - m_1\|_2 = \left\| \frac{1}{\sum_{j=1}^J \hat{\beta}^{2(j-1)}} \left[\sum_{j=1}^J \hat{\beta}^{(j-1)} \left(\hat{m}_j - m_j\right) + \sum_{j=1}^J \hat{\beta}^{j-1} \left(\beta_0^{j-1} - \hat{\beta}^{j-1}\right) m_1 \right] \right\|_2 .$$

For each j, $\|\hat{m}_j - m_j\|_2$ has order $O_P\left(h^p + \log n/\sqrt{nh}\right)$. Combining with the result that $\sum_{j=1}^J \hat{\beta}^{j-1}\left(\beta_0^{j-1} - \hat{\beta}^{j-1}\right) = O_P(1/\sqrt{n}) = o_P\left(h^p + \log n/\sqrt{nh}\right)$ from Theorem 2.2, we can easily obtain that $\|\hat{m}_1^* - m_1\|_2$ is with the order $O_P\left(h^p + \frac{\log n}{\sqrt{nh}}\right)$, so is $\|\hat{m}_1^* - m_1\|_{2,n}$. ∎

2.7.1 Proof of Theorem 2.1

Note that the risk function $R(\beta)$ given in (2.2.2) is locally convex on β and hence, consistency for β can be implied by $\sup_{\beta\in[\beta_1,\beta_2]}\left|\hat{R}(\beta)-R(\beta)\right|\to 0$ *a.s.*, where $\hat{R}(\beta)$ is given in (2.2.4). Note that

$$\begin{aligned}\hat{R}(\beta) &= \sum_{j=1}^{J}\left\|\hat{m}_j - m_j + \beta^{j-1}m_1 - \beta^{j-1}\hat{m}_1\right\|_{2,n}^2 + \sum_{j=1}^{J}\left\|m_j-\beta^{j-1}m_1\right\|_{2,n}^2 \\ &\quad + \sum_{j=1}^{J} 2\left\langle m_j-\beta^{j-1}m_1, \hat{m}_j - m_j + \beta^{j-1}m_1 - \beta^{j-1}\hat{m}_1\right\rangle_{2,n} = P_1(\beta)+P_2(\beta)+P_3(\beta).\end{aligned}$$

By (2.7.4), we have $\sup_{\beta\in[\beta_1,\beta_2]} P_1(\beta) = O_{a.s.}(h^{2p}+\log^2 n/nh)$, and

$$\sup_{\beta\in[\beta_1,\beta_2]} P_3(\beta) \le 2J \max_{1\le j\le J}\left\{\left\|\hat{m}_j - m_j + \beta^{j-1}m_1 - \beta^{j-1}\hat{m}_1\right\|_{2,n} \sup_{x\in[a,b]}\left|m_j(x)-\beta^{j-1}m_1(x)\right|\right\},$$

which is of the order $O_{a.s.}\left(h^p+\log n/\sqrt{nh}\right)$. Thus,

$$\sup_{\beta\in[\beta_1,\beta_2]}\left|\hat{R}(\beta)-P_2(\beta)\right| = O_{a.s.}\left(h^p+\log n/\sqrt{nh}\right).$$

While

$$\begin{aligned}
&\sup_{\beta\in[\beta_1,\beta_2]}|P_2(\beta)-R(\beta)| \\
\leq\ & \left|\frac{1}{n}\sum_{t=J+1}^{T}\left\{\sum_{j=1}^{J}m_j^2(Y_t)\right\}-E\left\{\sum_{j=1}^{J}m_j^2(Y_t)\right\}\right|+\frac{1-\beta_2^{2J}}{1-\beta_2^2}\left|\frac{1}{n}\sum_{t=J+1}^{T}m_1^2(Y_t)-Em_1^2(Y_t)\right| \\
&+\frac{2(1-\beta_2^{2J})}{1-\beta_2^2}\left|\frac{1}{n}\sum_{t=J+1}^{T}\left\{\sum_{j=1}^{J}m_j(Y_t)m_1(Y_t)\right\}-E\left\{\sum_{j=1}^{J}m_j(Y_t)m_1(Y_t)\right\}\right|.
\end{aligned}$$

By a strong law of large numbers for mixing processes, $\sup_{\beta\in[\beta_1,\beta_2]}|P_2(\beta)-R(\beta)| = o_{a.s.}(1)$. Thus

$$\sup_{\beta\in[\beta_1,\beta_2]}\left|\hat{R}(\beta)-R(\beta)\right|\leq\sup_{\beta\in[\beta_1,\beta_2]}\left|\hat{R}(\beta)-P_2(\beta)\right|+\sup_{\beta\in[\beta_1,\beta_2]}|P_2(\beta)-R(\beta)|=o_{a.s.}(1),$$

and $\hat{\beta}$ converges to β_0 $a.s.$ is followed.

2.7.2 Proof of Theorem 2.2

We make a Taylor expansion about $\frac{d}{d\beta}\hat{R}(\beta)$ at β_0,

$$\sqrt{n}\frac{d}{d\beta}\hat{R}\left(\hat{\beta}\right)=\sqrt{n}\left.\frac{d}{d\beta}\hat{R}(\beta)\right|_{\beta=\beta_0}+\left.\frac{d^2}{d\beta^2}\hat{R}(\beta)\right|_{\beta=\tilde{\beta}}\sqrt{n}\left(\hat{\beta}-\beta_0\right),$$

where $\tilde{\beta}$ is between $\hat{\beta}$ and β_0. Thus one has

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}-\beta_0\right) &= \sqrt{n}\left\{\frac{d^2}{d\beta^2}\hat{R}\left(\beta\right)\Big|_{\beta=\tilde{\beta}}\right\}^{-1}\left\{\frac{d}{d\beta}\hat{R}\left(\hat{\beta}\right)-\frac{d}{d\beta}\hat{R}\left(\beta\right)\Big|_{\beta=\beta_0}\right\} \\ &= -\sqrt{n}\left\{\frac{d^2}{d\beta^2}\hat{R}\left(\beta\right)\Big|_{\beta=\tilde{\beta}}\right\}^{-1}\frac{d}{d\beta}\hat{R}\left(\beta\right)\Big|_{\beta=\beta_0}. \qquad (2.7.5)\end{aligned}$$

We need the following two lemmas to deal with $-\sqrt{n}\,\frac{d}{d\beta}\hat{R}\left(\beta\right)\Big|_{\beta=\beta_0}$ and $\frac{d^2}{d\beta^2}\hat{R}\left(\beta\right)$ respectively.

Lemma 2.2. *Under Assumptions (A1)-(A5),*

$$-\frac{\sqrt{n}}{2}\,\frac{d}{d\beta}\hat{R}\left(\beta\right)\Big|_{\beta=\beta_0} = n^{-1/2}\sum_{t=J+1}^{T}\epsilon_t H(\beta_0, m_1(Y_t)), \qquad (2.7.6)$$

where $H(\beta_0, m_1(Y_t))$ is given in (2.7.9).

Proof. Note that

$$\begin{aligned}&-\frac{\sqrt{n}}{2}\,\frac{d}{d\beta}\hat{R}\left(\beta\right)\Big|_{\beta=\beta_0} = n^{-1/2}\sum_{t=J+1}^{T}\sum_{j=1}^{J}\left(j-1\right)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}\hat{m}_1\left(Y_t\right) \\ =\;& n^{-1/2}\sum_{t=J+1}^{T}\sum_{j=1}^{J}\left(j-1\right)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}\left[\hat{m}_1\left(Y_t\right)-m_1\left(Y_t\right)\right] \\ &+n^{-1/2}\sum_{t=J+1}^{T}\sum_{j=1}^{J}\left(j-1\right)\beta_0^{j-2}\left\{\hat{m}_j(Y_t)-\beta_0^{j-1}\hat{m}_1(Y_t)\right\}m_1\left(Y_t\right) \\ =\;& I+II,\end{aligned}$$

and the first term I can be written as

$$\begin{aligned} I &= n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1)\, \beta_0^{j-2} \left\{ \hat{m}_j(Y_t) - m_j(Y_t) \right\} \left\{ \hat{m}_1(Y_t) - m_1(Y_t) \right\} \\ &\quad -n^{-1/2} \sum_{t=J+1}^{T} \sum_{j=1}^{J} (j-1)\, \beta_0^{2j-3} \left\{ \hat{m}_1(Y_t) - m_1(Y_t) \right\}^2 \\ &= I_1 + I_2. \end{aligned}$$

As in the proof of Theorem 2.1, we have both I_1 and I_2 with order $O_{a.s.}(h^{2p} + \frac{\log^2 n}{nh})$. With the order of h in Assumption (A5), we have $I = O_{a.s.}\left\{ n^{1/2} \left(h^{2p} + \frac{\log^2 n}{nh} \right) \right\} = o_{a.s.}(1)$. For part II, noting that $\|\tilde{m}_1 - m_1\|_{2,n} = O(h^p),\ a.s.$, so by (2.7.3) and Assumption (A5) we have

$$\begin{aligned} II &= n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} (j-1)\, \beta_0^{j-2} \left\{ \tilde{m}_j(Y_{t'}) + \tilde{\epsilon}_j(Y_{t'}) - \beta_0^{j-1} \tilde{m}_1(Y_{t'}) - \beta_0^{j-1} \tilde{\epsilon}_1(Y_{t'}) \right\} m_1(Y_{t'}) \\ &= n^{-1/2} \sum_{t'=J+1}^{T} \sum_{j=1}^{J} (j-1)\, \beta_0^{j-2} \left\{ \tilde{\epsilon}_j(Y_{t'}) - \beta_0^{j-1} \tilde{\epsilon}_1(Y_{t'}) \right\} m_1(Y_{t'}) + o_{a.s.}(1). \end{aligned}$$

Let

$$\widehat{\boldsymbol{V}} = \frac{1}{n} \boldsymbol{B}^{\mathrm{T}} \boldsymbol{B} = \begin{pmatrix} 1 & 0 \\ 0 & \langle B_{j,k}, B_{j',k'} \rangle_{2,n} \end{pmatrix}_{\substack{1 \le j,j' \le J, \\ 1-p \le k,k' \le N}}, \quad \boldsymbol{V} = \begin{pmatrix} 1 & 0 \\ 0 & \langle B_{j,k}, B_{j',k'} \rangle_{2} \end{pmatrix}_{\substack{1 \le j,j' \le J, \\ 1-p \le k,k' \le N}}.$$

With $\tilde{\epsilon}_j$ defined in (2.7.2), the main term in II is

$$n^{-1/2}\sum_{t'=J+1}^{T}\sum_{j=1}^{J}\left[(j-1)\,\beta_0^{j-2}m_1\left(Y_{t'}\right)\left[\left\{\mathbf{B}\left(Y_{t'}\right)\mathbf{\Lambda}_j-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_j\right\}\right.\right.$$
$$\left.\left.-\beta_0^{j-1}\left\{\mathbf{B}\left(Y_{t'}\right)\mathbf{\Lambda}_1-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_1\right\}\right]\widehat{\boldsymbol{V}}^{-1}\left\{\frac{1}{n}\sum_{t=J+1}^{T}B_{j,k}(Y_t)\epsilon_t\right\}_{j,k}\right] \qquad (2.7.7)$$

According to Lemma A.10 in Wang and Yang (2007), we can replace $\widehat{\boldsymbol{V}}$ by $\boldsymbol{V}$ in equation (2.7.7) with a negligible term $O_{a.s.}\left\{n^{-1/2}N(\log n)^2\right\}$. Next we interchange the indices t and t' of (2.7.7), thus the main term in II can be approximated by

$$n^{-1/2}\sum_{t=J+1}^{T}\sum_{j=1}^{J}\left[(j-1)\,\beta_0^{j-2}\epsilon_t\left[\left\{\mathbf{B}\left(Y_t\right)\mathbf{\Lambda}_j-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_j\right\}\right.\right.$$
$$\left.\left.-\beta_0^{j-1}\left\{\mathbf{B}\left(Y_t\right)\mathbf{\Lambda}_1-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_1\right\}\right]\boldsymbol{V}^{-1}\left\{\frac{1}{n}\sum_{t'=J+1}^{T}B_{j,k}(Y_{t'})m_1\left(Y_{t'}\right)\right\}_{j,k}\right]. \qquad (2.7.8)$$

Denote

$$H(\beta_0,m_1(Y_t))=\sum_{j=1}^{J}\left[(j-1)\,\beta_0^{j-2}\left[\left\{\mathbf{B}\left(Y_t\right)\mathbf{\Lambda}_j-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_j\right\}\right.\right.$$
$$\left.\left.-\beta_0^{j-1}\left\{\mathbf{B}\left(Y_t\right)\mathbf{\Lambda}_1-\frac{1}{n}\mathbf{1}_n^{\mathrm{T}}\boldsymbol{B}\mathbf{\Lambda}_1\right\}\right]\boldsymbol{V}^{-1}\left\{\frac{1}{n}\sum_{t'=J+1}^{T}B_{j,k}(Y_{t'})m_1\left(Y_{t'}\right)\right\}_{j,k}\right]. \qquad (2.7.9)$$

and equation (2.7.8) can be written as $n^{-1/2}\sum_{t=J+1}^{T}\epsilon_t H(\beta_0,m_1(Y_t))$, which leads to (2.7.6). ∎

Lemma 2.3. *Under (A1)-(A5),* $\frac{d^2}{d\beta^2}\hat{R}(\beta) = Em_1^2(Y_t)\sum_{j=2}^{J}\left\{(j-1)\beta^{(j-2)}\right\}^2 + o_{a.s.}(1)$.

Proof. Note that

$$\begin{aligned}\frac{d^2}{d\beta^2}\hat{R}(\beta) &= n^{-1}\sum_{t=J+1}^{T}\sum_{j=2}^{J}\hat{m}_1^2(Y_t)\left\{(j-1)\beta^{(j-2)}\right\}^2 \\ &\quad +n^{-1}\sum_{t=J+1}^{T}\sum_{j=2}^{J}(j-1)(j-2)\beta^{j-3}\hat{m}_1(Y_t)\left\{\hat{m}_j(Y_t)-\beta^{j-1}\hat{m}_1(Y_t)\right\} \\ &= I_1 + I_2,\end{aligned}$$

where $I_2 = o_{a.s.}(1)$ similarly as for I in Lemma (2.2), and for I_1,

$$\begin{aligned} & n^{-1}\sum_{t=J+1}^{T}\hat{m}_1^2(Y_t) - n^{-1}\sum_{t=J+1}^{T}m_1^2(Y_t) = n^{-1}\sum_{t=J+1}^{T}\left\{\hat{m}_1(Y_t)-m_1(Y_t)\right\}\left\{\hat{m}_1(Y_t)+m_1(Y_t)\right\} \\ \leq & \left\{n^{-1}\sum_{t=J+1}^{T}\left(\hat{m}_1(Y_t)-m_1(Y_t)\right)^2\right\}^{1/2}\left\{n^{-1}\sum_{t=J+1}^{T}\left(\hat{m}_1(Y_t)+m_1(Y_t)\right)^2\right\}^{1/2} \\ \leq & \left\|\hat{m}_1(x)-m_1(x)\right\|_{2,n}\sup_x\left|\sqrt{6}m_1(x)\right| = O_{a.s.}\left(h^p+\frac{\log n}{\sqrt{nh}}\right) = o_{a.s.}(1). \qquad (2.7.10)\end{aligned}$$

By a law of large numbers, we have $n^{-1}\sum_{t=J+1}^{T}m_1^2(Y_t) \to E\left[m_1^2(Y_t)\right]$, as n goes to infinity. Combining with (2.7.10), we have $\lim_{n\to\infty} n^{-1}\sum_{t=J+1}^{T}\hat{m}_1^2(Y_t) = E\left[m_1^2(Y_t)\right]$. ∎

Now, we continue the proof of Theorem 2.2. Combining (2.7.5), (2.7.6) and Lemma 2.3, and noting that as $n\to\infty$, $\sum_{j=2}^{J}(j-1)^2\tilde{\beta}^{2j-4} \to \sum_{j=2}^{J}(j-1)^2\beta_0^{2j-4}$, *a.s.*, we have,

$$n^{1/2}\left(\hat{\beta}-\beta_0\right) = n^{-1/2}\sum_t \epsilon_t H(\beta_0, m_1(Y_t))\left\{\sum_{j=2}^{J}(j-1)^2\beta_0^{2j-4}E\left[m_1^2(Y_t)\right]\right\}^{-1} + o_{a.s.}(1).$$

Asymptotic normality of $n^{1/2}\left(\hat{\beta}-\beta_0\right)$ follows from a Slutsky theorem and a central limit theorem for strongly mixing sequences (see, e.g., Bosq (1996), Theorem 1.7). We have to verify that for some $\nu > 2$, $E\left|\epsilon_t H(\beta_0, m_1(Y_t))\right|^{\upsilon} < \infty$, which can be obtained with our Assumptions (A2) and (A4).

2.8 References

Bosq, D. (1996) Nonparametric Statistics for Stochastic Processes. New York: Springer-Verlag.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *computational statistics & data analysis* **40**, 665-683.

Carroll, R., Hardle, W. and Mammen, E. (2002). Estimation in an additive model when the components are linked parametrically. *Econometric Theory* **18**, 886-912.

de Boor, C. (2001) A Practical Guide to Splines. New York: Springer-Verlag.

Engle, R. F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

Engle, R. F. and Ng, V. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* **48**, 1749-1778.

Glosten, L. R., Jaganathan, R. and Runkle, D. E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779-1801.

Härdle, W. and Tsybakov, A. B. (1997) Locally polynomial estimators of the volatility Function. *Journal of Econometrics* **81**, 223-242.

Härdle, W., Tsybakov, A. B. and Yang, L. (1998) Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* **68**, 221-245.

Hengartner, N. W. and Sperlich, S. (2005) Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis* **95**, 246-272.

Huang, J. Z. (2003) Local asymptotics for polynomial spline regression. *Annals of Statistics* **31**, 1600-1635.

Huang, J. Z. and Yang, L. (2004) Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society: Series B* **66**, 463-477.

Hafner, C. (1998) Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility. Heidelberg: Physica-Verlag.

Hastie, T. J. and Tibshirani, R. J. (1990) Generalized Additive Models. London: Chapman and Hall.

Linton, O. B. and Nielsen, J. P. (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.

Linton, O. B. and Mammen, E. (2005) Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* **73**, 771-836.

Masry, E. and Tjøstheim, D. (1995) Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econometric Theory* **11**, 258-289.

Opsomer, J. D. and Ruppert, D. (1997) Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics* **25**, 186-211.

Pagan, A. R. and Schwert, G. W. (1990) Alternative models for conditional stock volatility. *Journal of Econometrics* **45**, 267–290.

Song, Q. and Yang, L. (2009a) Spline confidence bands for variance function. *Journal of Nonparametric Statistics* **21**, 589-609.

Song, Q. and Yang, L. (2009b) Simultaneous confidence band for nonlinear additive autoregression model via spline-backfitted spline smoothing. *Journal of Multivariate analysis*, in press.

Stone, C. J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* **13**, 689-705.

Wang, L. and Yang, L. (2007) Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* **35**, 2474-2503.

Yang, L. (2002) Direct estimation in an additive model when the components are proportional. *Statistica Sinica* **12**, 801-821.

Yang, L. (2006) Semiparametric GARCH model and foreign exchange volatility. *Journal of Econometrics* **130**, 365-384.

Yang, L., Härdle, W. and Nielsen, J. P. (1999) Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* **20**, 597-604.

Table 2.1: Monto Carlo performance results based on 200 replications ($J_{\text{model}} = 5$). The values outside and inside the parentheses are the results based on the fitted J_{fit} and the oracle $J_{\text{oracle}} = 5$.

Size	Estimator	Parametric component			Nonparametric component	J_{fit} mean(median)
		MEAN	STD	MSE	AMSE	
A						
500	GARCH-ADD	0.68(0.63)	0.17(0.14)	0.035(0.033)	0.026(0.025)	3.7(3)
	GARCH-LS	0.81(0.78)	0.15(0.15)	0.027(0.024)	0.034(0.031)	5.8(5)
	GARCH-NL	0.80(0.79)	0.16(0.15)	0.027(0.024)	0.033(0.032)	5.8(5)
1000	GARCH-ADD	0.74(0.70)	0.14(0.11)	0.019(0.017)	0.016(0.017)	3.9(4)
	GARCH-LS	0.80(0.80)	0.11(0.10)	0.014(0.013)	0.025(0.024)	6.2(6)
	GARCH-NL	0.80(0.82)	0.11(0.10)	0.015(0.016)	0.025(0.026)	6.1(5)
2000	GARCH-ADD	0.75(0.73)	0.10(0.08)	0.009(0.008)	0.011(0.012)	4.1(4)
	GARCH-LS	0.78(0.80)	0.08(0.08)	0.008(0.008)	0.017(0.017)	6.2(5)
	GARCH-NL	0.79(0.81)	0.08(0.08)	0.008(0.010)	0.018(0.019)	6.2(5)
3000	GARCH-ADD	0.77(0.75)	0.07(0.07)	0.005(0.004)	0.011(0.011)	4.6(5)
	GARCH-LS	0.78(0.80)	0.08(0.07)	0.007(0.007)	0.015(0.015)	6.3(5)
	GARCH-NL	0.79(0.81)	0.08(0.06)	0.008(0.008)	0.015(0.016)	6.2(5)
B						
500	GARCH-ADD	0.73(0.71)	0.15(0.14)	0.022(0.022)	0.413(0.348)	5.1(5)
	GARCH-LS	0.81(0.80)	0.13(0.13)	0.021(0.019)	0.173(0.170)	6.0(5)
	GARCH-NL	0.81(0.82)	0.12(0.11)	0.019(0.016)	0.175(0.145)	5.9(5)
1000	GARCH-ADD	0.77(0.77)	0.10(0.10)	0.011(0.010)	0.177(0.163)	5.5(5)
	GARCH-LS	0.81(0.81)	0.10(0.09)	0.013(0.012)	0.100(0.108)	6.1(5)
	GARCH-NL	0.80(0.83)	0.09(0.07)	0.011(0.011)	0.095(0.089)	6.1(5)
2000	GARCH-ADD	0.77(0.78)	0.08(0.07)	0.006(0.006)	0.118(0.116)	5.4(5)
	GARCH-LS	0.79(0.81)	0.08(0.07)	0.008(0.008)	0.079(0.088)	6.4(5)
	GARCH-NL	0.79(0.82)	0.07(0.05)	0.007(0.008)	0.074(0.076)	6.3(5)
3000	GARCH-ADD	0.78(0.79)	0.06(0.06)	0.005(0.005)	0.093(0.098)	5.6(5)
	GARCH-LS	0.79(0.81)	0.07(0.06)	0.007(0.007)	0.075(0.082)	6.2(5)
	GARCH-NL	0.79(0.82)	0.06(0.04)	0.006(0.007)	0.069(0.072)	6.3(5)
C						
500	GARCH-ADD	0.57(0.55)	0.16(0.12)	0.042(0.036)	0.179(0.079)	2.7(2)
	GARCH-LS	0.79(0.72)	0.18(0.18)	0.040(0.031)	0.106(0.085)	5.3(5)
	GARCH-NL	0.76(0.71)	0.19(0.19)	0.038(0.034)	0.102(0.083)	5.3(5)
1000	GARCH-ADD	0.60(0.58)	0.16(0.11)	0.036(0.029)	0.147(0.055)	2.6(2)
	GARCH-LS	0.75(0.71)	0.15(0.13)	0.026(0.018)	0.075(0.057)	5.7(5)
	GARCH-NL	0.73(0.71)	0.15(0.14)	0.024(0.020)	0.069(0.057)	5.6(5)
2000	GARCH-ADD	0.64(0.61)	0.13(0.09)	0.021(0.016)	0.107(0.033)	2.7(2)
	GARCH-LS	0.73(0.71)	0.10(0.04)	0.011(0.010)	0.044(0.038)	5.8(5)
	GARCH-NL	0.72(0.71)	0.11(0.05)	0.013(0.012)	0.046(0.040)	5.8(5)
3000	GARCH-ADD	0.67(0.64)	0.11(0.07)	0.013(0.009)	0.069(0.020)	2.9(3)
	GARCH-LS	0.71(0.70)	0.09(0.08)	0.007(0.006)	0.030(0.027)	6.2(5)
	GARCH-NL	0.71(0.71)	0.09(0.08)	0.008(0.007)	0.033(0.030)	6.2(5)

Table 2.2: Coverage probabilities from 500 replications.

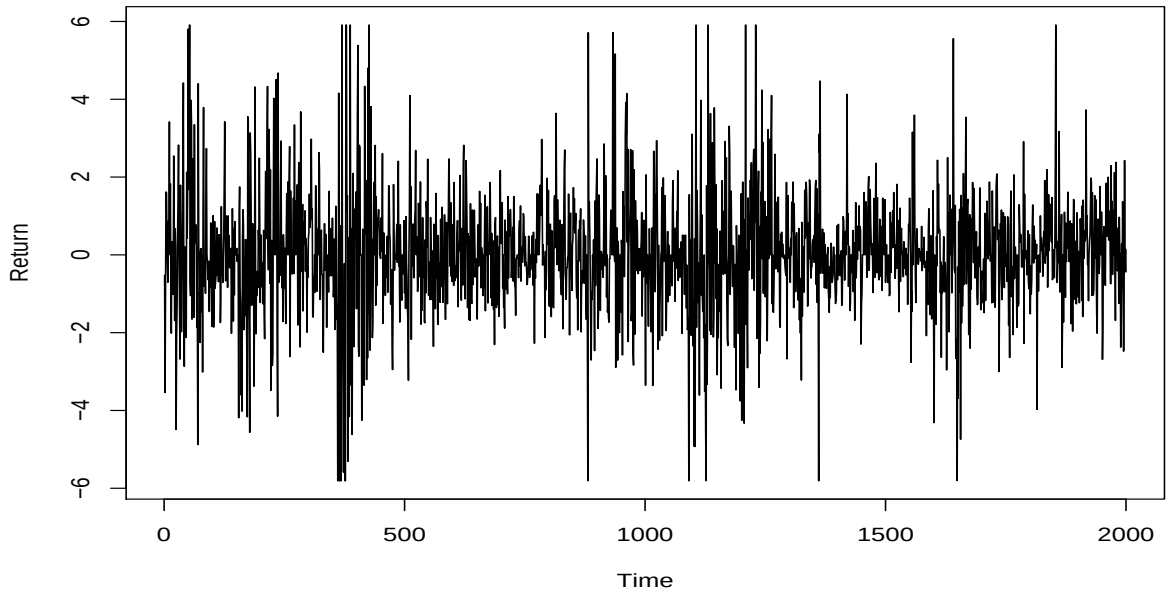
Sample Size	Model A		Model B		Model C	
	95%	99%	95%	99%	95%	99%
500	0.982	0.998	0.924	0.982	0.904	0.982
1000	0.970	0.996	0.962	0.990	0.848	0.950
2000	0.982	0.996	0.956	0.984	0.868	0.934
3000	0.976	0.996	0.932	0.986	0.892	0.954

Table 2.3: Monto Carlo performance results based on 200 replications ($J_{\text{model}} = \infty$).

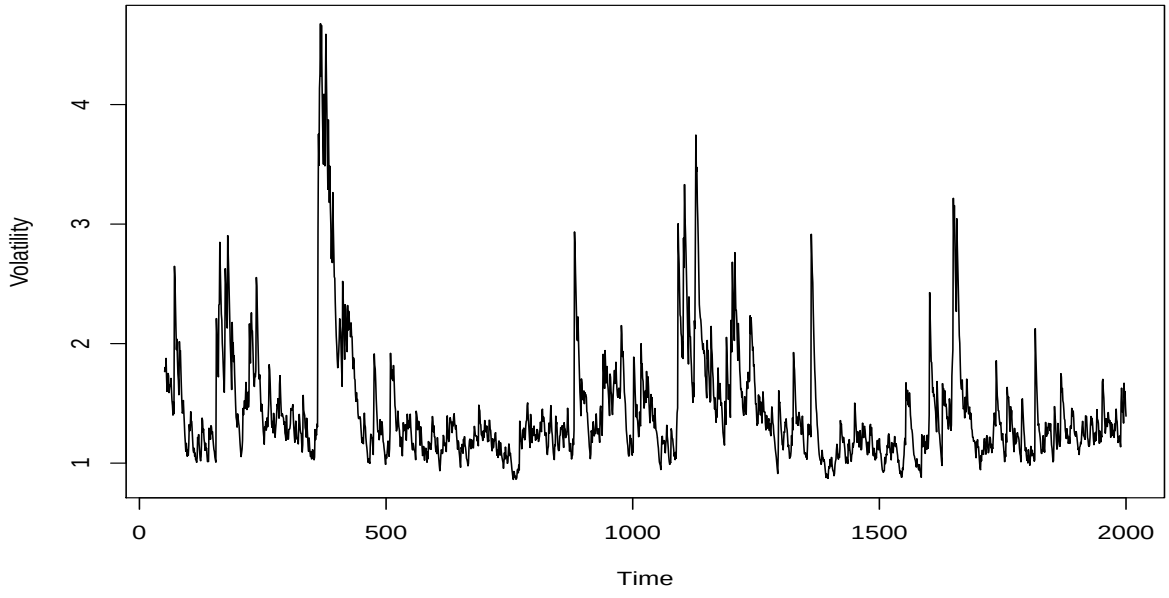
	Size	Estimator	Parametric component			Nonparametric component	Time (secs)
			MEAN	STD	MSE	AMSE	
A	1000	GARCH(1,1)	0.77	0.04	0.002	0.008	1.3
		GJR(1,1)	0.77	0.04	0.002	0.008	7.4
		GARCH-ADD	0.71	0.10	0.012	0.016	2.6
		GARCH-LS	0.81	0.06	0.008	0.030	11.1
		GARCH-NL	0.82	0.06	0.008	0.031	11.2
	2000	GARCH(1,1)	0.77	0.03	0.001	0.006	6.2
		GJR(1,1)	0.77	0.03	0.001	0.006	12.9
		GARCH-ADD	0.75	0.06	0.004	0.009	5.6
		GARCH-LS	0.80	0.04	0.004	0.023	23.2
		GARCH-NL	0.81	0.04	0.005	0.025	22.6
	3000	GARCH(1,1)	0.77	0.02	0.001	0.005	11.8
		GJR	0.77	0.02	0.001	0.006	19.0
		GARCH-ADD	0.76	0.04	0.002	0.008	8.0
		GARCH-LS	0.80	0.03	0.004	0.017	32.0
		GARCH-NL	0.81	0.03	0.005	0.019	32.8
B	1000	GARCH(1,1)	0.76	0.06	0.004	0.015	3.6
		GJR(1,1)	0.76	0.03	0.001	0.009	6.2
		GARCH-ADD	0.76	0.09	0.009	0.141	2.6
		GARCH-LS	0.82	0.06	0.010	0.066	11.1
		GARCH-NL	0.84	0.04	0.010	0.058	11.2
	2000	GARCH(1,1)	0.76	0.02	0.001	0.013	6.1
		GJR(1,1)	0.76	0.02	0.001	0.006	12.3
		GARCH-ADD	0.78	0.06	0.005	0.060	5.2
		GARCH-LS	0.82	0.04	0.007	0.035	21.0
		GARCH-NL	0.83	0.04	0.008	0.039	20.7
	3000	GARCH(1,1)	0.76	0.02	0.001	0.012	11.4
		GJR(1,1)	0.76	0.02	0.001	0.005	15.9
		GARCH-ADD	0.79	0.04	0.004	0.033	7.8
		GARCH-LS	0.82	0.03	0.006	0.028	31.0
		GARCH-NL	0.83	0.02	0.007	0.030	30.4

Table 2.4: Fitting the BMW daily returns

Model	−Log-likelihood	Volatility prediction error
GARCH(1,1)	3394.667	22.589
GJR(1,1)	3387.449	22.065
GARCH-ADD	3387.310	21.759

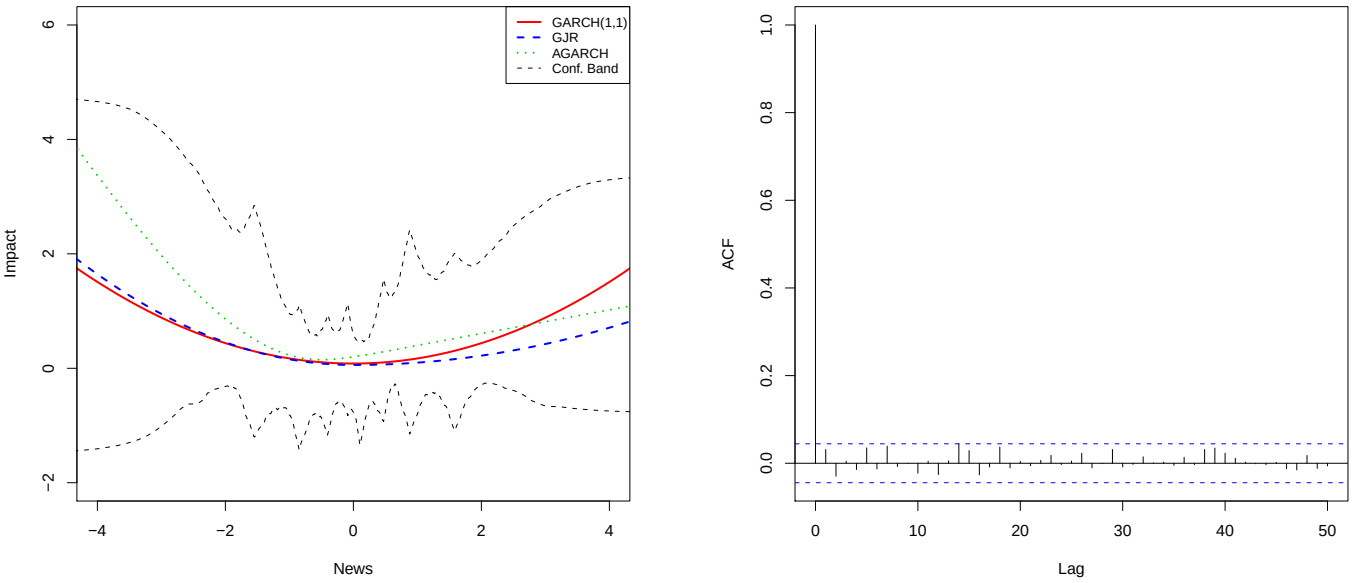


(a)



(b)

Figure 2.1: BMW daily returns: (a) original series; (b) the estimated volatility function.



(a) (b)

Figure 2.2: Spline ARCH(∞) modelling of BMW daily returns: (a)estimated news impact curve; (b) the estimated ACF along with 95% Bartlett intervals for $\hat{\epsilon}^2$.

Chapter 3

Modeling and Forecasting the Functional Time Series[1]

[1]Feng, C., Wang, L. and Seymour, L. (2012+). Modeling and forecasting the time series of treasury bond yield curves. To be submitted.

Abstract

A novel method is proposed for forecasting time series of smooth curves, using functional principal component (FPC) analysis in combination with time series modeling and FPC scores forecasting. We achieve smoothing, dimension reduction and prediction simultaneously with expedient computation. The work is motivated by the demand to forecast the time series of economic functions, such as Treasury bond yield curves. Extensive simulation studies have been carried out to compare the prediction accuracy of our method with other existing methods. The proposed methodology is applied to forecasting the yield curves of US Treasury bond.

KEYWORDS: functional time series, eigenfunctions, principal component analysis, spline smoothing, trend

3.1 Introduction

In financial economics, a yield curve is a graphic representation of the relationship between market remuneration rates and the remaining time to maturity of debt securities, also known as the term structure of interest rates. Economists often use the yield curve to capture the overall movement of interest rates, to forecast the path of the economy, and to estimate the likelihood of recessions and inflations. Yield curve prediction is also valuable in many other applications such as, bond portfolio selection, risk management and derivative pricing.

In the past, many efforts have been made to develop theoretical models of term structure interest rates. There are two main types of modeling approaches: one focuses on fitting the term structure at a certain time point to avoid any arbitrage opportunities; see Hull and Wright (1990) and Heath *et al.* (1992); the other focuses on the affine term structure model to capture the dynamics of the interest rate; see Vasicek (1977) and Cox *et al.* (1985). However, there are limitations to the above two approaches: the former one performs well in terms of fitting the historical cross-sectional term structure, but it can not provide references to the future time horizons; the performance of the second approach is relatively poor, even compared with the simple random walk.

Recently more research efforts have been devoted to yield curve forecasting. For example, Diebold and Li (2006) propose a dynamic three-factor Nelson-Siegel framework to model the yield curve, in which the three factors are treated as unobserved stochastic processes and fitted by traditional time series model. Koopman *et al.* (2010) improve the above model by introducing the time-varying factor loading and time-varying volatility. Without applying the Nelson-Siegel framework, Bowsher and Meeks (2008) developed a functional signal plus noise (FSN) model to treat the yield curve as the cubic spline functions. However, it lost economic intuition as it used the knots for splines as factors; see Koopman *et al.* (2010). Despite these powerful methods, modeling the time series of a cross-section of yields is still very challenging, especially when the cross-sectional dimension is large. Feasible methods for studying their dynamics are still in their infancy. In this paper, we consider both modeling and forecasting of Treasury bond yield curves.

The method introduced in this paper inspire the way to view the bond yield curves in the functional space.

Functional data analysis (FDA) has become a popular area in statistics research in recent years. As stated by Ramsey and Silverman (2005), statistical analysis has increasingly depended on functional data. In this article, we treat the observed data as random curves rather than finite-dimensional vectors. The dimensionality of the vector time series of yield curves is usually high. To achieve dimension reduction, Bowsher and Meeks (2008) propose modeling the continuous yield functions via natural cubic spline whose dynamic evolution is driven by a cointegrated vector autoregression for the ordinates at the knots of the spline. In an application of forecasting future call arrival rate profiles to telephone customer service centers, Shen (2009) proposes a low-dimensional smooth factor model (SFM) that constrains the factors to be smooth. Park et al. (2009) also propose a very interesting approach to reduce the high-dimensional time series to low-dimensional problem by using factor model while the factor and factor loading are estimated by the semi-parametric methods.

In this article, we apply one of the most commonly-used dimension reduction approaches in FDA, the functional principal component analysis (FPCA), to achieve the dimension reduction. Hyndman and Ullah (2007) study the age-specific mortality rate or fertility rate curves over past years, and use the FPCA to forecast the future mortality/fertility rate curves. Our proposed method differs from the existing proposals. Hyndman and Ullah (2007) first smooth each discrete curve before decomposing the smoothed curve by a basis function expansion, and then do the prediction. In this article, instead of pre-

smoothing the data curve-by-curve, we first decompose the original time series into a polynomial trend surface and a stationary noise surface, then we combine the smoothing, FPCA and time series forecasting in one model. When the FPCA is done, forecasting future yield curves reduces to forecasting time series of scores. Univariate time series models can be built for each factor score series to produce time series forecasts of the yield curve.

Our method produces multi-step ahead forecasts that outperform many other models, including the SFM in Shen (2009) and the random walk forecast. In addition, the FPCA is a novel exploratory tool which provides unique insight into the economical interpretation of the yield curve.

The rest of this paper is organized as follows. In Section 3.2, we state our model and describe the proposed forecasting method and the corresponding algorithm for functional time series. Section 3.3 reports findings from a simulation study. In Section 3.4, we illustrate the proposed method using the yields data of US Treasury bonds and compare the forecasting performance of our method with the rival models. Section 3.5 concludes.

3.2 Methodology

3.2.1 Models

Let $\{Y_t(u_j)\}_{t=1}^{n}$ be the observed historical data on some equally or unequally spaced grid $\{u_j\}_{j=1}^{m}$ at time points $t = 1, ..., n$. At the t^{th} time point, its sample path $\{u_j, Y_{tj}\}$

is the noisy realization of a random curve $X_t(u_j)$ in the sense that

$$Y_{tj} = X_t(u_j) + \sigma(u_j)\varepsilon_{tj},\ j = 1, ..., m,\ t = 1, ..., n \tag{3.2.1}$$

with independent errors ε_{tj} satisfying $E(\varepsilon_{tj}) = 0, E(\varepsilon_{tj}^2) = 1$. The stochastic process is modeled by

$$X_t(u) = \mu(u) + \alpha_t + \xi_t(u), \tag{3.2.2}$$

where $\mu(u)$ is some smooth but unknown function of $u \in \mathcal{U}$, α_t is the trend in the functional time series of $X_t(u)$, and the process $\xi_t(u)$ is a mean zero stationary stochastic process with $E\left[\int_{\mathcal{U}} \xi^2(u)du\right] < +\infty$. The trend term in (3.2.2), α_t, can be allowed to depend on some unknown parameters, but at this stage it is assumed to be known and set equal to zero without any loss of generality.

We are interested in predicting the future smooth curve $X_{n+h}(u)$, for some $h \geq 1$ based on the history curves. To achieve this, we model samples of random functions $X_1(u), ..., X_n(u)$ through functional principal components (FPC) analysis. For the process $\{\xi(u), u \in \mathcal{U}\}$, define the covariance function $G(u, u') = \text{cov}\{\xi(u), \xi(u')\}$. Let sequences $\{\lambda_k\}_{k=1}^{\infty}, \{\phi_k(u)\}_{k=1}^{\infty}$ be the eigenvalues and eigenfunctions of $G(u, u')$, respectively, in which $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, and $\{\phi_k\}_{k=1}^{\infty}$ form an orthonormal basis in the L^2 sense, such that $G(u, u') = \sum_{k=1}^{\infty} \lambda_k \phi_k(u)\phi_k(u')$.

The process $\{\xi_t(u), u \in \mathcal{U}\}$ allows the Karhunen-Loève L^2 representation (Rice and Silverman, 1991)

$$\xi_t(u) = \sum_{k=1}^{\infty} \beta_{t,k}\phi_k(u),$$

where $E(\beta_{t,k}) = 0$, $\text{var}(\beta_{t,k}) = \lambda_k$, $\text{cov}(\beta_{t,k}, \beta_{t,k'}) = 0$ for any fixed $t \geq 1$, $k \neq k' \geq 1$. In what follows, we assume that each series of the scores $\left\{\beta_{t,k}, t \geq 1\right\}$, $k = 1, 2...$, are dependent and predictable. We further assume that $\lambda_k = 0$, for $k > \kappa$, where κ is a positive integer or $+\infty$. The data process can now be modeled though the following

$$Y_{tj} = \mu(u_j) + \sum_{k=1}^{\kappa} \beta_{t,k}\phi_k(u_j) + \sigma(u_j)\varepsilon_{tj}. \tag{3.2.3}$$

3.2.2 Spline Estimators

To estimate the mean function $\mu(\cdot)$ and the covariance function $G(\cdot,\cdot)$, we apply the polynomial spline approximation. Let $\{B_{l,p}, l = 1-p, ..., N\}$ be B-spline basis functions of order p; see de Boor (2001) for more details of spline smoothing.

We propose to smooth the mean function $\mu(\cdot)$ by aggregating the observed observations from all sample trajectories. The estimator $\hat{\mu}(\cdot)$ is obtained from

$$\hat{\mu}(u) = \underset{g(\cdot)\in\mathcal{H}^{(p-2)}}{\text{argmin}} \sum\nolimits_{t=1}^{n} \sum\nolimits_{j=1}^{m} \{Y_{tj} - g(u_j)\}^2 = \sum_{l=1-p}^{N_\mu} \hat{a}_l B_{l,p}(u),$$

where N_μ is the number of interior knots of the B-spline basis and the coefficients are:

$$\{\hat{a}_l, l = 1-p, ..., N_\mu\} = \underset{R^{N_\mu+p}}{\operatorname{argmin}} \sum_{t=1}^{n}\sum_{j=1}^{m}\left\{Y_{tj} - \sum_{l=1-p}^{N_\mu} a_l B_{l,p}(u_j)\right\}^2. \tag{3.2.4}$$

Let $\hat{R}_{\cdot jj'} = n^{-1}\sum_{t=1}^{n}\{Y_{tj} - \hat{\mu}(u_j)\}\{Y_{tj'} - \hat{\mu}(u_{j'})\}$, $1 \le j \ne j' \le m$. We estimate the covariance function $G(u, u')$ using the tensor product spline approach by Cao, et al (2011). The spline estimator of $G(u, u')$ is defined as

$$\hat{G}(u, u') = \sum\nolimits_{l,l'=1-p}^{N_G} \hat{b}_{ll'} B_{l,p}(u) B_{l',p}(u'),$$

where N_G is the number of interior knots used to build the tensor product B-spline basis and the spline coefficients

$$\left\{\hat{b}_{ll'}\right\}_{l,l'=1-p}^{N_G} = \underset{R^{N_G+p}\otimes R^{N_G+p}}{\operatorname{argmin}} \sum_{j\ne j'}^{m}\left\{\hat{R}_{\cdot jj'} - \sum_{1-p\le l,l'\le N_G} b_{ll'} B_{l,p}(u_j) B_{l',p}(u_{j'})\right\}^2.$$

3.2.3 Estimation of the Eigenvalues and Eigenfunctions

The estimates of eigenfunctions and eigenvalues correspond to the solutions of $\hat{\phi}_k$ and $\hat{\lambda}_k$ of the eigenequations,

$$\int_{\mathcal{U}} \hat{G}(u, u')\hat{\phi}_k(u)\, du = \hat{\lambda}_k \hat{\phi}_k(u'),$$

where the $\hat{\phi}_k$s are subject to $\int_{\mathcal{U}} \hat{\phi}_k(u)^2 du = 1$ and $\int_{\mathcal{U}} \hat{\phi}_k(u)\hat{\phi}_{k'}(u) du = 0$ for $k' < k$; see Bathia *et. al* (2010). To solve this equation, we define a fine grid $(v_1, ..., v_{n_v})$ consisting of n_v points equally spaced on [0,1], We estimate the FPC scores by $\hat{\beta}_{t,k} = \sum_{j=1}^{n_v}(Y_{tj} - \hat{\mu}(v_j))\hat{\phi}_k(v_j)(v_j - v_{j-1})$, setting $v_0 = 0$.

Finally, to choose the number of eigenfunctions that provide a reasonable approximation to the infinite-dimensional process, we apply a simple criterion in Müller (2009), i.e. $\kappa = \text{argmin}_{1\leq q\leq K}\left\{\sum_{k=1}^{q}\hat{\lambda}_k / \sum_{k=1}^{K}\hat{\lambda}_k > 0.95\right\}$. Other methods like the cross-validation in Yao, Müller and Wang (2005) might also work. We used the simple method and it works well in simulations.

3.2.4 Curve Forecasting

Given the historical observations $\{Y_{tj}, 1 \leq t \leq n, 1 \leq j \leq m\}$, we now forecast the h-step ahead underlying smooth curve $X_{n+h}(u)$ for some $h \geq 1$.

We first forecast the FPC scores. For each $k = 1, ..., \kappa$, one can use the classical time series models, such as AR models or ARMA models, to forecast the future scores $\left\{\beta_{t,k}, t \geq n+1\right\}$ based on the estimated values $\left\{\hat{\beta}_{t,k}, t = 1, ..., n\right\}$. For any $k = 1, ..., \kappa$, the h-step ahead forecasts of $\beta_{t,k}$ is denoted by $\hat{\beta}_{t+h,k}$. Then for any fixed u, the h-step ahead forecast of $X_t(u)$ is

$$\hat{X}_{t+h}(u) = \hat{\mu}(u) + \sum_{k=1}^{\kappa}\hat{\beta}_{t+h,k}\hat{\phi}_k(u). \tag{3.2.5}$$

Finally, we summarize our algorithm to forecast $\hat{X}_{t+h}(u)$ as the following:

1. Estimate and remove the trend in the functional time series;
2. For the detrended time series, estimate the mean function $\mu(u)$ and covariance function $G(u, u')$;
3. Estimate the eigenfunctions $\phi_k(u)$'s and eigenvalues λ_k's;
4. Estimate the FPC scores by $\hat{\beta}_{t,k}$ and forecast the future scores $\hat{\beta}_{t+h,k}$, k=1, ..., κ;
5. Forecast the future curve $\hat{X}_{t+h}(u)$.

3.3 Simulation

To illustrate the forecasting performance of the proposed approach, we generate data from model

$$Y_{tj} = \mu(u_j) + \sum_{k=1}^{2} \beta_{t,k}\phi_k(u_j) + \sigma\varepsilon_{tj}, \quad 1 \le j \le 30,\ 1 \le t \le n, \tag{3.3.1}$$

where $u = (3, 6, ..., 60, 63, 69, ..., 87, 93, 105, 117, 129, 141)^T/141$ is a 30-dimensional grids unequally spaced on interval $[0, 1]$, and functions $\mu(u) = \sin\{2\pi(u - 1/2)\} + 10u + 10$, $\phi_1(u) = -\sqrt{2}\cos\{\pi(u - 1/2)\}$, $\phi_2(u) = \sqrt{2}\sin\{\pi(u - 1/2)\}$. The noise sequence ε_{tj}'s are i.i.d. r.v.'s $\sim N(0, 1)$ and the error standard deviation $\sigma = 0.5$, 1.0 and 1.5. The scores $\beta_{t,1}$ and $\beta_{t,2}$ are simulated from an AR(2) process and an ARMA(1,1) process,

respectively,

$$\beta_{t,1} = 0.75\beta_{t-1,1} + 0.20\beta_{t-2,1} + \epsilon_t, \quad \beta_{t,2} = 0.80\beta_{t-1,2} + 0.30\epsilon_{t-1} + \epsilon_t,$$

where ϵ_t's and ϵ_{t-1}'s are i.i.d. r.v.'s $\sim N(0,1)$. We rescaled $\beta_{t,1}$ and $\beta_{t,2}$ such that $\text{var}(\beta_{t,2}) = \lambda_1 = 0.4$ and $\text{var}(\beta_{t,2}) = \lambda_2 = 0.1$.

For each $n = 30, 50, 100$, $n + 1000 + 30$ time series of curves are generated according to model (3.3.1). For each realization, the last $n + 30$ observations are kept as our data for inference. Truncating the first 1000 observations off the series ensures that the remaining series behaves like a stationary series. Figure 3.1 illustrates one realization of simulated time series of 100 curves based on model (3.3.1). We split each replication into two datasets: the training set of size n and the test set of size 30.

We compare our approach with the simple random walk forecasts (RW) and the smooth-factor-model (SFM) approach of Shen (2009). For the number of eigenfunctions in our modeling, we try both $\kappa_{fit} = 2$ and the estimated κ_{fit} based on the simple criterion (Section 4.2.3), and we call the former the FPCAR approach and the latter the FPCAR* approach.

We always use the first n simulated curves in the training set to estimate the model and the successive 30 curves as the out-of-sample testing period. For the FPCAR and FPCAR* approaches, we estimate the mean function $\mu(\cdot)$ and the covariance function $G(\cdot,\cdot)$ using cubic B-splines. In the mean estimation, the number of knots N_μ was selected by the generalized cross-validation (GCV). To smooth the covariance functions

$N_G = n^{1/(2p)} \log(\log(n))$ was set as suggested in Cao et al. (2010). The knots of the spline basis functions are set at equally spaced sample quantiles of u_js.

After the model fitting, we roll the data series 1-curve forward and perform a 1-curve ahead forecast of the FPC scores via AR models and obtain the curve forecasts $\hat{X}_{t+1}(u)$ in (3.2.5). For quantifying the goodness of fit, we compute the forecasting root mean square error (RMSE) and average percent error (APE) for the $(t+1)$th curve in the data set:

$$\begin{aligned} RMSE_{t+1} &= \sqrt{\frac{1}{m}\sum\nolimits_{j=1}^{m}\left(X_{t+1}(u_j) - \hat{X}_{t+1}(u_j)\right)^2}, \\ APE_{t+1} &= \frac{100}{m}\sum\nolimits_{j=1}^{m}\frac{\left|X_{t+1}(u_j) - \hat{X}_{t+1}(u_j)\right|}{X_{t+1}(u_j)}, \end{aligned}$$

for $t = n, ..., n+29$.

We repeat the above simulation procedure 100 times. For each simulated dataset, we calculate the mean APE and mean RMSE of 30 forecasted curves. The simulation results are represented in boxplot format (see Fig 3.2 and Fig 3.3) with respect to different sample sizes and errors, which display the median, upper quartile, lower quartile, outliers of all approaches (FPCAR, FPCAR*, SFM and RW from left to right). The proposed FPCAR and FPCAR* approaches consistently outperform other competitor approaches without regard to the training sample size or the error rate (refer to Table 3.1 and Table 3.2). In addition, there seems to be no noticeable difference between the FPCAR approach and the FPCAR* approach, which indicates the suggested κ selection works

fairly well. We also observe that accuracy performance of the FPCAR and FPCAR* are very robust even when the data become very noisy.

3.4 Application to the Yield Curves of US Treasury Bonds

In this section, we provide a study of a time series of yield curves. We investigate a sample of monthly price quotes of U.S. Treasury bonds from July 1986 through October 1999, a total of 160 curves. A similar dataset was analyzed by Diebold and Li (2006). To avoid missing values, the data set comprises monthly real yield forward curves for maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months, where a month is defined as 30.4375 days. The upper graph in Figure 3.4 shows a three-dimensional plot of the yield curve data.

The purpose here is to forecast the future yield curves based on the historical curves. We use the first 150 curves (i.e. from July 1986 to December 1998) as the training set and the successive remaining 10 curves as out-of-sample test data.

Before applying the proposed FPCAR method to the training set, we first use the polynomial regression to estimate the trend of the time series. Then we remove the temporal structure of the time series, and work with the detrended series. Figure 3.4 shows the estimated trend (middle graph) in terms of time and the detrended time series (lower graph).

We then estimated mean function and the covariance function using cubic B-splines. Seven knots are selected to smooth the mean function according to the GCV, and three knots are used to smooth the covariance function. The estimates of the mean function and the covariance surface are presented in Figures 3.5 and 3.6. From Figure 3.5, one sees that the mean yield curve exhibits the typical upward sloping shape, which indicates long-term yields are higher than the short-term yields. One also notices that there is sharp increase in the peak at short maturities from month 3 to month 18, then it grows gradually till the end of maturity 120 months. This suggests that the gap between short-term and medium-term rates is greater than that between the medium-rate and long-term rates. The estimated covariance surface we obtained using tensor-product spline smoothing is shown to have very low volatility at the origin followed by a rapid growth.

The criterion introduced in Section 2.3 suggests that first two principal components are sufficient enough to explain the modes of variability. These two principal component functions account for 99.11% of total variation. The estimated eigenvalues are $\hat{\lambda}_1 = 0.3162$ and $\hat{\lambda}_2 = 0.1037$. The estimated eigenfunctions are shown in Figure 3.7. The first eigenfunction indicates that a large proportion of the variability along the time axis is in the same direction of the amplitude of mean curve, as it seems to have similar shape to the mean curve. We note that 74.63% of total variability is explained by the first eigenfucntion, which is dominant. The second FPC contributes 24.48% to the total variation, which can be interpreted as a slope factor, which may cause the very long-term yield to fall due to its negative values.

We next make the 1 to 10-step ahead forecasts of the yield curves. Figure 3.8 shows the forecasted FPC scores. Figure 3.9 shows the forecasts of yield curves from January to October of year 1999.

The dynamic Nelson-Siegel model (DNS) method of Diebold and Li (2006) is widely used for forecasting yield curves. We compare the forecast performance of our method with the DNS method of Diebold and Li (2006), SFM method of Shen (2008) and the random walk model. Table 3.3 shows the mean APE and mean RMSE of the above four forecasting methods. Our method consistently outperforms the widely used DNS method of Diebold and Li (2006), Shen (2009) and random-walk forecasts described by Fama (1965) on the basis of both mean APE and mean RMSE.

3.5 Conclusion

Motivated by the demand to forecast the time series of economic functions, we propose a feasible and effective method to modeling and forecasting time series of various functions/curves such as the yield curves. We propose efficient functional principal component-based methods for modeling and forecasting time series of Treasury yield curves. Without applying the traditional econometric model, our methodology combines the functional data analysis and time series model, to achieve more accurate prediction performance and more expedient computing speed than competing methods. Our new approach is based on the analysis of time series of stochastic and continuous functions, which is significantly different from the standard approaches of multivariate time series

or panel data analysis. The latter can work poorly when the cross-sectional dimension of the data is large. Our approach does not require pre-smoothing of the data and naturally combines nonparametric smoothing with functional principal component analysis. In addition, the method proposed consistently outperforms both the widely used DNS method of Diebold and Li (2006), SFM method of Shen (2008) and random-walk forecasts.

Our methods are motivated by and applied to yield curve data. However, our method can be generally applied to many data which are temporally dependent functional data or time series of smooth functions.

3.6 References

Bathia, N., Q. Yao, and F. Zieglemann (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics* **38**, 3352-3386.

Cai, T. and Hall, P. (2005). Prediction in functional linear regression. *Annals of Statistics*, **34**, 2159-2179.

Cao, G., Yang, L. and Todem, D. (2012) Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics* **24**, 359-377.

Cao, G., Wang, L., Li, Y. and Yang, L. (2011). Spline confidence envelopes for covariance function in dense functional/longitudinal data. *Manuscript.*

de Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.

Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**, 337-364.

Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal* **21**, 55-59.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice.* Springer Series in Statistics, Springer: Berlin.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493-1517.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis* **51**, 4942-4956.

James, G. M., Hastie, T. and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika* **87**, 587-602.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *Journal of the American Statistical Association* **100**, 565-576.

Koopman, S. J., Mallee, M. and van der Wel, M. (2005). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business & Economic Statistics* **28**, 329-343.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics* **38**, 3321–3351.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68**, 179-199.

Müller, H. G. (2009). Functional modeling of longitudinal data. *In: Longitudinal Data Analysis (Handbooks of Modern Statistical Methods)*, Ed. Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., Wiley, New York, 223-252.

Park, B., Mammen, E., Hardle, W. and Borak, S. (2009). Time series modeling with semiparametric factor dynamics. *Journal of the American Statistical Association*, **104**, 284-298.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society B, **53**, 233-243.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Second Edition. Springer Series in Statistics. Springer: New York.

Shen, H. (2009). On modeling and forecasting time series of curves. Technometrics 51, 227-238.

Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B* **68**, 3-25.

Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873-2903.

Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590.

Zhao, X., Marron, J. S. and Wells, M. T. (2004) The functional data analysis view of longitudinal data. *Statistica Sinica* **14**, 789-808.

Zhou, L., Huang, J. and Carroll, R. J. (2008). Joint modeling of paired sparse functional data using principal components. *Biometrika* **95**, 601-619.

Table 3.1: Summary statistics of the mean APE (%) based on 100 replications.

Size	Error	FPC-AR*				SFM				RW			
n	σ	Q1	median	mean	Q3	Q1	median	mean	Q3	Q1	median	mean	Q3
30	0.5	2.448	2.646	2.728	2.933	3.790	4.081	4.245	4.696	3.722	3.913	4.016	4.253
		(2.445)	(2.636)	(2.726)	(2.931)								
	1.0	2.905	3.215	3.242	3.504	4.057	4.378	4.532	5.003	6.521	6.674	6.746	6.967
		(2.877)	(3.189)	(3.231)	(3.515)								
	1.5	3.457	3.740	3.779	4.117	4.427	4.782	4.892	5.312	9.448	9.653	9.707	9.934
		(3.411)	(3.729)	(3.752)	(4.066)								
50	0.5	2.112	2.406	2.392	2.662	3.251	3.700	3.736	4.206	3.616	3.839	3.818	3.988
		(2.104)	(2.401)	(2.389)	(2.658)								
	1.0	2.582	2.927	2.918	3.234	3.572	4.074	4.064	4.448	6.433	6.687	6.636	6.826
		(2.575)	(2.905)	(2.909)	(3.243)								
	1.5	3.092	3.529	3.516	3.917	3.945	4.430	4.471	4.888	9.367	9.696	9.644	9.912
		(3.113)	(3.495)	(3.498)	(3.865)								
100	0.5	2.015	2.199	2.234	2.445	3.105	3.387	3.472	3.770	3.592	3.720	3.756	3.908
		(2.019)	(2.202)	(2.233)	(2.445)								
	1.0	2.473	2.739	2.744	2.995	3.397	3.784	3.801	4.072	6.432	6.594	6.617	6.804
		(2.450)	(2.752)	(2.747)	(3.037)								
	1.5	2.887	3.259	3.289	3.591	3.818	4.190	4.201	4.521	9.392	9.632	9.649	9.903
		(2.886)	(3.260)	(3.294)	(3.599)								

$*$ Values inside ( ) are the APEs based on optimal $\kappa = 2$; the values outside ( ) are the APEs based on selected κ by the simple criterion described in Section 2.3.

Table 3.2: Summary statistics of the mean RMSE based on 100 replications.

Size	Error	FPC-AR*				SFM				RW			
n	σ	Q1	median	mean	Q3	Q1	median	mean	Q3	Q1	median	mean	Q3
30	0.5	0.3516	0.3809	0.3952	0.4222	0.5429	0.5953	0.6215	0.6658	0.5860	0.6138	0.6306	0.6526
		(0.3510)	(0.3808)	(0.3945)	(0.4212)								
	1.0	0.4347	0.4703	0.4744	0.5065	0.5964	0.6447	0.6679	0.7176	1.0400	1.0640	1.0700	1.0960
		(0.4287)	(0.4669)	(0.4712)	(0.5050)								
	1.5	0.5161	0.5565	0.5578	0.5982	0.6630	0.7137	0.7243	0.7725	1.508	1.539	1.541	1.571
		(0.5049)	(0.5456)	(0.5502)	(0.5945)								
50	0.5	0.3113	0.3431	0.3437	0.3746	0.4874	0.5483	0.5453	0.5874	0.5704	0.5942	0.5995	0.6262
		(0.3086)	(0.3420)	(0.3429)	(0.3739)								
	1.0	0.3855	0.4260	0.4220	0.4625	0.5470	0.5945	0.5967	0.6348	1.0300	1.0460	1.0490	1.0730
		(0.3834)	(0.4227)	(0.4200)	(0.4622)								
	1.5	0.4578	0.5146	0.5094	0.5607	0.6118	0.6575	0.6587	0.7143	1.5070	1.5260	1.5250	1.5510
		(0.4510)	(0.5040)	(0.5049)	(0.5587)								
100	0.5	0.2871	0.3171	0.3208	0.3562	0.4534	0.5096	0.5096	0.5571	0.5677	0.5874	0.5911	0.6114
		(0.2861)	(0.3161)	(0.3202)	(0.3561)								
	1.0	0.3594	0.3956	0.3974	0.4338	0.5119	0.5658	0.5621	0.6065	1.0310	1.0470	1.0480	1.0650
		(0.3552)	(0.3990)	(0.3965)	(0.4330)								
	1.5	0.4250	0.4797	0.4789	0.5246	0.5633	0.6254	0.6240	0.6763	1.5050	1.5300	1.5290	1.5480
		(0.4253)	(0.4796)	(0.4767)	(0.5182)								

$*$ Values inside ( ) are the RMSEs based on optimal $\kappa = 2$; the values outside ( ) are the RMSEs based on selected κ by the simple criterion described in Section 2.3.

Table 3.3: Mean APE(%) and RMSE for US yield curve forecasting.

	FPCAR	SFM	DNS	RW
APE	3.7412	9.9379	3.6876	12.9360
RMSE	0.2247	0.5924	0.2261	0.7626

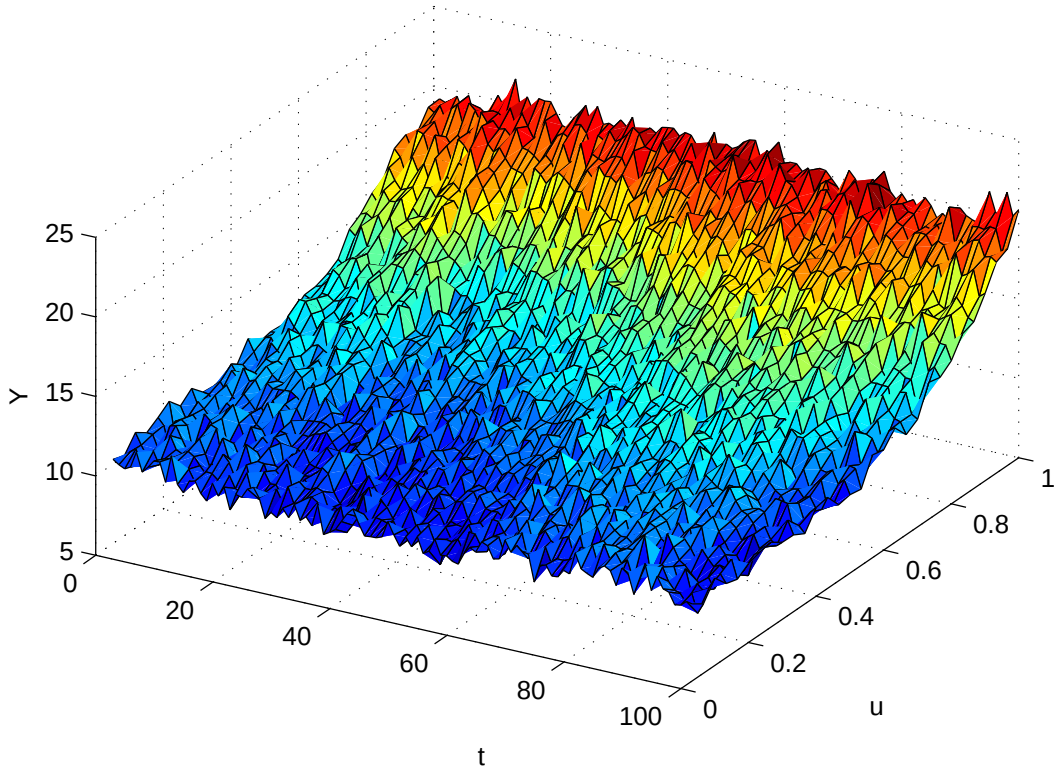


Figure 3.1: One realization of the simulated dataset.

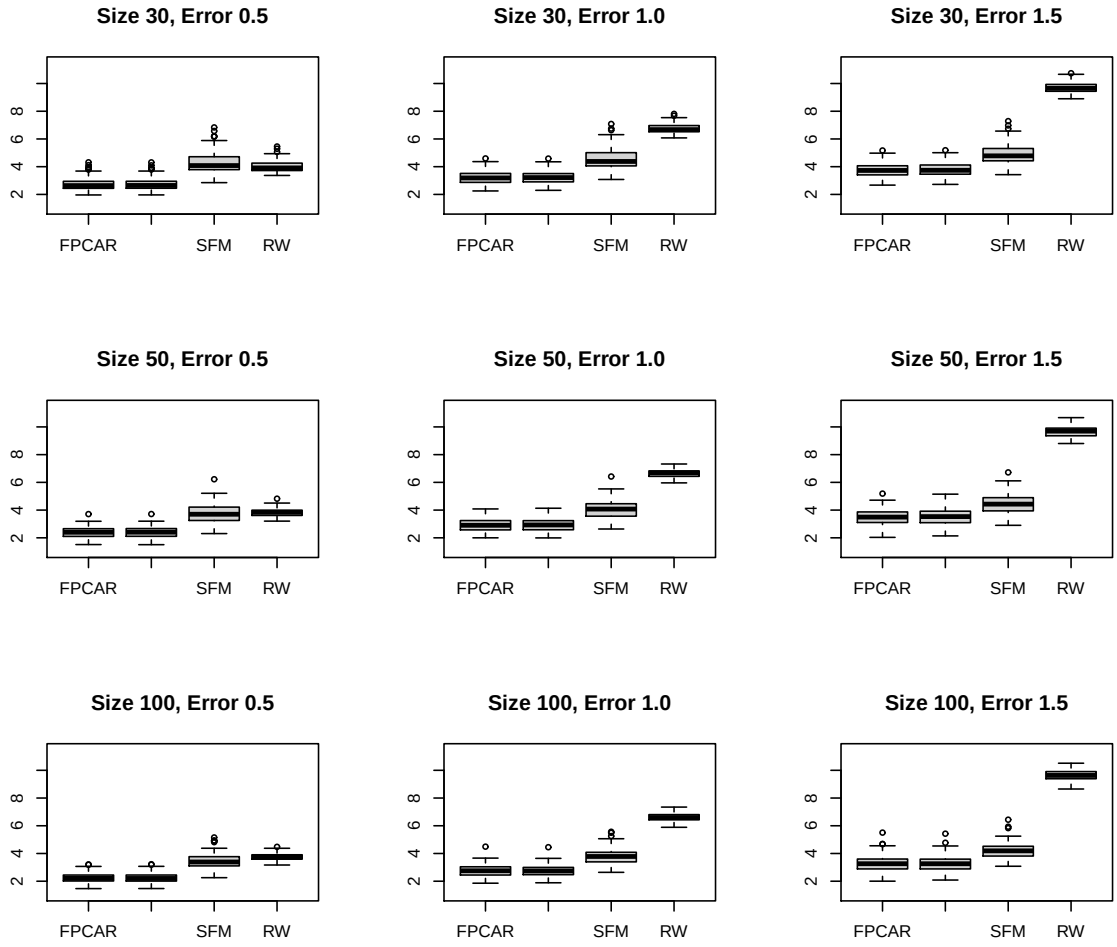


Figure 3.2: Boxplot of the mean APEs using FPCAR, FPCAR*, SFM, RW.

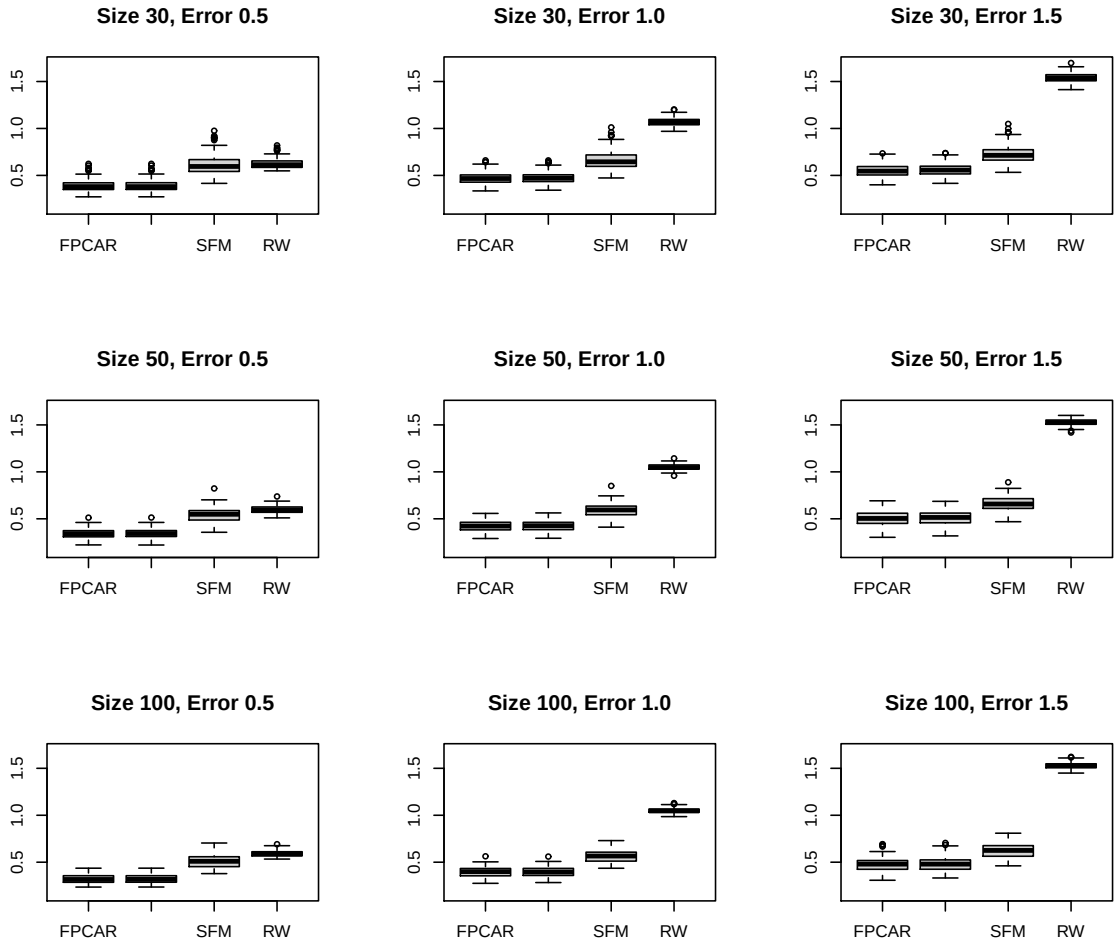


Figure 3.3: Boxplot of the mean RMSEs using FPCAR, FPCAR*, SFM, RW.

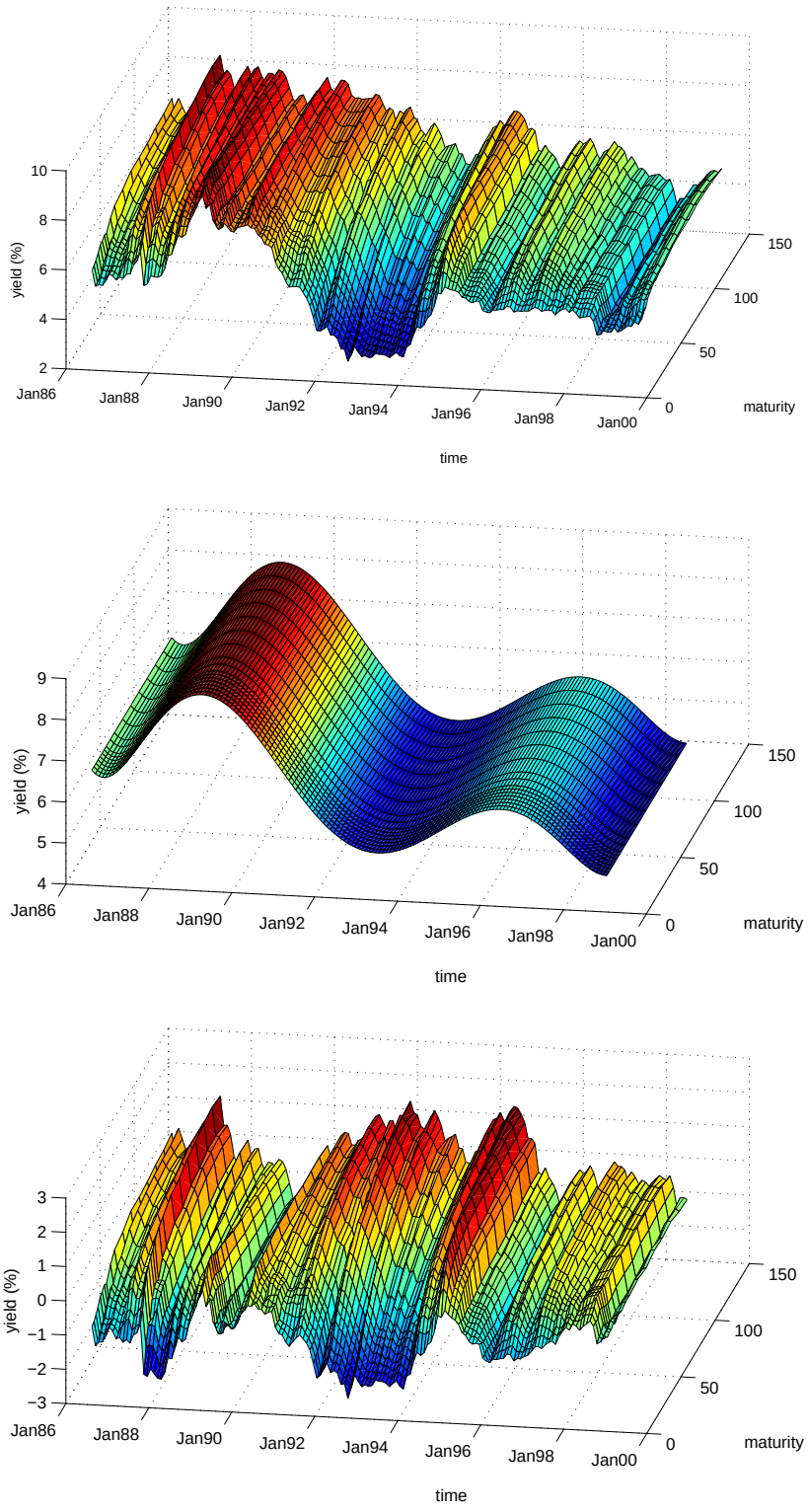


Figure 3.4: Original (upper), trend (middle) and detrended (lower) US Treasury bonds.

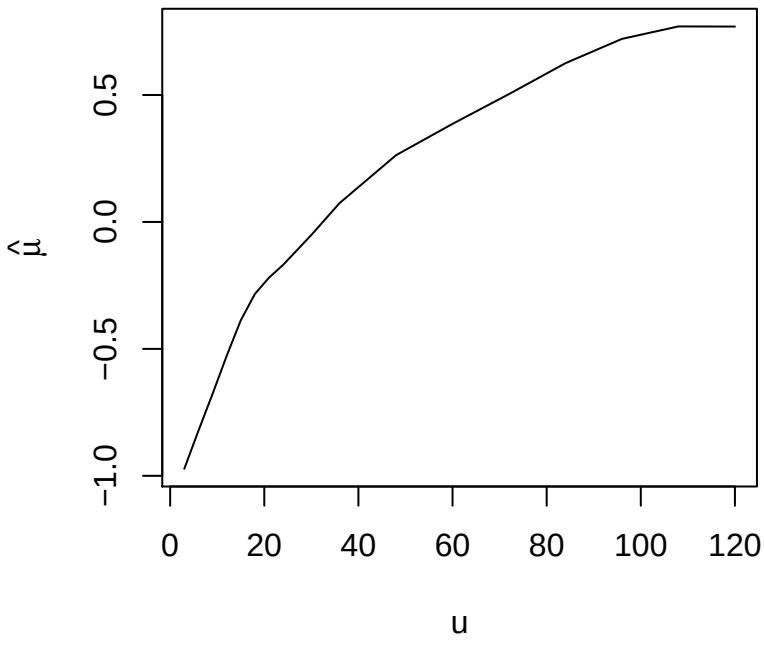


Figure 3.5: Cubic spline estimate of the mean function.

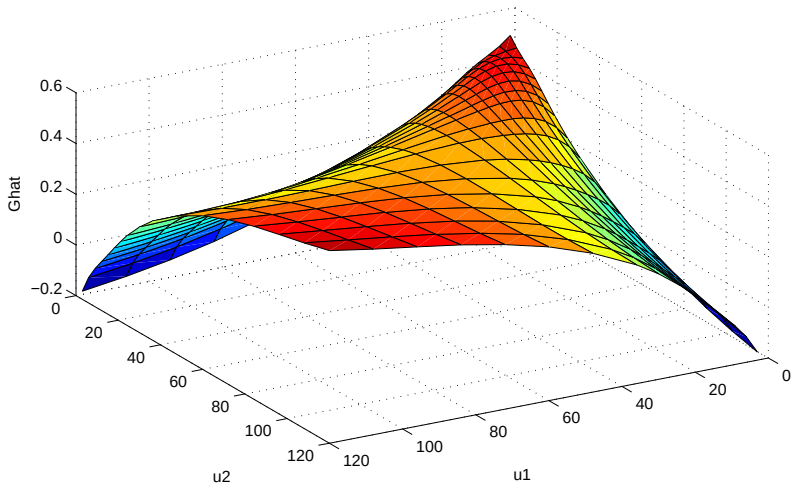


Figure 3.6: Cubic spline estimate of the covariance surface.

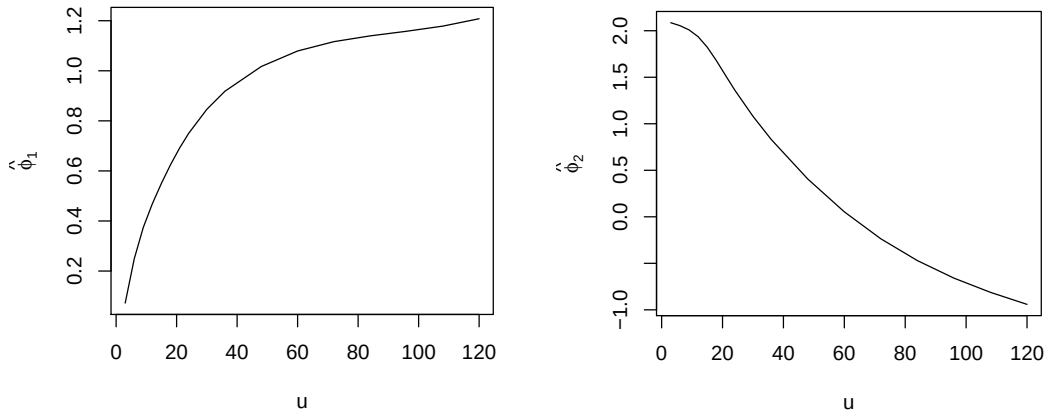


Figure 3.7: Estimated eigenfunctions $\hat{\phi}_1(u)$ and $\hat{\phi}_2(u)$.

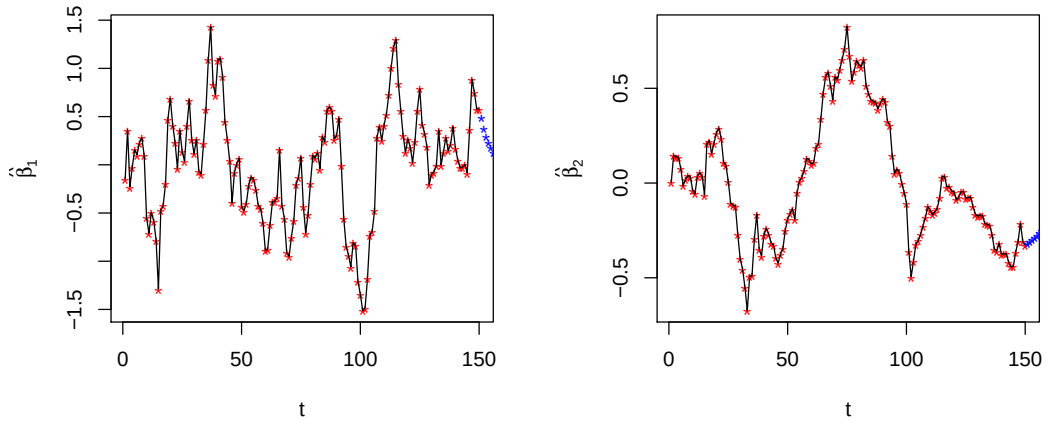


Figure 3.8: Estimated FPC scores and their 1 to 10-step ahead forecasts.

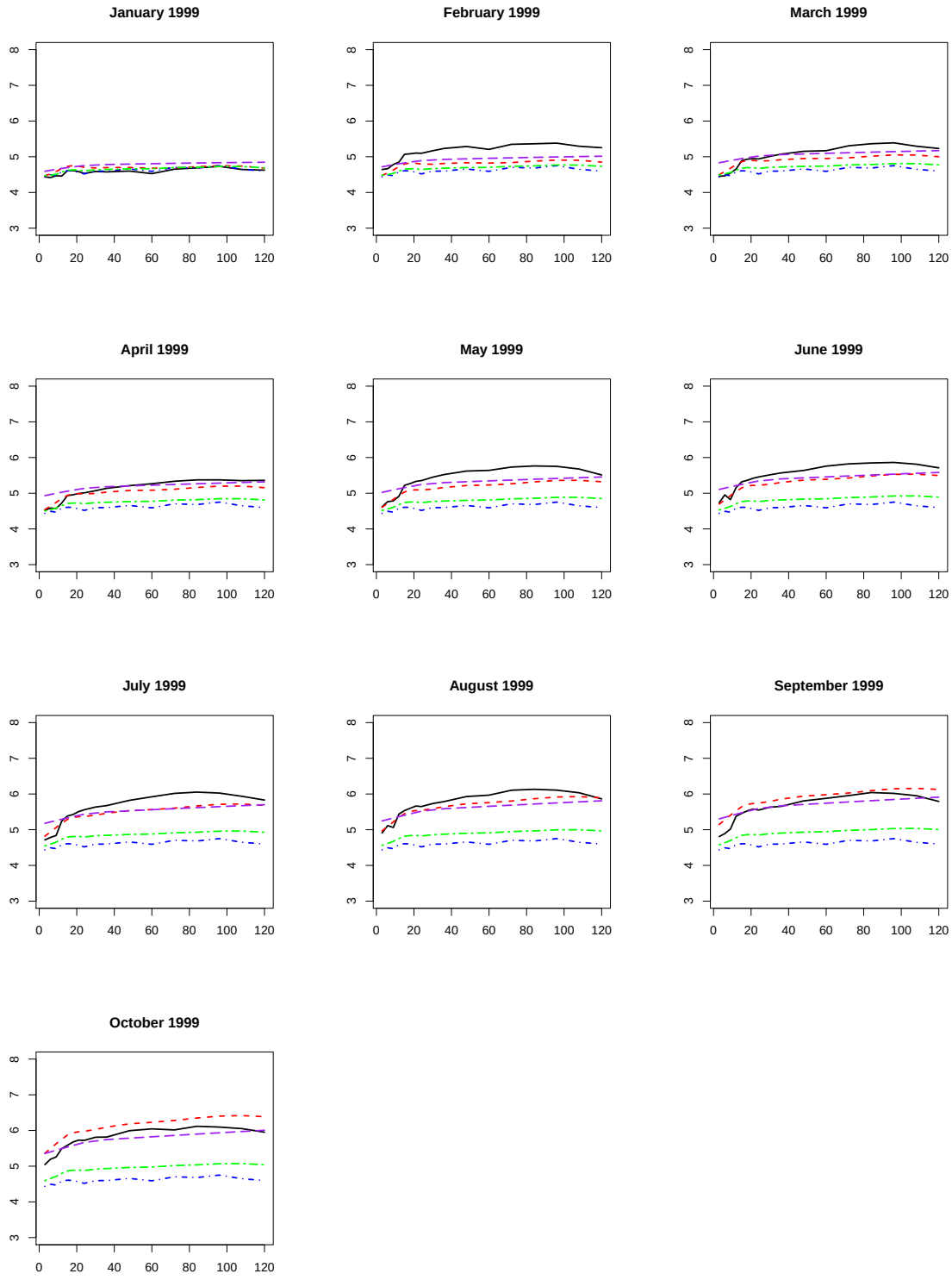


Figure 3.9: 1 to 10 step-ahead forecasts: real (black), FPCAR* (red), SFM (green), DNS (purple) and RW (blue).

Chapter 4

Two-Sample Comparison for Functional Derivatives[1]

[1]Feng, C., Wang, L., Seymour, L., Stooksbury, D. (2012+). Two Sample Comparison for Functional Derivatives with an Application to Temperature Data. To be submitted.

Abstract

We develop a novel method to construct simultaneous confidence bands for the difference of derivatives of regression functions from two groups. This method was derived to answer a question arising from climate: Are the temperature transitions in Athens, Georgia different under the substantial global atmospheric pressure oscillations? We show that our proposed procedure has desirable theoretical properties. In particular, we show that the proposed spline confidence bands are asymptotically efficient as if there is no measurement error. In the end, the performance of the confidence bands is illustrated through numerical simulation studies and a temperature data collected in Athens, GA, in US.

KEYWORDS: B-spline, ENSO, functional data analysis, Karhunen-Loève representation, simultaneous confidence band

4.1 Introduction

Anecdotally, one may hear the question "How did temperature change in response to El Niño?" as a general observation about the weather. However, statistically, how can we establish whether the temperature transition is significantly different under these major atmospheric pressure oscillations? We propose to answer this question by examining the curves defined by the average daily temperature between October and September in the next year, indexed by year. Then, the climate question reduces to: Is the derivative of the

temperature curve between October and next September changing under different global atmospheric situations? To study the temperature dynamics, we target the two-group comparison of functional derivatives of mean curves by constructing the simultaneous confidence band for the difference of derivatives.

Functional data analysis (FDA) is a very useful tool to address the challenges posed by increasing complex data subjects, which are typically curves or surfaces. Functional derivatives of mean functions are of great importance in many real-life applications, such as studies of climate change, growth potential and acceleration.

Estimation and inference of derivatives of the mean functions has already attracted attention from statisticians; see Liu and Müller (2009) and Hall, Müller and Yao (2009). These methodologies do not require densely observed data for estimation and provided the pointwise confidence band for derivatives. Research for studying the simultaneous confidence band of mean curves can be seen from Degras (2011), and Ma, Yang and Carroll (2012). Cao et al (2012) developed a procedure to construct simultaneous confidence bands for the derivatives of one sample mean function in functional data analysis. In this paper, we extend Cao et al (2012) to the difference of derivatives of regression functions from two groups. To our knowledge, there is not any methodology that provides simultaneous confidence bands for the difference of functional derivatives in FDA.

The rest of the paper is organized as follows. Section 4.2 introduces the model and proposed spline estimator for the mean functions and their derivatives based on two samples. Section 4.3 presents the simultaneous confidence bands for the difference of

derivatives of mean curves from two groups. In Section 4.4, we report the coverage rate of proposed confidence bands in an extensive simulation study. An application to a temperature data set is given in Section 4.5.

4.2 Methodology

4.2.1 Models and Data

Let $X_{A1}(u)$, $X_{A2}(u)$, ..., $X_{An_A}(u)$ and $X_{B1}(u)$, $X_{B2}(u)$, ..., $X_{Bn_B}(u)$ be iid realizations of $X_A(u)$ and $X_B(u)$ defined on the continuous interval $\mathcal{U}$ for groups A and B. We assume that $X_H(u), u \in \mathcal{U}$ is a $L^2(\mathcal{U})$ process, i.e. $E[\int_{\mathcal{U}} X_H^2(u)]du < +\infty$, for $H = A$ and B. For each group, the square integrable stochastic process $X_{Ht}(u)$ is modeled by

$$X_{Ht}(u) = \mu_H(u) + \eta_{Ht}(u), \ t = 1, ..., n_H, \ H = A, B, \tag{4.2.1}$$

where $\mu_H(u)$ is some smooth but unknown function of $u \in \mathcal{U}$, and the process $\eta_{Ht}(u)$ is a mean zero stationary stochastic process with $E\left[\int_{\mathcal{U}} \eta_H^2(u)du\right] < +\infty$. We define the covariance function $G_H(u, u') = \text{cov}\{\eta_H(u), \eta_H(u')\}$ for the process $\{\eta_H(u), u \in \mathcal{U}\}$, $H = A, B$.

The problem considered in this paper is of testing the hypothesis that the two groups of random curves have the same ν-th order derivative functions for their mean functions against the alternative that the ν-th order derivative functions are different for two

groups; In symbols: $H_0 : \mu_A^{(\nu)} = \mu_B^{(\nu)}$ v.s. $H_a : \mu_A^{(\nu)} \neq \mu_B^{(\nu)}$. For this objective, we construct the confidence bands for $\mu_A^{(\nu)} - \mu_B^{(\nu)}$.

We apply functional principal components analysis to random functions $\{X_{At}(u)\}_{t=1}^{n_A}$ and $\{X_{Bt}(u)\}_{t=1}^{n_B}$ for groups A and B. For $H = A, B$, let sequences $\{\lambda_{Hk}\}_{k=1}^{\infty}, \{\psi_{Hk}(u)\}_{k=1}^{\infty}$ be the eigenvalues and eigenfunctions of $G_H(u, u')$, in which $\lambda_{H1} \geq \lambda_{H2} \geq \cdots \geq 0$, $\sum_{k=1}^{\infty} \lambda_{Hk} < \infty$, and $\{\psi_{Hk}\}_{k=1}^{\infty}$ form an orthonormal basis, such that $G_H(u, u') = \sum_{k=1}^{\infty} \lambda_{Hk} \psi_{Hk}(u) \psi_{Hk}(u')$.

The process $\{\eta_{Ht}(u), u \in \mathcal{U}\}, H = A, B$ has the following Karhunen-Loève L^2 representation (Rice and Silverman, 1991)

$$\eta_{Ht}(u) = \sum_{k=1}^{\infty} \xi_{Htk} \psi_{Hk}(u),$$

where $E(\xi_{Htk}) = 0$, $\text{var}(\xi_{Htk}) = \lambda_{Hk}$, $\text{cov}(\xi_{Htk}, \xi_{Htk'}) = 0$ for any fixed $t \geq 1$, $k \neq k' \geq 1$. We further assume that $\lambda_{Hk} = 0$, for $k > \kappa$, where κ is a positive integer or $+\infty$. The eigenfunction ψ_{Hk} are referred to as functional principal components (FPCs) with FPC scores ξ_{Htk}.

Let Y_{Atj} and Y_{Btj} be the jth observation of the random functions $X_{At}(\cdot)$ and $X_{Bt}(\cdot)$ made at random point u_j for groups A and B.

$$\begin{aligned} Y_{Htj}(u_j) &= X_{Ht}(u_j) + \sigma_H(u_j)\varepsilon_{Htj}, \ j = 1, ..., m, \ t = 1, ..., n_H \\ &= \mu_H(u_j) + \sum_{k=1}^{\kappa} \xi_{Ht,k} \psi_{Hk}(u_j) + \sigma_H(u_j)\varepsilon_{Htj}, \end{aligned} \tag{4.2.2}$$

where ε_{Htj} are the additional iid measurement errors satisfying $E[\varepsilon_{Htj}] = 0$, $E[\varepsilon_{tj}^{(H)}]^2 = 1$, for $H = A, B$.

4.2.2 Spline Estimators

The estimation of the mean functions μ_H and the covariance functions G_H, where $H = A, B$, for each group of random curves can be achieved by the B-spline approximation. Let $\{B_{l,p}, l = 1-p, ..., N\}$ be B-spline basis functions of order p. Let $\mathcal{S}^{(p-2)}$ be polynomial spline space of order p on $\mathcal{U}$. Following Cao, Yang and Todem (2010), the estimator $\hat{\mu}_H(\cdot)$ for the first sample is obtained by

$$\hat{\mu}_H(u) = \underset{g(\cdot)\in\mathcal{S}^{(p-2)}}{\operatorname{argmin}} \sum_{t=1}^{n}\sum_{j=1}^{m} \{Y_{Htj} - g(u_j)\}^2 = \sum_{l=1-p}^{N_\mu} \hat{a}_l B_{l,p}(u),$$

where N_μ is the number of interior knots of the B-spline basis. Due to the nature of the least square, we can write it into matrix form:

$$\hat{\mu}_H(u) = \mathbf{B}_p(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{Y}_H,$$

in which $\mathbf{B}_p = \left(B_{1-p,p}(u), ..., B_{N_\mu,p}(u)\right)$ and $\mathbf{B} = \left(\mathbf{B}_p^T(u_1), ..., \mathbf{B}_p^T(u_m)\right)^T$.

By differentiation, $\mu_H^{(\nu)}(u)$ is the ν-th order derivative of $\mu_H(u)$ with respect to u. For any $\nu = 1, ..., p-2$,

$$\hat{\mu}_H^{(\nu)}(u) = \mathbf{B}_p^{(\nu)}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{Y}_H,$$

in which $\mathbf{B}_p^{(\nu)} = \left(B_{1-p,p}^{(\nu)}(u), ..., B_{N_\mu,p}^{(\nu)}(u)\right)$. By de Boor (2001), for $p > 2$ and $p-2 \leq l \leq N_\mu - 1$,

$$\frac{d}{du}B_{l,p}(u) = (p-1)\left(\frac{B_{l,p-1}(u)}{\theta_{l+p-1}-\theta_l} - \frac{B_{l+1,p-1}(u)}{\theta_{l+p}-\theta_{l+1}}\right).$$

Note that $\mathbf{B}_p^{(\nu)}(u) = \mathbf{B}_{p-\nu}(u)\,\mathbf{D}_{(\nu)}$, where $\mathbf{D}_{(\nu)} = \mathbf{D}_\nu^T\mathbf{D}_{\nu-1}^T\cdots\mathbf{D}_1^T$. We denote

$$\mathbf{D}_s = (p-s)\begin{pmatrix} \frac{-1}{\theta_1-\theta_{1-p+s}} & 0 & 0 & \cdots & 0 & 0 \\ \frac{-1}{\theta_1-\theta_{1-p+s}} & \frac{-1}{\theta_2-\theta_{2-p+s}} & 0 & \cdots & 0 & 0 \\ 0 & \frac{-1}{\theta_2-\theta_{2-p+s}} & \frac{-1}{\theta_3-\theta_{3-p+s}} & \cdots & 0 & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & 0 & \cdots & 0 & \frac{-1}{\theta_{N_\mu+p-s}-\theta_{N_\mu}} \end{pmatrix}$$

as given by Cao et al. (2012).

Let $\hat{R}_{\cdot jj'} = n^{-1}\sum_{t=1}^n \{Y_{Htj} - \hat{\mu}_H(u_j)\}\{Y_{Htj'} - \hat{\mu}_H(u_{j'})\}$, $1 \leq j \neq j' \leq m$. We estimate the covariance function $G_H(u,u')$ using the tensor product spline approach by Cao, Wang, Wang and Yang (2012). The spline estimator of $G(u,u')$ is defined as

$$\hat{G}_H(u,u') = \sum_{l,l'=1-p}^{N_G} \hat{b}_{ll'}B_{l,p}(u)\,B_{l',p}(u'), \tag{4.2.3}$$

where N_G is the number of interior knots used to build the tensor product B-spline basis and the spline coefficients can be computed by the following minimization:

$$\left\{\hat{b}_{ll'}\right\}_{l,l'=1-p}^{N_G} = \underset{R^{N_G+p}\otimes R^{N_G+p}}{\operatorname{argmin}} \sum_{j\neq j'}^{m} \left\{\hat{R}_{\cdot jj'} - \sum_{1-p\leq l,l'\leq N_G} b_{ll'} B_{l,p}\left(u_j\right) B_{l',p}\left(u_{j'}\right)\right\}^2.$$

For these estimators, the theoretical properties are shown in Cao, Todem and Yang (2012) and Cao et al. (2012).

4.2.3 Estimation of the Eigenvalues and Eigenfunctions

The estimates of eigenfunctions and eigenvalues for curves from group H correspond to the solutions of $\hat{\psi}_{Hk}$ and $\hat{\lambda}_{Hk}$ of the eigen equations,

$$\int_{\mathcal{U}} \widehat{G}_H\left(u,u'\right)\hat{\psi}_{Hk}\left(u\right)du = \hat{\lambda}_{Hk}\hat{\psi}_{Hk}\left(u'\right),$$

where the $\hat{\psi}_{Hk}$s are subject to $\int_{\mathcal{U}}\hat{\psi}_{Hk}\left(u\right)^2du = 1$ and $\int_{\mathcal{U}}\hat{\psi}_{Hk}\left(u\right)\hat{\psi}_{Hk'}\left(u\right)du = 0$ for $k' < k$. We define a fine grid $(f_0, f_1, ..., f_{n_f})$ consisting of $n_f + 1$ points equally spaced on $\mathcal{U}$, We estimate the FPC scores by $\hat{\xi}_{Htk} = \sum_{j=1}^{n_f+1}\{Y_{Htj} - \hat{\mu}_H(f_j)\}\hat{\psi}_{Hk}(f_j)(f_j - f_{j-1})$.

For constructing the confidence band, we also need to estimate the function $\Sigma_H\left(u,u'\right)$ through the ν-th derivative of eigenfunctions $\psi_{Hk}^{(\nu)}$ which can be obtained from derivatives of $G_H\left(u,u'\right)$. And $\widehat{G}_H$ and G_H are also asymptotically equivalent at the derivatives by the proof shown in Cao et al. (2012). We also refered to the derivative definition in

Cao et al. (2012): $G_H^{(0,\nu)}(u,u') = \frac{\partial^\nu}{\partial(u')^\nu} G_H(u,u')$ and $\widehat{G}_H^{(0,\nu)}(u,u') = \frac{\partial^\nu}{\partial(u')^\nu}\widehat{G}_H(u,u') = \sum_{l,l'=1-p}^{N_G} \hat{b}_{ll'} B_{l,p}(u) B_{l',p}^{(\nu)}(u')$.

We estimate the ν-th derivative of the k-th eigenfunction $\psi_{1k}^{(\nu)}$ by Liu and Müller (2009) based on the following equation:

$$\hat{\psi}_{Hk}^{(\nu)} = \frac{1}{\hat{\lambda}_{Hk}} \frac{\partial^\nu}{\partial (u')^\nu} \int_0^1 G_H(u,u') \hat{\psi}_{Hk}(u)\, du = \frac{1}{\hat{\lambda}_{Hk}} \int_0^1 \frac{\partial^\nu}{\partial (u')^\nu} G_H(u,u') \hat{\psi}_{Hk}(u)\, du \tag{4.2.4}$$

The derivative of the integral in (4.2.4) can be approximated by the discrete sum:

$$\frac{1}{m} \sum_{j=1}^{m} {G_H}^{(0,\nu)}\left(u_j, u_j'\right) \hat{\psi}_{Hk}(u_j).$$

Therefore, $\Sigma_H(u,u')$ is estimated by

$$\hat{\Sigma}_H(u,u') = \sum_{k=1}^{\kappa} \hat{\lambda}_{Hk} \hat{\psi}_{Hk}^{(\nu)}(u) \hat{\psi}_{Hk}^{(\nu)}(u'). \tag{4.2.5}$$

We also need to choose the number of eigenfunctions that provide a reasonable approximation to the infinite-dimensional process. We can apply a simple criterion in Müller (2009), i.e. $\kappa_H = \text{argmin}_{1\le q\le K}\left\{\sum_{k=1}^{q} \hat{\lambda}_{Hk} / \sum_{k=1}^{K} \hat{\lambda}_{Hk} > 0.95\right\}$. This method works well in the practice.

4.3 Confidence Band

4.3.1 Asymptotic Confidence Band

For $H = A, B$, let $\Sigma_H(u, u') = \sum_{k=1}^{\kappa} \lambda_{Hk} \psi_{Hk}^{(\nu)}(u) \psi_{Hk}^{(\nu)}(u')$ be positive definite function and $\hat{\mu}_H^{(\nu)}$ be the spline estimates for the ν-th derivative of group mean function $\mu_H^{(\nu)}$.

For any $r \in (0, 1]$, we denote $C^{q,r}[0, 1]$ as the space of Hölder continuous functions on $[0, 1]$, $C^{q,r}[0, 1] = \left\{\eta : \|\eta\|_{q,r} = \sup_{t \neq s, t, s \in [0,1]} \left|\eta^{(q)}(t) - \eta^{(q)}(s)\right| / |t - s|^r < +\infty\right\}$. The technical assumptions are in the following:

(A1) *The regression functions* $\mu_A,\ \mu_B \in C^{p-1,1}(\mathcal{U})$;

(A2) *The standard deviation functions* $\sigma_A, \sigma_B \in C^{0,\delta}(\mathcal{U})$ *for some* $\delta \in (0, 1]$*;*

(A3) *The number of observations for each trajectory* $m \gg n^{\theta}$ *for some* $\theta > \frac{1+2\nu}{2(p-\nu)}$*; the number of interior knots satisfies* $n^{\frac{1}{2(p-\nu)}} \ll N_\mu \ll (m/\log(n))^{\frac{1}{1+2\nu}}$, $n^{\frac{1}{2p}} \ll N_G \ll n^{\frac{1}{2+2\nu}}$;

(A4) *There exists a constant* $C > 0$ *such that* $\Sigma_A(u, u) > C$, $\Sigma_B(u, u) > C$, *for any* $u \in \mathcal{U}$*;*

(A5) *For* $k \in \{1, \ldots, \kappa\}$, $\nu = 0, 1, \ldots, p-2$, $\psi_{Hk}^{(\nu)}(u) \in C^{0,\delta}(\mathcal{U})$, *for some* $\delta \in (0, 1]$, $\sum_{k=1}^{\kappa} \sqrt{\lambda_{Hk}} \|\psi_{Hk}^{(\nu)}\|_\infty < \infty$; *and for a sequence* $\{\kappa_n\}_{n=1}^{\infty}$ *of increasing integers with* $\lim_{n\to\infty} \kappa_n = \kappa$, $N_\mu^{-\delta} \sum_{k=1}^{\kappa_n} \sqrt{\lambda_{Hk}} \|\psi_{Hk}^{(\nu)}\|_{0,\delta} = o(1)$*;*

(A6) For $H = A, B$, $k \geq 1$, $\{\xi_{Htk}, t \geq 1\}$ is a strictly stationary double-side ϕ-mixing random variables, that is

$$\phi(n) = \sup_k \phi(\mathcal{F}_{-\infty}^k, \mathcal{F}_{k+n}^\infty) \to 0, \ n \to \infty,$$

where $\mathcal{F}_n^m = \sigma(\xi_t, n \leq t \leq m)$ and $\phi(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A)>0} |P(B|A) - P(B)|$. The decay rate $\phi(n) = O(n^{-b})$ for some $b > 2$ and $E|\xi_{Htk}|^{2+\delta_1}$ *for some positive constant* δ_1.

(A7) For $H = A, B$, $1 \leq j \leq m$, $\{\varepsilon_{Htj}, t \geq 1\}$ is a white noise sequence and $E|\varepsilon_{Htj}|^{2+\delta_2}$ *for some positive constant* δ_2.

Let $\hat{r} = n_A/n_B$ and $V(u, u') = \Sigma_A(u, u') + r\Sigma_B(u, u')$, where $r = \lim n_A \to \infty \hat{r}$. Denote $W(u)$, $u \in \mathcal{U}$ a standardized Gaussian process such that $EW(u) = 0$, $EW^2(u) = 1$ with covariance

$$E[W(u) W(u')] = V^{-1/2}(u, u) V(u, u') V^{-1/2}(u', u').$$

Denoted by Q_α the $(1-\alpha)$-th quantitle of the absolute maxiam deviation of $W(u)$, $u \in \mathcal{U}$.

Theorem 4.1. *Under Assumptions (A1)-(A6), for any* $\alpha \in (0, 1)$, *as* $n_A \to \infty$, $\hat{r} \to r > 0$,

$$P\left\{\sup_{u \in \mathcal{U}} \frac{n_A^{1/2} \left|\left(\hat{\mu}_A^{(\nu)} - \hat{\mu}_B^{(\nu)}\right)(u) - \left(\mu_A^{(\nu)} - \mu_B^{(\nu)}\right)(u)\right|}{\{\Sigma_A(u, u) + r\Sigma_B(u, u)\}^{1/2}} \leq Q_\alpha\right\} \to 1 - \alpha.$$

4.3.2 Implementation

To construct the confidence bands, we need to estimate Q_α. We simulate:

$$\widehat{W}_t = \widehat{V}^{-1/2}(u,u)\sum_{k=1}^{\kappa}\{\sqrt{\lambda_{Ak}}Z_{Ak,t}\psi_{Ak}^{(\nu)}(u) + \sqrt{r\lambda_{Bk}}Z_{Bk,t}\psi_{Bk}^{(\nu)}(u)\}, u \in \mathcal{U}.$$

Q_α can be estimated by $100\,(1-\alpha)$-th percentile of $\left\{\sup_{u\in\mathcal{U}}|\widehat{W}_t(u)|\right\}_{t=1}^{5000}$. Therefore, in application we can propose the band as

$$\left(\hat{\mu}_A^{(\nu)} - \hat{\mu}_B^{(\nu)}\right)(u) \pm n_A^{-1/2}\widehat{V}(u,u)^{1/2}\,\hat{Q}_\alpha, \tag{4.3.1}$$

as confidence band for $\mu_A^{(\nu)} - \mu_B^{(\nu)}$.

4.4 Simulation

In order to illustrate the performance of the confidence band in (4.3.1), we conduct the simulation study with the data generated from the following model:

$$Y_{tj} = \mu(j/N) + \sum_{k=1}^{\kappa}\xi_{tk}\psi_k(j/N) + \varepsilon_{tj},$$

where noise sequence ε_{tj}'s are i.i.d. r.v.'s $\sim N(0, 0.5^2)$. We do the pairwise comparison of groups of curves with different mean functions and different first order derivatives of mean functions. We generate three groups of curves from the following specification:

Group A: $\mu(u) = 2\sin(\pi(u-.5)) + 2u$, $\psi_1(u) = -\sqrt{2}\cos(2\pi(u-.5))$,

$\psi_2(u) = \sqrt{2}\sin(4\pi(u-.5))$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\kappa = 2$;

Group B: $\mu(u) = 2\cos(\pi(u-.5)) + 2$, $\psi_1(u) = -\sqrt{2}\cos(2\pi(u-.5))$,

$\psi_2(u) = \sqrt{2}\sin(4\pi(u-.5))$, $\lambda_1 = 2$, $\lambda_2 = 1$, $\kappa = 2$;

Group C: $\mu(u) = 2\cos(\pi(u-.5)) + 2$, $\psi_1(u) = -\cos(\pi u/10)/\sqrt{5}$,

$\psi_2(u) = \sin(\pi u/10)/\sqrt{5}$, $\lambda_1 = 4$, $\lambda_2 = 1$, $\kappa = 2$.

The scores ξ_{t1} and ξ_{t2} are simulated from an AR(2) process and an ARMA(1,1) process, respectively,

$$\xi_{t1} = 0.75\xi_{t-1,1} + 0.20\xi_{t-2,1} + \epsilon_t, \quad \xi_{t2} = 0.80\xi_{t-1,2} + 0.30\epsilon_{t-1} + \epsilon_t,$$

where ϵ_t's and ϵ_{t-1}'s are i.i.d. r.v.'s $\sim N(0,1)$. We rescaled ξ_{t1} and ξ_{t2} such that $\text{var}(\xi_{t1}) = \lambda_1$ and $\text{var}(\xi_{t2}) = \lambda_2$. Note that the curves in groups A and B have different mean functions but share the same covariance functions; the curves in groups A and C have different mean functions and different covariance functions; the curves in groups B and C have the same mean function but different covariance functions. We construct the confidence bands for the difference of derivative functions of mean functions in three scenarios: 1) Group A vs. Group B; 2) Group A vs. Group C; 3) Group B vs. Group C. Here, we consider two confidence levels: 0.99 and 0.95. The number of the trajectories n is set to be 30, 50, 100, 200, 400 and 1000. For each n, we try different numbers of observations on the curve. We run 500 replications for each simulation. We apply generalized cross-validation to select the number of knots N_μ (from 2 to 20) for estimating

the mean function. According to Cao et al. (2012b), the number of knots for smoothing the covariance function can be determined by the formula: $N_G = \left[cn^{1/2p}\log(\log(n))\right]$, where c is a constant (here we use $c = 1$).

For each replication, the true function is the difference of the first derivative functions for mean curves of two groups: $\mu_A^{(1)} - \mu_B^{(1)}$, $\mu_A^{(1)} - \mu_C^{(1)}$ and $\mu_B^{(1)} - \mu_C^{(1)}$. We repeat the test for 500 times to check if true functions are covered by the confidence bands at 200 equally spaced point on $\mathcal{U}$. For the first scenario (refer to Table 4.1), we see that coverage rate of estimated 95% bands exceed the nominal levels in some cases. This occurs mostly because that the curves in two groups share the same covariance structure. This rarely happens in the real life. For the second and third scenarios, as n and N increase, the coverage rates approach to the nominal levels. It is noted that when n is large ($n = 1000$), the coverage rates appear to be very stable.

(Insert Table 4.1 about here.)

4.5 Application

We apply the proposed methodology to the temperature data collected from the weather station in Athens, Georgia. The data set contains 55 temperature curves from October to September of the next year during 1948/10-2003/9, where on each curve we have 365 daily average temperature observations. Our proposed approach helps the climatologists to examine if temperature transition throughout the year in Athens,GA differs signifi-

cantly in terms of some of the major atmospheric pressure oscillations that take place over the globe. The most notable and most powerful of these is the "El Niño". El Niño refers to a phase during which a region in the central Pacific ocean exhibits unusually warm waters. If the waters in this region are unusually cold, that is called "La Niña". If the waters are in neither phase, this is referred to as Neutral. Associated with the temperature fluctuations are oscillations in the atmospheric pressure above this region of the Pacific. The atmospheric pressure phenomenon is known as the Southern Oscillation. Thus, ENSO is short for the El Niño - Southern Oscillation, and is the technical name of the phenomenon. The ENSO is observed every October, and is classified as La Niña (Cold Phase), Neutral or El Niño (Warm Phase). The ENSO phases back to 1868 can be checked at: http://www.coaps.fsu.edu/jma.shtml. Therefore, we define the year in our data set as "ENSO year". For example, the ENSO year 1948 starts October 1948 and ends September 1949. We can divide 55 yearly temperature curves into three groups according to three ENSO Phases. There are 14 curves in Cold Phase, 28 curves in Neutral Phase and 13 curves in Warm Phase. Figure 4.1 shows a three-dimensional plot of the temperature data.

Before applying the proposed method to the temperature data set, we first use the B-Spline method in Shao and Yang (2011) to estimate the trend of the temperature curves. Then, we remove this temporal dynamic, and work with the residual data. Figure 4.1 shows the estimated trend (middle graph) in terms of time and the detrended climate data (lower graph).

In our application, we denote X_H by the random temperature curve in H ENSO Phase where $H = A, B, C$ (A - Cold, B - Neutral, C - Warm). Let $u_j = j/365$ be the daily index, where $j = 1, ..., 365$, and let μ_H be the mean temperature function of X_H for H ENSO Phase ENSO Phase.

$$X_{Ht}(u_j) = \mu_H(u_j) + \sum_{k=1}^{\kappa} \xi_{Ht,k}\psi_{Hk}(u_j), \ t = 1, ..., n_H$$

Denote by Y_{Htj} the noisy observation on the j-th day of the t-th year in the H ENSO Phase, which contains the additional noise ε_{Htj}; Then we can model

$$Y_{Htj}(u_j) = \mu_H(u_j) + \sum_{k=1}^{\kappa} \xi_{Ht,k}\psi_{Hk}(u_j) + \sigma_H(u_j)\varepsilon_{Htj},$$

We first construct the 99% confidence band for the mean curve and its first order derivative function of 55 temperature curves in Fig 4.2 . For the mean curve, we see that temperature looks like a sine curve which is consistent with our common sense. For derivative curve, the temperature changes sharply from December to March, exhibits a flat shape until May, and drops sharply after June. Fig 4.3 shows the mean functions for the three ENSO Phase. We can see that mean curves for warm phase and neutral phase are very similar. There are obvious difference between Dec and Mar for the three mean functions. Fig 4.4 shows the functional derivatives for mean curves of the three groups. We can see the similar pattern in Fig 4.3 that derivatives for warm phase and neutral phase are very close, but apart from the one for cold phase. In order to see if the temperature transition differs significantly across ENSO phases, we develop

three pairwise hypothesis tests based on two samples, for example: $H_0 : \mu_A^{(1)} = \mu_B^{(1)}$ *v.s.* $H_a : \mu_A^{(1)} \neq \mu_B^{(1)}$, where $\mu_A^{(1)}$ and $\mu_B^{(1)}$ are the first order derivatives of the mean temperature curves observed in cold phase and neutral phase. By repeating this procedure, we can conduct the comparison between group A and C: $H_0 : \mu_A^{(1)} = \mu_C^{(1)}$ *vs.* $H_a : \mu_A^{(1)} \neq \mu_C^{(1)}$ and the comparison between group B and C: $H_0 : \mu_B^{(1)} = \mu_C^{(1)}$ *vs.* $H_a : \mu_B^{(1)} \neq \mu_C^{(1)}$. Fig 4.5 shows the confidence interval at confidence levels 0.99 and 0.95 for the difference of the first order derivative functions for two groups. For example, the upper graph displays the 0.99 and 0.95 confidence intervals for $\mu_A^{(1)} - \mu_B^{(1)}$, with the center dashed-dotted line representing the spline estimator $\hat{\mu}_A^{(1)} - \hat{\mu}_B^{(1)}$ and a solid line representing zero line. Since the zero lines cut across the confidence bounds in all three cases, this indicates that temperature transition throughout the year in Athens differs significantly under the influence of ENSO.

4.6 Appendix

We cite a strong approximation result for a sequence of stationary φ-mixing random variables from Lin and Li (2008), which plays an important role through our proof shown below.

Lemma 4.1. *[Lin and Li (2008), Lemma 2.1] Let* $\{\xi_t, t \geq 1\}$ *be a sequence of strictly* φ*-mixing random variables with* $E(\xi_t) = 0$ *and* $Var(\xi_t) = 1$ *for each* $t \geq 1$*. Suppose condition (A5) holds, then we can redefine* $\xi_t, t \geq 1$ *on a richer probability space together*

with a sequence of independent N(0,1) random variables $\{Z_t, t \geq 1\}$ *such that*

$$\max_{1\leq k\leq n}\left|\sum_{t=1}^{k}\xi_t - \sum_{t=1}^{k} Z_t\right| = o\left(n^{\frac{1}{2+\delta_1}}(\log n)^{1+\varepsilon+\frac{1+\lambda}{2+\delta_1}}\right) \quad a.s. \tag{4.6.1}$$

for any $\varepsilon > 0$, *where* $\lambda = \frac{2\log 3}{\log\theta - 1}$ *and* $\theta = 1 - \frac{2(b-1)}{b(2+\delta_1)}$.

4.6.1 Proof of Theorem 4.1

Proof. Following Cao et al. (2012), denote the signal, noise and eigenfunction vectors by $\boldsymbol{\mu}_H = (\mu_H(u_1), \cdots, \mu_H(u_m))^T$, $\mathbf{e}_H = (\sigma_H(u_1)\bar{\varepsilon}_{H\cdot 1}, \cdots, \sigma_H(u_m)\bar{\varepsilon}_{H\cdot m})^T$, and $\boldsymbol{\psi}_{Hk} = (\psi_{Hk}(u_1), \cdots, \psi_{Hk}(u_m))^T$, where $\bar{\varepsilon}_{H\cdot j} = n_H^{-1}\sum_{i=1}^{n_H}\varepsilon_{Hij}$, $H = A, B$. We can decompose the spline estimator $\hat{\mu}_H(\cdot)$ into three terms:

$$\hat{\mu}_H^{(\nu)}(u) = \tilde{\mu}_H^{(\nu)}(u) + \tilde{e}_H^{(\nu)}(u) + \tilde{\xi}_H^{(\nu)}(u), \tag{4.6.2}$$

where $\tilde{\mu}_H^{(\nu)}(u) = \Omega^{(\nu)}(u)\boldsymbol{\mu}_{\mathbf{H}}$, $\tilde{e}_H^{(\nu)}(u) = \Omega^{(\nu)}(u)\mathbf{e}$ and $\tilde{\xi}_H^{(\nu)}(u) = \sum_{k=1}^{\kappa}\bar{\xi}_{H\cdot k}\Omega^{(\nu)}(u)\boldsymbol{\psi}_{Hk}$ with $\Omega^{(\nu)}(u) = \mathbf{B}_p^{(\nu)}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$, and $\bar{\xi}_{H\cdot k} = n_H^{-1}\sum_{t=1}^{n_H}\xi_{Hik}$, $1 \leq k \leq \kappa$, $H = A, B$.

Therefore, asymptotic error $\left(\hat{\mu}_A^{(\nu)} - \hat{\mu}_B^{(\nu)}\right) - \left(\mu_A^{(\nu)} - \mu_B^{(\nu)}\right)$ can be decomposed into three components:

$$\left(\tilde{\mu}_A^{(\nu)} - \tilde{\mu}_B^{(\nu)} - \mu_A^{(\nu)} + \mu_B^{(\nu)}\right) + \left(\tilde{e}_A^{(\nu)} - \tilde{e}_B^{(\nu)}\right) + \left(\tilde{\xi}_A^{(\nu)} - \tilde{\xi}_B^{(\nu)}\right).$$

The proof for $\sqrt{n}$ asymptotic efficiency of first two components are trivial by using Proposition 1 in Cao et al (2012). Here, we focus on the third component.

According to Lemma 4.1, one can find iid $Z_{Hik,\xi} \sim N(0,1)$, $i = 1, ..., n_H$ such that $\max_{1\leq k\leq \kappa} \left|\bar{\xi}_{H\cdot k} - \sqrt{\lambda_{Hk}}\bar{Z}_{H\cdot k,\xi}\right| = O_{a.s.}\left(n_H^{\tau-1}\right)$, where $\tau \in (1/(2+\delta_1), 1/2)$, and $\bar{Z}_{H\cdot k,\xi} = n_H^{-1}\sum_{i=1}^{n_H} Z_{Hik,\xi}$. Likewise, for the white noise sequence $\{\varepsilon_{Hij}, i \geq 1\}$, one can also find iid $Z_{Hik,\varepsilon} \sim N(0,1)$, $i = 1, ..., n_H$ such that $\max_{1\leq k\leq \kappa} \left|\bar{\varepsilon}_{H\cdot j} - \sqrt{\lambda_{Hj}}\bar{Z}_{H\cdot j,\varepsilon}\right| = O_{a.s.}\left(n_H^{\beta-1}\right)$, where $\beta \in (1/(2+\delta_2), 1/2)$.

Next, we define a stochastic process

$$\begin{aligned}\widetilde{W}(u) &= n_A^{1/2}\left[\sum\nolimits_{k=1}^{\kappa}\left\{\lambda_{Ak}\left(\psi_{Ak}^{(\nu)}(u)\right)^2 + r\lambda_{Bk}\left(\psi_{Bk}^{(\nu)}(u)\right)^2\right\}\right]^{-1/2}\sum_{k=1}^{\kappa}\widetilde{W}_k(u) \\ &= n_A^{1/2}V(u,u)^{-1/2}\sum_{k=1}^{\kappa}\widetilde{W}_k(u),\end{aligned}$$

where $\widetilde{W}_k(u) = \sqrt{\lambda_{Ak}}\bar{Z}_{A\cdot k}\psi_{Ak}^{(\nu)}(u) - \sqrt{\lambda_{Bk}}\bar{Z}_{B\cdot k}\psi_{Bk}^{(\nu)}(u)$, $k = 1, ..., \kappa$. It is clear that, for any $u \in \mathcal{U}$, $\widetilde{W}(u)$ is Gaussian with mean 0 and variance 1, and the covariance

$$E[\widetilde{W}(u)\widetilde{W}(u')] = V(u,u)^{-1/2}V(u,u')V(u',u')^{-1/2}.$$

That is, the distribution of $\widetilde{W}(u)$, $u \in \mathcal{U}$ and the distribution of $W(u), u \in \mathcal{U}$ are identical. We denote $\widetilde{\psi}_{Hk}^{(\nu)} = \Omega^{(\nu)}(u)\boldsymbol{\psi}_{Hk}$, and according to (A.5) in Cao, et al. (2012),

$\sup_{u\in\mathcal{U}}|\widetilde{\psi}_{Hk}^{(\nu)}-\psi_{Hk}^{(\nu)}|\leq CN_{\mu}^{-(p-v)}$ for some positive constant C. Note that

$$\begin{aligned}
&\sup_{u\in\mathcal{U}}\left|\widetilde{W}(u)-n_A^{1/2}V(u,u)^{-1/2}\left\{\tilde{\xi}_A^{(\nu)}(u)-\tilde{\xi}_B^{(\nu)}(u)\right\}\right| \\
= \;& n_A^{1/2}\sup_{u\in\mathcal{U}}V(u,u)^{-1/2}\left|\sum_{k=1}^{\kappa}\widetilde{W}_k(u)-\left\{\tilde{\xi}_A^{(\nu)}(u)-\tilde{\xi}_B^{(\nu)}(u)\right\}\right| \\
\leq \;& n_A^{1/2}\sup_{u\in\mathcal{U}}V(u,u)^{-1/2} \\
& \times\sum_{k=1}^{\kappa}\left\{\left|\sqrt{\lambda_{Ak}}\bar{Z}_{A\cdot k}-\bar{\xi}_{A\cdot k}\right|\left|\psi_{Ak}^{(\nu)}(u)\right|+\left|\bar{\xi}_{A\cdot k}\right|\left|\psi_{Ak}^{(\nu)}(u)-\widetilde{\psi}_{Ak}^{(\nu)}(u)\right|\right. \\
& \left.+\left|\sqrt{\lambda_{Bk}}\bar{Z}_{B\cdot k}-\bar{\xi}_{B\cdot k}\right|\left|\psi_{Bk}^{(\nu)}(u)\right|+\left|\bar{\xi}_{B\cdot k}\right|\left|\psi_{Bk}^{(\nu)}(u)-\widetilde{\psi}_{Bk}^{(\nu)}(u)\right|\right\} \\
= \;& O_P\left(n_A^{\tau-1/2}+n_B^{\tau-1/2}r^{1/2}+2N_{\mu}^{-(p-\nu)}\right)=o_P(1).
\end{aligned}$$

Theorem 4.1 follows directly. ∎

4.7 References

Cao, G., Yang, L. and Todem, D. (2012a) Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics* **24**, 359-377.

Cao, G., Wang, J., Wang, L. and Todem, D. (2012b). Spline confidence bands for functional derivatives. *Journal of Statistical Planning and Inference,* forthcoming.

Degras, D. A. (2011) Simultaneous confidence bands for nonparametric regression with functional data. Statistica Sinica **21**, 1735-1765.

Hall, P., Müller, H. G. and Yao, F. (2009) Estimation of functional derivatives. *Annals of Statistics* **37**, 3307-3329.

Lin, Z. and Li, D. (2008) Strong approximation for moving average processes under dependence assumptions. *Acta Mathematica Scientia* **28**, 217-224.

Liu, B. and Müller, H. G. (2009) Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association* **104**, 704-717.

Ma, S., Yang, L. and Carroll, R. J. (2012) A simultaneous confidence band for spase longitudinal regression. *Statistica Sinica* **22**, 95-122.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. Journal of the Royal Statistical Society B, **53**, 233-243.

Shao, Q. and Yang, L. (2011) Autoregressive coefficient estimation in nonparametric analysis. *Journal of Time Series Analysis* **32**, 587-597.

Table 4.1: Coverage rates of spline confidence bands for the difference of first derivatives of mean functions

n	N	Group A-B		Group A-C		Group B-C	
		95%	99%	95%	99%	95%	99%
30	30	0.892	0.932	0.862	0.898	0.860	0.896
	60	0.972	0.984	0.936	0.960	0.940	0.958
50	50	0.896	0.924	0.866	0.888	0.858	0.888
	100	0.960	0.978	0.936	0.952	0.940	0.954
100	100	0.962	0.980	0.924	0.952	0.922	0.952
	200	0.962	0.976	0.928	0.954	0.934	0.954
200	200	0.952	0.974	0.938	0.944	0.934	0.944
	400	0.994	0.994	0.980	0.990	0.984	0.990
400	200	0.972	0.978	0.938	0.962	0.940	0.964
	400	0.982	0.992	0.976	0.982	0.968	0.980
1000	200	0.978	0.986	0.936	0.962	0.938	0.956
	400	0.978	0.988	0.956	0.972	0.956	0.970

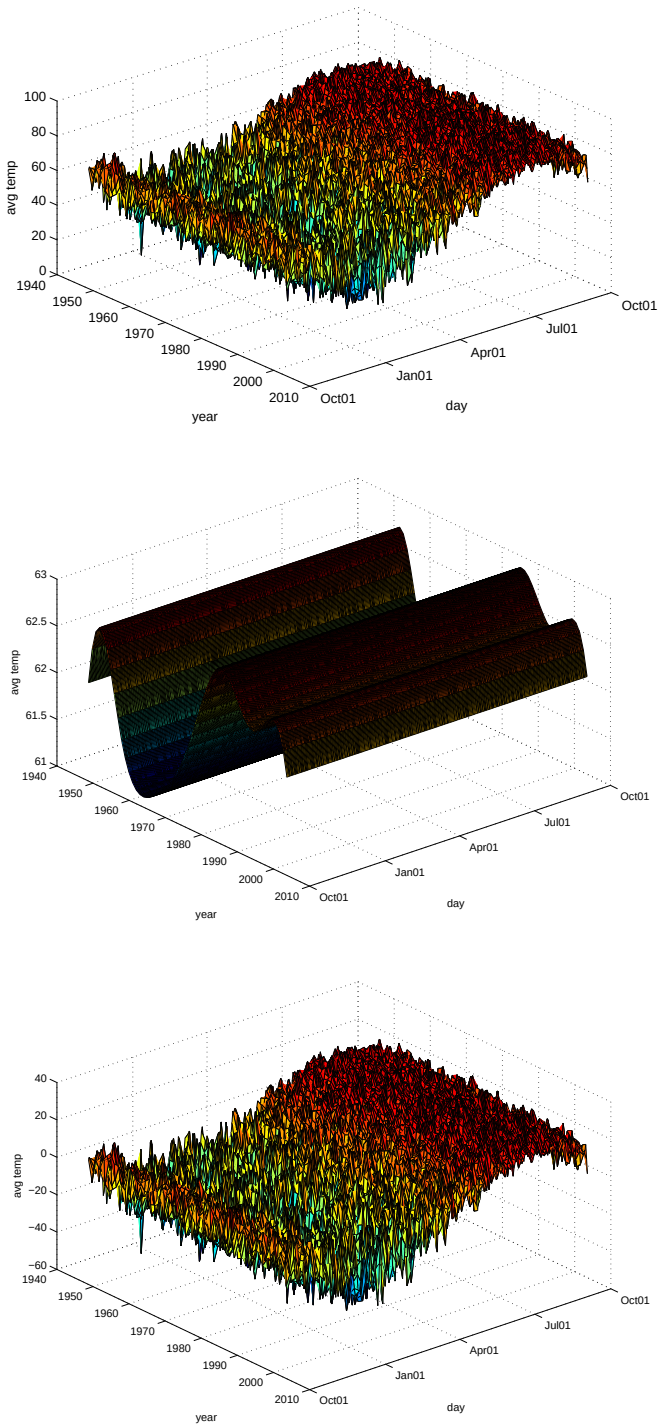


Figure 4.1: 3-D display of Temperature Data for 1948.10-2003.9. The data set contains 55 temperature curves from October to September in next year during 1948/10-2003/9, where on each curve we have 365 daily average temperature observation

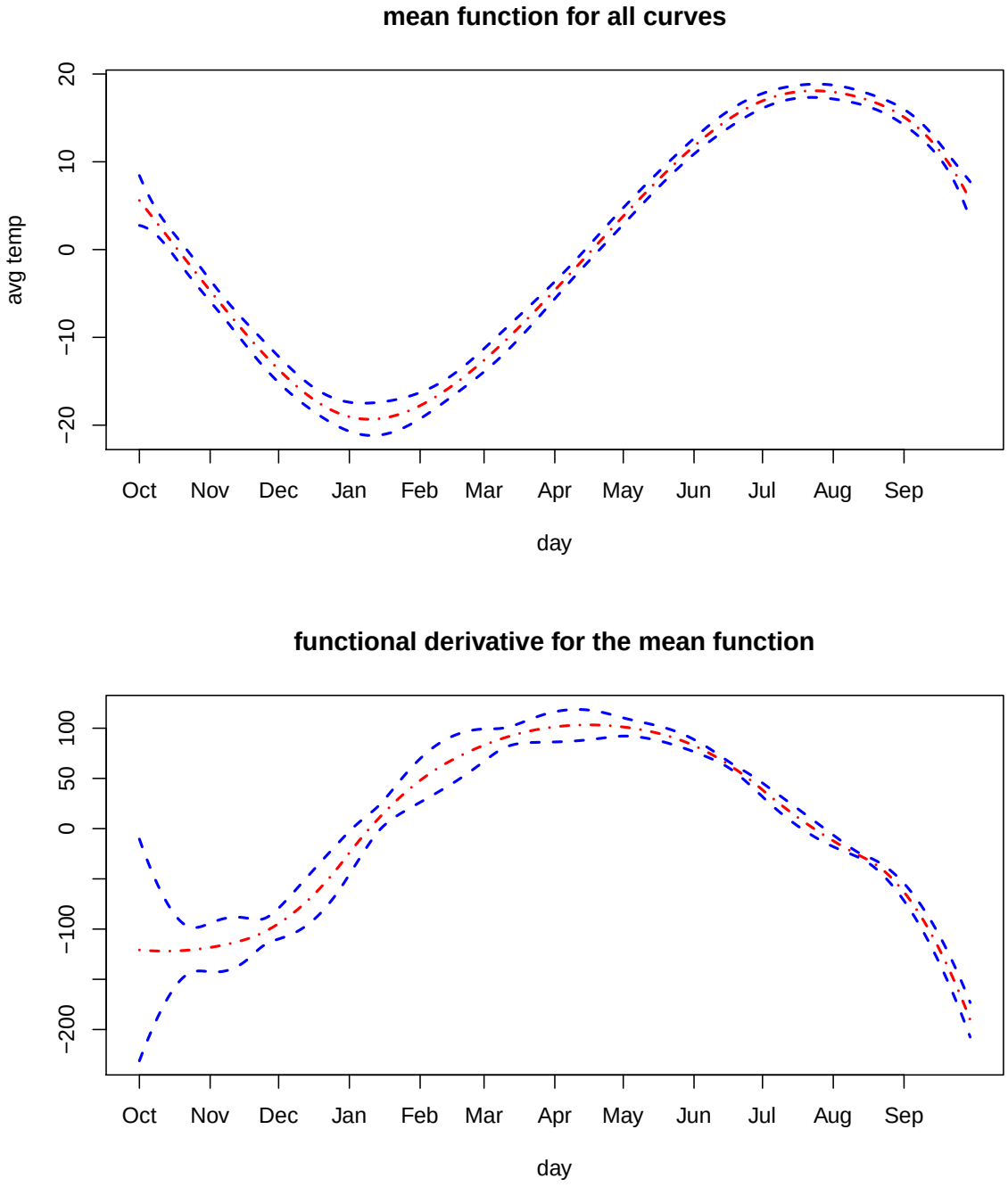


Figure 4.2: Plots of the cubic spline estimators (dotted-dashed line) and 99% confidence bands (upper and lower dashed lines) of the mean function and its first order derivatives. The labels on x-axis denote the first day of the month.

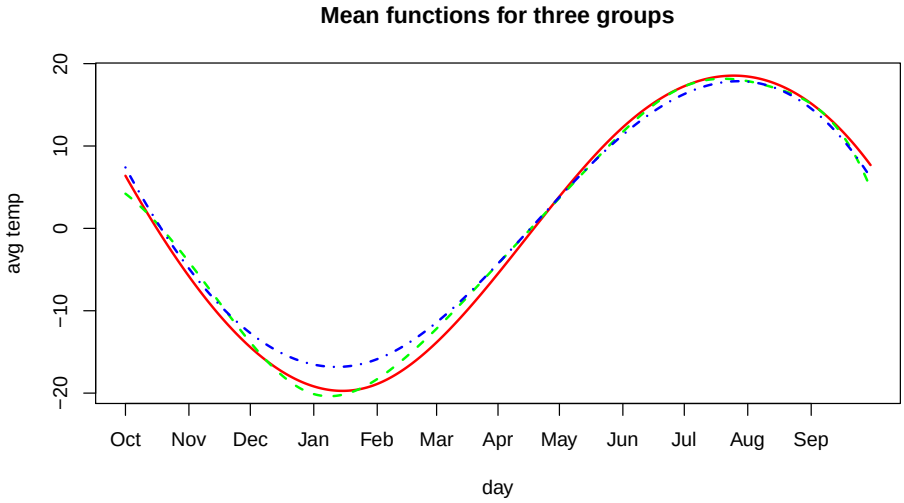


Figure 4.3: Plots of the mean functions for Group A (dotted-dashed line), B (dashed line) and C (solid line). The labels on x-axis denote the first day of the month.

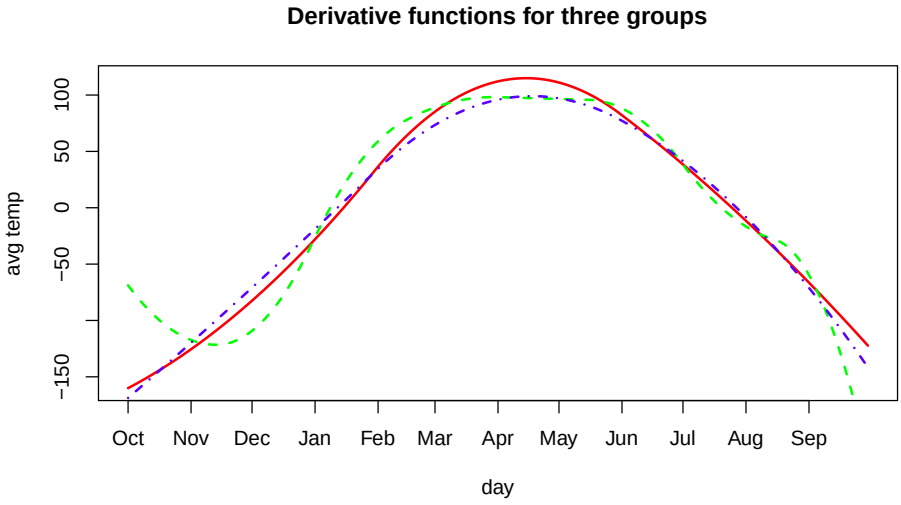


Figure 4.4: Plots of the functional derivative of mean functions for Group A (dotted-dashed line), B (dashed line) and C (solid line). The labels on x-axis denote the first day of the month.

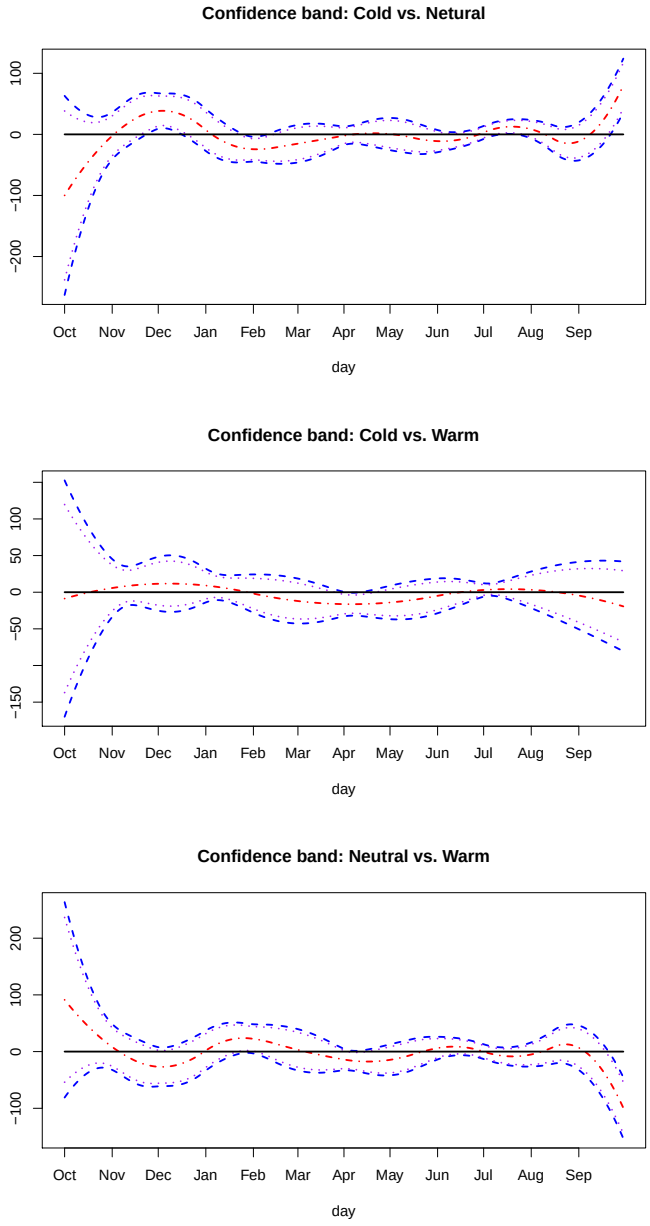


Figure 4.5: Plots of the cubic spline estimators (dotted-dashed line), 99% confidence bands (upper and lower dashed lines) and 95% confidence bands (upper and lower dotted lines) of $\mu_A^{(1)} - \mu_B^{(1)}$, $\mu_A^{(1)} - \mu_C^{(1)}$ and $\mu_B^{(1)} - \mu_C^{(1)}$ for three groups of temperature curves.

Chapter 5

Conclusion

In this dissertation, novel non/semi-parametric models have been developed respectively for financial volatility, term structure of Treasury Bond yield curves, and temperature data. The proposed spline estimators for semiparametric GARCH models have good theoretical properties with fast computation. We also constructed the simultaneous confidence bands for the derived news impact curve. We developed the methodology for modeling and forecasting the weakly dependent functional data, which has superior performance in the real data application. The simultaneous confidence bands for two-group comparison of functional derivatives in functional time series were derived to study the temperature transition problem. There are still many open questions in time series with complex features; for example, I would like to investigate the spatio-temporal correlation structure for multiple locations of climate time series data. This will provide the climatologists an effective tool to evaluate climate change over a region. Other closely related

topics in which I am interested for my future research are model selection for functional regression and inference for covariance function in weakly dependent functional data. I am also interested in extending the traditional time series models, such as AR, MA and ARMA models to the functional framework.

Bibliography

[1] Bathia, N., Q. Yao, and F. Zieglemann (2010). Identifying the finite dimensionality of curve time series. *Annals of Statistics* **38**, 3352-3386.

[2] Bosq, D. (1998). Nonparametric Statistics for Stochastic Processes. New York: Springer-Verlag.

[3] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307-327.

[4] Bollerslev, T. and Engle, R. (1993). Common persistence in conditional variances. *Econometrica* **61**, 167-186.

[5] Bühlmann, P. and McNeil, A. J. (2002). An algorithm for nonparametric GARCH modelling. *Computational Statistics & Data Analysis* **40**, 665-683.

[6] Cai, T. and Hall, P. (2005). Prediction in functional linear regression. *Annals of Statistics* **34**, 2159-2179.

[7] Cao, G., Yang, L. and Todem, D. (2012a). Simultaneous inference for the mean function of dense functional data. *Journal of Nonparametric Statistics,* in press.

[8] Cao, G., Wang, J., Wang, L. and Todem, D. (2012b). Spline confidence bands for functional derivatives. *Journal of Statistical Planning and Inference,* forthcoming.

[9] Cao, G., Wang, L., Li, Y. and Yang, L. (2011). Spline confidence envelopes for covariance function in dense functional/longitudinal data. *Manuscript.*

[10] Carroll, R., Hardle, W. and Mammen, E. (2002). Estimation in an additive model when the components are linked parametrically. *Econ. Theory* **18**, 886-912.

[11] de Boor, C. (2001). A Practical Guide to Splines. New York: Springer-Verlag.

[12] Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. Statistica Sinica **21**, 1735- 1765.

[13] Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* **130**, 337-364.

[14] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1007.

[15] Engle, R. F. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* **48**, 1749-1778.

[16] Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal* **21**, 55-59.

[17] Fan, J. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. Chapman and Hall: London.

[18] Fan, J. and Yao, Q. (2003). Nonlinear Time Series: Nonparametric and Parametric Methods. New York: Springer.

[19] Ferraty, F. and Vieu, P. (2006). Nonparametric Functional Data Analysis: Theory and Practice. Springer Series in Statistics, Springer: Berlin.

[20] Glosten, L. R., Jaganathan, R. and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779-1801.

[21] Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics* **34**, 1493–1517.

[22] Hall, P., Müller, H. G. and Yao, F. (2009). Estimation of functional derivatives. *Annals of Statistics* **37**, 3307-3329.

[23] Härdle, W. and Tsybakov, A. B. (1997). Locally polynomial estimators of the volatility Function. *Journal of Econometrics* **81**, 223-242.

[24] Härdle, W., Tsybakov, A. B. and Yang, L. (1998). Nonparametric vector autoregression. *Journal of Statistical Planning and Inference* **68**, 221-245.

[25] Hengartner, N. W. and Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis* **95**, 246-272.

[26] Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics* **31**, 1600-1635.

[27] Huang, J. Z. and Yang, L. (2004). Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society: Series B* **66**, 463-477.

[28] Hafner, C. (1998). Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility. Heidelberg: Physica-Verlag.

[29] Hastie, T. J. and Tibshirani, R. J. (1990). Generalized Additive Models. London: Chapman and Hall.

[30] Hörmann, S. and Kokoszka, P.(2010). Weakly dependent functional data. *The Annals of Statistics* **38**, 1845-1884.

[31] Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis* **51**, 4942-4956.

[32] James, G. M., Hastie, T. and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika* **87**, 587-602.

[33] Koopman, S. J., Mallee, M. and van der Wel, M. (2005). Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters. *Journal of Business & Economic Statistics* **28**, 329-343.

[34] Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Annals of Statistics* **38**, 3321-3351.

[35] Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.

[36] Linton, O. B. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* **73**, 771-836.

[37] Liu, B. and Müller, H. G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to online auction dynamics. *Journal of the American Statistical Association* **104**, 704-717.

[38] Masry, E. and Tjøstheim, D. (1995). Nonparametric estimation and identification of nonlinear ARCH time series: strong convergence and asymptotic normality. *Econometric Theory* **11**, 258-289.

[39] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68**, 179-199.

[40] Müller, H. G. (2009). Functional modeling of longitudinal data. *In: Longitudinal Data Analysis (Handbooks of Modern Statistical Methods)*, Ed. Fitzmaurice, G., Davidian, M., Verbeke, G., Molenberghs, G., Wiley, New York, 223-252.

[41] Nelson, D. (1991). Conditional heteroscedasticity in asset pricing: a new approach. *Econometrica* **59**, 347-370.

[42] Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Annals of Statistics* **25**, 186-211.

[43] Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* **45**, 267-290.

[44] Park, B., Mammen, E., Hardle, W. and Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, **104**, 284-298.

[45] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis.* Second Edition. Springer Series in Statistics. Springer: New York.

[46] Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society B*, **53**, 233-243.

[47] Shao, Q. and Yang, L. (2011) Autoregressive coefficient estimation in nonparametric analysis. *Journal of Time Series Analysis* **32**, 587-597.

[48] Shen, H.(2009). On modeling and forecasting time series of curves. *Technometrics* **51**, 227-238.

[49] Song, Q. and Yang, L. (2009a). Spline confidence bands for variance function. *Journal of Nonparametric Statistics* **21**, 589-609.

[50] Song, Q. and Yang, L. (2009b). Simultaneous confidence band for nonlinear additive autoregression model via spline-backfitted spline smoothing. *Journal of Multivariate Analysis*, in press.

[51] Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics* **13**, 689-705.

[52] Tsay, R. S. (2005). Analysis of financial time series. Wiley, New York.

[53] Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* **35**, 2474-2503.

[54] Yang, L. (2002). Direct estimation in an additive model when the components are proportional. *Statistica Sinica* **12**, 801-821.

[55] Yang, L. (2006). Semiparametric GARCH model and foreign exchange volatility. *Journal of Econometrics* **130**, 365-384.

[56] Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *Journal of Time Series Analysis* **20**, 597-604.

[57] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B* **68**, 3-25.

[58] Yao, F., Müller, H. G. and Wang, J. L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics* **33**, 2873-2903.

[59] Yao, F., Müller, H. G. and Wang, J. L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577-590.

[60] Zhao, X., Marron, J. S. and Wells, M. T. (2004). The functional data analysis view of longitudinal data. *Statistica Sinica* **14**, 789-808.

[61] Zhou, L., Huang, J. and Carroll, R. J. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601-619.